In press at *Social Development*

Children's restorative justice in an intergroup context

Xin Yang [a]

Zhen Wu [a]

Yarrow Dunham [b]

a. Department of Psychology, Tsinghua University, Beijing, China

b. Department of Psychology, Yale University, New Haven, USA

Please send correspondence to Zhen Wu, Department of Psychology, Tsinghua University, Beijing, China, 100084 (zhen-wu@mail.tsinghua.edu.cn).

Author Note:

Xin Yang is currently affiliated with Yale University.

Abstract

The present study investigated 3- to 6-year-old (total n = 158) children's restoration behaviors both when they were second-party victims (Experiment 1) and when they were third-party bystanders (Experiment 2) of transgressions. We also explored how group membership (based on color preference) affects these behaviors. We found that children preferred restoration to punishment, and that they emphasized restorative justice not only for themselves but also for others. Furthermore, when they were victims of transgressions, the tendency to choose restoration over punishment was stronger in older than younger children. Second-party restoration behavior was influenced by group concerns in that children treated ingroup transgressors more leniently than outgroup and unaffiliated transgressors, but third-party restoration behavior was not. Our research challenges the view that punishment is the standard response to transgressions and suggests that alternative options like restoration are sometimes preferred over punishment by young children.

Keywords: restorative justice; restoration; punishment; group concern; minimal groups

Restorative justice in children in an intergroup context

Imagine someone takes something from you or from another person. How will you react? You might angrily seek revenge for yourself or for the victim, for example by taking things away from the transgressor to teach them a lesson. Alternatively, you might simply try to take your lost item back or return it to the victim without further punishing the transgressor. These actions reflect people's relative concerns between punishing the transgressor and protecting the victim. The former action focuses on severe punishment, while the latter reflects restorative justice that focuses on undoing harm and canceling out ill-gotten gains. Despite the prevalence and potential benefits of restorative measures in our society (e.g., Saulnier & Sivasubramaniam, 2015; Weitekamp, 1993), the literature has focused on studying punishment with economic games like the ultimatum game or third-party punishment games (e.g., Fehr & Fischbacher, 2004; Jordan, McAuliffe, & Warneken, 2014; McAuliffe, Jordan, & Warneken, 2015; Wu & Gao, 2018). Very few studies have investigated people's sense of restorative justice (with few exceptions; Chavez & Bicchieri, 2013; FeldmanHall, Sokol-Hessner, Van Bavel, & Phelps, 2014; Heffner & FeldmanHall, 2019; Riedl, Jensen, Call, & Tomasello, 2015).

When adults are asked to choose between different response options including both restorative and punitive options, they sometimes prefer more restorative actions both when their own welfare is affected (FeldmanHall et al., 2014; especially when the transgression is not too severe, see Heffner & FeldmanHall, 2019) and when they are third-parties of the transgression (Chavez & Bicchieri, 2013; cf. FeldmanHall et al., 2014). Does such a preference develop after experiences with justice restoration, or is it an early-emerging preference prior to prolonged social learning? Testing young children before formal schooling would offer insights into this

issue, as they likely have less experience compared to older children and adults. More specifically, past work looking at children's early sense of justice and fairness have shown that second-party punishment (children are victims) emerges around ages 3 and 4 (Wu & Gao, 2018), and third-party punishment (children are bystanders) emerges between ages 5 and 6 (e.g., McAuliffe, Jordan, & Warneken, 2015). In the current study, we therefore explore how 3-6-year-old Chinese preschoolers choose between restoration and punishment. We also investigate how social relationships, i.e. whether the transgressor belongs to the same or different social groups as the victim, might influence children's behavior across second-party and third-party contexts.

As far as we know, there has been only one developmental study looking at children's restoration behavior (Study 2 in Riedl et al., 2015). In their study, after an object was taken away from the victim, three-year-olds could choose to restore the object (i.e., returning it to the victim), discard the object (i.e., throwing it away, which they call punishment), or do nothing. The critical finding was that three-year-olds overwhelmingly chose to restore the object to the victim rather than punish the transgressor (very few children chose to do nothing). While this finding was taken to indicate a preference for restoration over punishment, there were confounding factors that hamper strong interpretations. In Riedl et al. (2015)'s study, participants essentially chose between removing the object such that no one gets it (punishment option), and returning this object to the victim (restoration option). This design makes the punishment option similar to the restoration option (in that the transgressor always lost the object) and fairly mild (i.e., the transgressor did not lose anything owned in the first place). This might favor the restoration option: why would people throw the object away when they can return this object to the rightful owner? Therefore, it leaves open questions as to children's developing preferences

for restoration versus punishment when the punishment option delivers a stronger sanction to the transgressor (thus increasing the deterrent effect of punishment).

Furthermore, preferences for restoration or punishment may vary according to the group membership of the transgressor, which was rarely tested in prior studies. In the context of punishment research, one influential view is people generally place higher expectations on ingroup members (such as the expectation to not harm but help and share resources with fellow group members; for developmental work, see e.g. Dunham, Baron, & Carey, 2011; Rhodes & Chalik, 2013) so they might punish ingroup transgressors more harshly to enforce group norms (Fehr & Fischbacher, 2004; Schmidt et al., 2012; for developmental work, see Yudkin, Bavel, & Rhodes, 2020). Alternatively, due to ingroup favoritism, i.e., preferential treatment of ingroups compared to outgroups, people might punish ingroup transgressors *less* harshly than outgroups (for a review of these two alternatives, see McAuliffe & Dunham, 2016). Ingroup favoritism is well-documented across various social groups (including gender, race, language, nationality, religion, etc.; e.g., Barret, 2007; Kinzler, Dupoux, & Spelke, 2007; Yee & Brown, 1994). It even robustly appears in lab-created, arbitrary, and meaningless "minimal" groups (when groups are defined with meaningless terms e.g. via a random assignment) when testing adults and children (Tajfel, 1970; for developmental work, see Abrams et al., 2008; Bigler et al., 1997; Dunham et al., 2011; Yang & Dunham, 2019; for a review, see Dunham, 2018). Some recent work suggests that this minimal ingroup favoritism appears in children as young as age three (Richter, Over, & Dunham, 2016). Therefore, our study examined whether group membership influences children's preference between restoration and punishment in 3-6-year-olds, when minimal group bias has been documented.

Which of these two accounts, ingroup expectations vs. ingroup favoritism, fits better with people's actual behaviors? Prior studies investigating this issue have found mixed results using classic economic games like the ultimatum game, which contrasts the option of punishment with doing nothing in response to fairness violations. Generally, when people are third-parties to unfair treatments, they show ingroup favoritism—they are less punitive towards ingroup transgressors (e.g., Bernhard, Fischbacher, & Fehr, 2006; Baumgartner, Götte, Gügler, & Fehr, 2012; Schiller, Baumgartner, & Knoch, 2014; but see Yudkin et al., 2020). A similar pattern was also found in 6- to 8-year-olds, suggesting that this tendency is present in children aged 6-8 (see Jordan et al., 2014), yet we know little about how such group biased third-party punishment develops in younger children. In addition, results are less consistent when people are directly affected (as second-parties). Some studies find ingroup favoritism (e.g., Kubota, Li, Bar-David, Banaji, & Phelps, 2013; Valenzuela & Srivastava, 2012), while others show the opposite pattern (e.g., McLeish & Oxoby, 2011; Mendoza, Lane, & Amodio, 2014; Wu & Gao, 2018; this is termed the "Black Sheep Effect", see Abrams, Palmer, Rutland, Cameron, & Van de Vyver, 2014), and yet another recent developmental study does not find group bias in either direction (McAuliffe & Dunham, 2017). These inconsistent findings may be due to different methodologies, cultures, or ages. Therefore, more studies are needed to clarify the directionality of group bias.

Overall, these above studies mainly focused on punitive response to unfairness. Meanwhile, humans also show strong tendency to restore justice, which centers on compensating the victim's welfare and cancelling out ill-gotten gains (e.g., FeldmanHall et al., 2014; Heffner & FeldmanHall, 2019). Studying the priority of punishment versus restoration will greatly enrich our understanding about the principles we use when dealing with injustice, and might offer more

insight into the underlying motives that drive people's behaviors in the face of unfairness. However, so far no studies have investigated how children choose between punishment and restorative options in a group context. To fill this gap, the current study directly compared children's choices both as victims (Experiment 1) and as bystanders (Experiment 2) in the face of ownership transgressions. In our study, we operationalized the punishment option such that the transgressor not only lost the object they had previously taken, but also had to pay an additional cost for the transgression. Punishment thus involved delivering a punitive sanction to the transgressor but did not provide restoration for the victim. In contrast, the restoration option involved returning the object to the victim without reducing the transgressor's original possession. Thus, this design allows us to look at children's developing preferences for restoration without *further* retribution (towards the transgressor's original possession) versus retribution without restoration.

Our primary hypothesis is that children would favor restorative options over punishment (based on past work in children, see Riedl et al., 2015, and adults, e.g., FeldmanHall et al., 2014; Heffner & FeldmanHall, 2019). This prediction is also consistent with past research on the development of children's sense of fairness (for a review, see McAuliffe et al., 2017), as children increasingly favor fair outcomes (even over having more than others) as they age. However, it is unclear whether children will choose restorative options more as victims or as bystanders, given the conflicting results in adults and little past work with children (see Chavez & Bicchieri, 2013; FeldmanHall et al., 2014). Another unexplored issue is how these preferences change with age, as past work only examined adults' behaviors and 3-year-olds' behaviors in separate studies (Chavez & Bicchieri, 2013; FeldmanHall et al., 2014; Riedl et al., 2015). Thirdly, the current study explored how the minimal group membership of the transgressor affects children's

behaviors. Given the inconsistent findings on group norm enforcement (leading to harsher punishment on ingroups) and ingroup favoritism (leading to less harsh punishment on ingroups), it remains unclear what patterns of group differences would emerge in our studies. Also, since minimal group studies are mostly conducted in Western, individualistic societies, it is important to extend this work to different cultures; in particular, in adults, collectivist cultures may show less ingroup bias in the minimal group setting (Falk, Heine, & Takemura, 2014). Overall, the present study aimed to fill in the gap by investigating how Chinese preschoolers choose between restoration and punishment when facing norm violation in a group context, and how such preferences differ between behaving as second-party victims and third-party bystanders.

Experiment 1

Methods

Participants

Power analysis using G*Power 3.1 showed that we needed $n = 90$ to be able to have >80% power to detect the main effect of children's preference between restoration and punishment (two tails, $\alpha = .05$, assuming small to medium effect size $d = .3$). We recruited 94 participants in order to account for potential exclusions. Participants were tested in a quiet room either in the lab or at local preschools in Beijing, China. We compared 3- to 4-year-olds and 5- to 6-year-olds, because past work finds that the latter age group (but not the former one) showed group biases in second-party punishment (Wu & Gao, 2018). Also, the ability to infer others' mental states (which plays an important role in children's fairness-related behaviors, e.g., Takagishi et al., 2010; Tsoi & McAuliffe, 2019; Wu & Su, 2014) undergoes critical developmental changes between the ages 4 and 5 (Wellman et al., 2006; Wimmer & Perner,

1983). The final sample included 86 preschool children, with 43 3- to 4-year-olds (M = 4.30yr, SD = 0.48, range = 2.96-4.95yr, 24 female) and 43 5- to 6-year-olds (M = 5.58yr, SD = 0.47, range = 5.00-6.55yr, 21 female). An additional 8 children were tested but excluded from data analyses for experimenter error (n = 4), refusal to complete the experiment (n = 3), or failure to pass the comprehension questions (n = 1). Experiments reported in this paper were approved by Tsinghua University Institutional Review Board (protocol #201602). Parental consent was obtained via the preschools in advance of all testing.

Design

We employed an experimental set-up that is at least somewhat more naturalistic and understandable to children as young as three, particularly in that it does not involve as much currency/market-based reasoning as economic games. This is an important consideration given that recent findings suggest an earlier emergence of certain forms of punishment behaviors (i.e., costly third-party punishment) in more naturalistic settings (Yudkin et al., 2020). There were two experimenters in the testing room: E1 was the main experimenter and E2 only operated puppets. Children were tested individually. They played 3 one-on-one games with 3 puppets in a counterbalanced order (Group, within-subject): one puppet was the child's ingroup member, one was an outgroup member, and one had not joined either group (i.e., hereafter "unaffiliated" partner; following adult work that introduces a group-neutral member, Schiller et al., 2014). There were two types of trials: At the beginning of each trial, the child and the puppet each got 2 tokens. In the partner-intervening trials, the puppet stole a token from the child without permission; in the experimenter-intervening trials E1 moved one token from the child to the puppet (but the child participants could not direct any action towards E1 in the game setting). Therefore, both types of trials resulted in 1 token for the child and 3 tokens for the puppet,

creating a similar unfair outcome for the child; however, the critical difference was that the partner violated ownership rule in the partner-intervening trials but not in the experimenter-intervening trials. We designed the experimenter-intervening trials as the control situation, as children's behaviors there demonstrated how frequently they wanted to try different response options (as discussed in McAuliffe et al., 2015) and how they preferred different distributions in the absence of antisocial behavior by the puppet. Therefore, the difference between the two trial types would more clearly reveal children's punitive and restorative motives. There were 3 trials of each type, resulting in 6 trials with each puppet. which were presented in a randomized order.

The child was given two options: restoring equality by taking the token back (leading to 2 tokens for each, the *restoration* option) or costly punishing the puppet such that the puppet lost all 3 tokens while the child lost 1 token (leading to the final outcome of 1 token left for the child but 0 for the puppet; the *punishment* option). These options were made by pushing one of the two cars placed on the apparatus, which enacted the choice in a clear, salient and fun manner (Figure 1). Children's choices were dichotomously coded (0 = punishment, 1 = restoration) for data analyses.

Apparatus

See Figure 1.

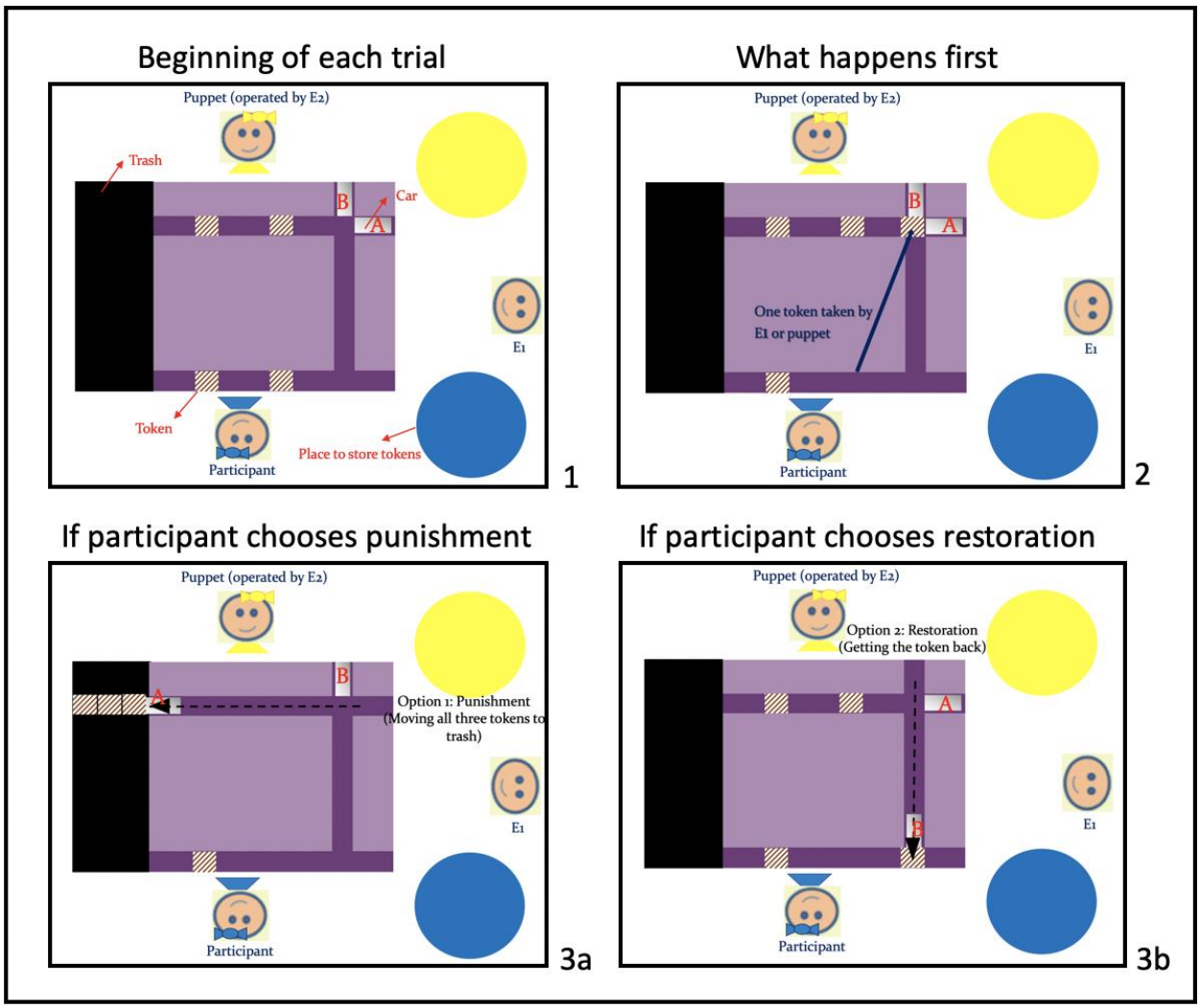

Figure 1. The layout of the apparatus. In this example, the participant was in the blue group, playing the game with an outgroup puppet. (1). Each trial began with E1 giving both sides 2 tokens. (2). Then either E1 (in the experimenter-intervening trials) or the puppet (in the partner-intervening trials) took a token away from the participant, resulting in a 1 (Participant): (3). Options and outcomes. (3a). If the child moved Car A on the right side of the apparatus (Option 1: punishment), it resulted in removing all 3 tokens (i.e., wooden blocks) to the inaccessible black area (final outcome: Participant got 1 but Puppet got 0); (3b). If the child moved Car B on the far side of the apparatus (Option 2: restoration), it resulted in restoring the lost token back to the child (final outcome: Participant got 2 and Puppet got 2).

Procedure

Group Assignment. Before the experiment, the child joined the "blue" or "yellow" team based on his or her self-reported color preference. The child picked the color s/he preferred from two pairs of little wooden shoes on the table (one pair was blue and the other pair was yellow), and then s/he was given the shoe s/he chose and assigned to that color group (52% participants chose blue). Next, the child put on a team shirt, a team waistband, and a team mascot (gender-matched headwear: girls got bowknots and boys got crowns), and s/he also got a storage board (all items were color-matched). Using multiple visual group markers was aimed to elicit an in-group bond in young children (Richter et al., 2016). We used the word "team" with children in order to help them better understand group memberships and to better induce group bias.

Group Introduction. E1 told the child that s/he would play one-on-one games with three "kids"—which were actually puppets operated by E2—one on the blue team, one on the yellow team, and one that was unaffiliated with either team (we always referred to the puppets as "kids" with the child; the unaffiliated puppet was described as someone who had not decided about which team to join). The blue-team and yellow-team puppets also had the color-matched group markers (i.e., toy shoes, shirts, waistbands, mascots, and the storage boards for blocks), but the unaffiliated puppet only had the storage board and importantly, it matched the apparatus' color (i.e., purple) in order not to give it any sort of group identity or involve any new color. In this phase, the child got familiarized with the three puppets. Puppet assignment (i.e., which puppet was assigned to which group) and sequence of group introduction were counterbalanced across participants.

Game Introduction. The game rule and the novel apparatus were introduced to both the child and the puppets by E1. At the start of each trial, both the child and the puppet had two tokens, which they knew they could exchange for stickers in a one-to-one fashion after the game.

To operationalize the randomized trial order in an engaging manner, E1 asked the child to pick one stick from 6 wooden sticks without replacement, which determined the order of trials. To demonstrate the outcome of two trial types, E1 explained that if the child picked a marked stick, then she would move one of the child's tokens to the puppet; by contrast, if the child picked an unmarked stick, then E1 said that she would not take any action, but the puppet would move one of the child's tokens without permission. Whichever case happened, the distribution of tokens changed from 2:2 (2 tokens for each one) to 1:3 (1 for the child and 3 for the puppet), creating disadvantageous inequality for the child.

The two trial types were demonstrated to children in a counterbalanced order and then E1 asked a set of comprehension questions (about the number of tokens himself/herself and the other puppet would get if s/he chose punishment or restoration options; "You can choose to push this car or that car. If you push this car [demonstration], how many tokens will you get? How many tokens will your partner get? If you push that car [demonstration], how many tokens will you get? How many tokens will your partner get?). E1 always asked children how many tokens they would get before asking how many tokens the puppet would get. Children had to pass all the comprehension questions before the *Experimental Session*. Only one participant failed the comprehension check and was excluded.

The words "punishment/punish", "restoration/restore", or "thief/steal" were never mentioned during the game; instead, E1 directly demonstrated the options and outcomes with cars and tokens.

Experimental Phase. The child played three one-on-one games following the same order as in *Group Introduction*. In each game, s/he was asked by E1 to greet the puppet (operated by E2) and was reminded of the puppet's group membership. The s/he played the game as described

above. More specifically, when the participant picked a stick with a mark, E1 said "This stick has a mark on it, so I will move one of your tokens" and then performed the action. When the participant picked a stick without a mark, E1 said "This stick does not have a mark on it, so I will not move your tokens". Next, the puppet partner said "I will take one of your tokens" while performing this action. After the tokens were moved, E1 asked the participant "Which car do you want to push?" and recorded behavioral response on a piece of paper. After each game, the child said goodbye to the puppet, exchanged for the stickers, and then started the game with the next puppet.

Group Bias Measures in Manipulation Check. After playing with all three puppets, E1 showed the child all puppets again and asked the following questions and explanations for his/her answers: 1) *Who do you like the best? [Explicit attitude]; 2) Here is a sticker and you can give it to one of them. Who do you want to give it to? [Resource allocation]; 3) One of these three kids helped her/his mom do chores today. She/he is a good kid. Who did the chores? [Behavioral attribution]*.

These three questions were used to check whether we successfully induced intergroup bias, as past research has shown that children of similar ages generally express explicit liking for ingroups, share more with ingroups, and attribute more positive behaviors to ingroups (Dunham, Baron, & Carey, 2011; Yang & Dunham, 2019).

Final Explanation. In the end, E1 asked the child why she pushed certain car(s) in the last trial and recorded their answers. We treated this as an exploratory measure and only reported descriptive results in Supplemental Materials.

Coding and reliability

E1 videotaped the experiment and manually coded the children's behavioral data (punishment = 0, restoration = 1) and answers to manipulation check questions during the experiment. A research assistant, who was unaware of the study design and hypothesis, independently coded 30% of the videos. Interrater agreement was 100% (for behavioral data, Cohen's $\kappa = 1.00$; for manipulation check questions, Cohen's $\kappa = 1.00$).

Results

We first report results on manipulation check questions, then main analyses examining children's choices between restoration and punishment options and the developmental trajectory. Preliminary analyses revealed no effect of participant gender, color group assignment, or order (all ps > .27), so these factors were not included in our models. Exploratory analyses of trial number and open-ended explanations are reported in Supplemental Materials.

Group bias in manipulation check

Children chose the ingroup puppet 47% of the time, the outgroup puppet 28% of the time, and the unaffiliated puppet 25% of the time. A chi-square test showed that they chose the ingroup puppet significantly more frequently than the other two options, $\chi^2(2, N = 86) = 22.07$, $p < .001$, indicating that our manipulation succeeded in inducing ingroup bias.

Restoration versus punishment choices

We fit a binomial mixed effects model predicting children's choices (0 = punishment, 1 = restoration) as a function of group (ingroup, outgroup, and unaffiliated), trial type (partner-intervening and experimenter-intervening), age group (3-4yr and 5-6yr), and their interactions, with a random intercept for participant. We found a significant group by trial type interaction (likelihood ratio test, $\chi^2(2, N = 86) = 8.16$, $p = .02$) and an effect of age group ($B = 1.25$, *SE*

= .42, 95% CI [.43, 2.07]). The effect of group was not significant (likelihood ratio test, $\chi^2(2, N = 86) = 2.05$, $p = .36$). Central to our primary focus, older children were more likely to prefer restoration over punishment. As shown in Figure 2, children by ages 5 and 6 clearly preferred restoration over punishment (all *p*s < .001, intercepts comparing to chance) while 3-4-year-olds did not show a clear preference. Analyses treating age as a continuous variable also showed similar results ($B = .90$, $SE = .26$, 95% CI [.38, 1.42]).

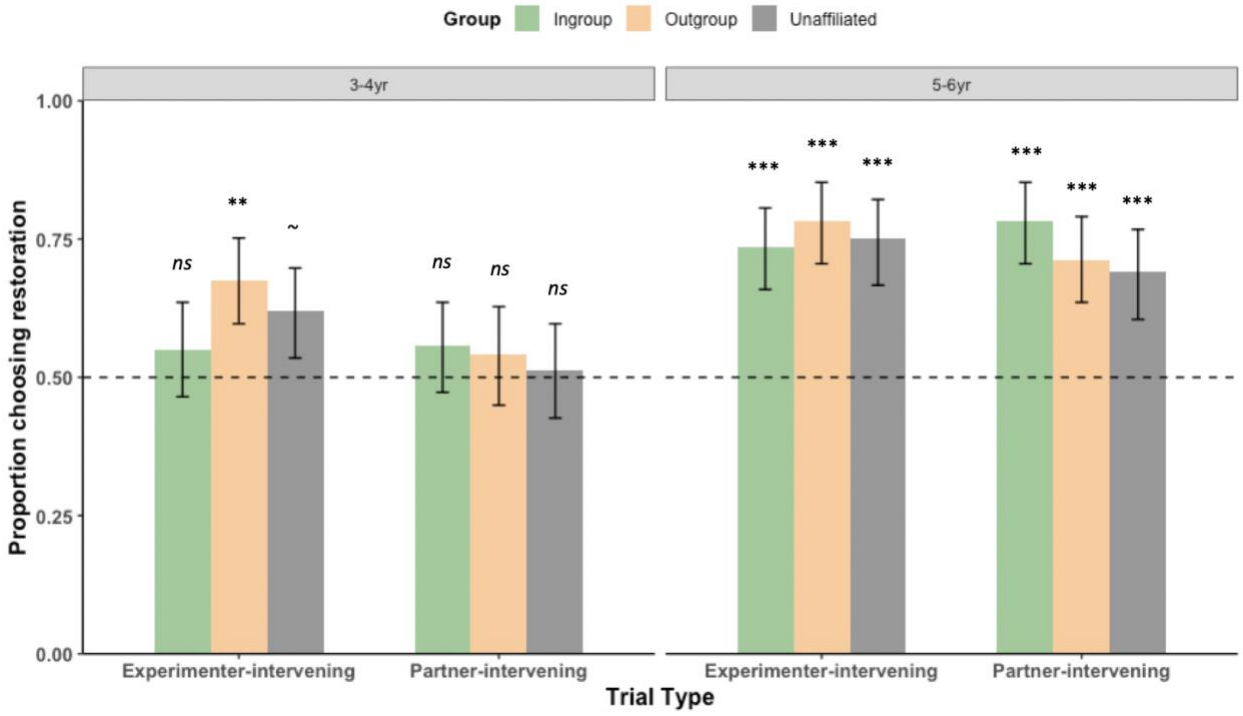

Figure 2. Experiment 1: Forced-choice between punishment and restoration (y-axis shows the proportion choosing restoration instead of punishment, possible range 0 to 1) across trial types (experimenter-intervening and partner-intervening trials), groups (ingroup, outgroup, and unaffiliated partners), and age groups (3-4yr and 5-6yr). Error bars are bootstrapped 95% confidence intervals.

The group by trial type interaction was driven by significant effects of trial type with outgroup and unaffiliated puppets but not with the ingroup puppet. As shown more clearly in Figure 3, compared to the experimenter-intervening trials, children directed restoration at outgroup and unaffiliated puppets less (directed punishment more) in the partner-intervening trials (outgroup, B = -.66, SE = .22, 95% CI [-1.10, -.22]; unaffiliated, B = -.54, SE = .22, 95% CI [-.97, -.11]), suggesting stronger punitive motives toward outgroup and unaffiliated partners in partner-intervening trials. However, this punitive motive weakened for the ingroup puppet, as children chose restoration and punishment similarly with the ingroup across the two trial types (B = .17, SE = .22, 95% CI [-.26, .60]). Unexpectedly, in experimenter-intervening trials children chose restoration less (punishment more) with ingroup partners than with outgroup ones, B = -.57, SE = .23, p = .01 (the other pairwise comparisons were not significant, ps > .19).

Adding children's ingroup bias levels shown in manipulation check questions (answering in favor of the ingroup puppet on 0 to 3 questions) to the model did not reveal an interaction between bias and group (p = .92). Children who showed different levels of ingroup bias on the manipulation check did not respond differently towards different groups in choosing restoration and punishment. To demonstrate this finding in a different way, we also classified children into three categories based on how they chose between punishment and restoration with ingroup and outgroup members: those who punished the ingroup less than the outgroup (N = 19), those who punished the ingroup more than the outgroup (N = 21), and those who punished the ingroup and the outgroup at the same degree (N = 46). The three categories of children did not differ in their levels of ingroup bias on the manipulation check (one-way ANOVA, p = .13). Taken together, children's ingroup biases on the manipulation check questions concerning ingroup preference did not seem to be related to their responses towards different groups during the game.

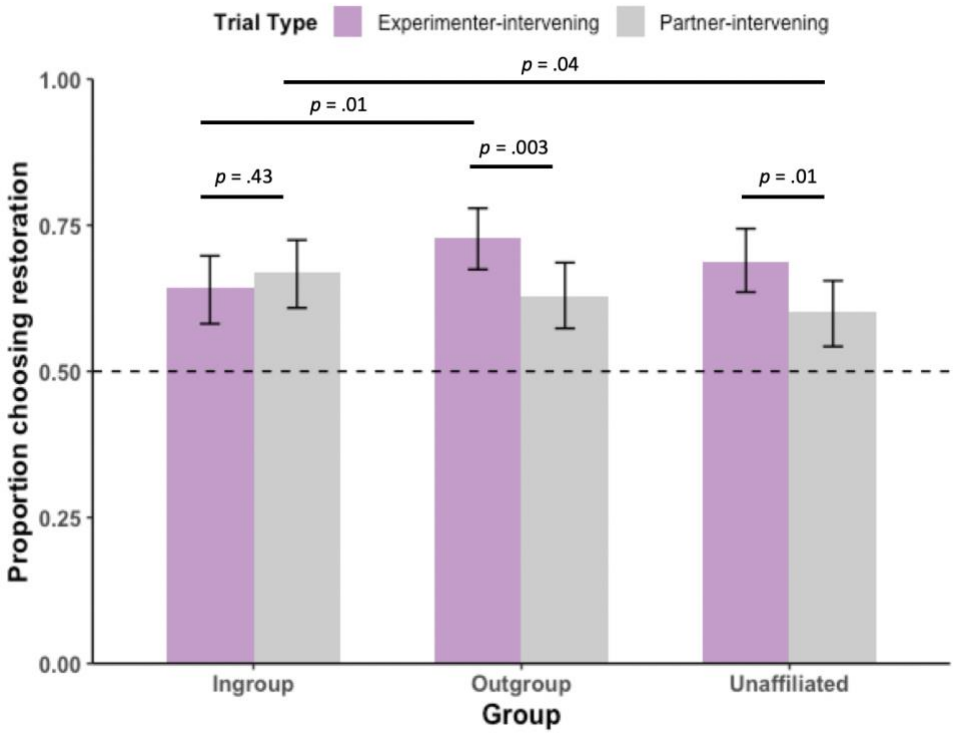

Figure 3. Experiment 1: Interaction between group and trial type (y-axis shows the proportion choosing restoration instead of punishment, possible range 0 to 1) averaged across participants (two age groups were combined because there was no interaction involving age group). Error bars are bootstrapped 95% confidence intervals.

Discussion

We successfully induced ingroup biases in 3- to 6-year-old Chinese preschoolers and measured their punishment and restoration behavior. Results showed that children chose restoration (i.e., getting their object back/the "2:2" option) more frequently than punishment (i.e., punishing the partner/the "1:0" option). We also found that 5- to 6-year-olds chose restoration more than did 3- to 4-year-olds, a developmental pattern first shown in the field. Furthermore, we showed that children's restoration and punishment behaviors were affected by group membership (via an interaction with trial types): when the puppet intervened and created an unfair outcome compared to when it did not, children were more likely to punish the outgroup

and unaffiliated puppets (i.e., they chose restoration less), but not the ingroup puppet. The direction of this group effect suggests ingroup love but not outgroup hate; that is, the difference between the ingroup and the outgroup was driven by the ingroup moving away from the neutral baseline, not the outgroup moving away from the neutral baseline.

Importantly, this is one of the first studies that demonstrate a preference for restoration over punishment in young children and the first study that reveals an age-related increase in that preference (see also Riedl et al., 2015). Children might develop an early preference for restoration because they view it as a better alternative to harsh punishment, i.e., it restores justice by delivering a moderate level of cost to the violator while also compensating the victim. Our findings also imply that, when restoration is available (as in our paradigm), children increasingly favor restoration with age. Our study suggests that it may be important to provide participants with both punishment (violator-oriented) and restoration (victim-oriented) options in order to get a more complete picture (see also Chavez & Bicchieri, 2013; FeldmanHall et al., 2014; Heffner & FeldmanHall, 2019).

This is also one of the first studies to demonstrate something much like a classic minimal group effect in non-Western children (see also Wu & Gao, 2018), and suggests that the tendency to immediately prefer social ingroups is culturally widespread.

Experiment 2

Experiment 1 found that when children's own interest was at stake, children aged 5-6, but not 3-4, robustly chose restoration over punishment. Experiment 2 aimed to replicate and extend the findings in the second-party context (children as victims) to a third-party context (children making choices for a victim). In a third-party context, children's choices may more clearly reveal

the balance between restorative and punitive motives without being affected by their self-interested concerns. Since intervening as a third-party appears later in development (e.g., between age 5 and 6 in third-party punishment; e.g., McAuliffe et al., 2015), here we focused on 5-year-olds and 6-year-olds.

Method

Participants

Power analysis using G*Power 3.1 showed that we needed $n = 71$ to be able to have >80% power to detect the main effect of children's preference for restoration over punishment (because we had prior prediction of direction based on Study 1 and Riedl et al., 2015; $\alpha = .05$, expecting small to medium effect size $d = .3$). We recruited 81 participants in order to account for potential exclusions. Participants were tested in a quiet place at a local preschool in Beijing, China. The final sample included 72 preschool children ($M = 5.99$yr, $SD = 0.38$, range = 5.13-6.68yr, 36 females). There were 36 5-year-olds ($M = 5.66$yr, $SD = 0.23$, range = 5.13-5.99yr, 19 females) and 36 6-year-olds ($M = 6.32$yr, $SD = 0.19$, range = 6.03-6.68yr, 17 females) by median split. An additional 9 children were tested but excluded from data analyses for experimenter error ($n = 3$), failure to complete the experiment ($n = 3$), or failure to pass the comprehension questions ($n = 3$). Parental consent was obtained via the preschool in advance of all testing.

Design

For simplicity of design, the victim puppet was always the child's ingroup member, while the other puppet partner was from an ingroup, an outgroup, or neither group (unaffiliated). The child was asked to help the ingroup victim puppet choose what to do. The child always got a

fixed amount of payoff for participation regardless of the choices they made. Other changes from Experiment 1 were detailed in Procedure.

Apparatus

Similar to Figure 1 except for minor procedural changes that are detailed in the following section.

Procedure

There was a slight change of order, in that we added a manipulation check before all testing to get baseline intergroup attitudes, and that we introduced the game rule before introducing the puppets (there were 4 puppets in this experiment). We made the partner-intervening trials more salient by highlighting the fact that the partner stole the possession of the victim against its will (see below). We used a blocked design for the two trial types rather than a trial-by-trial lot-drawing procedure to avoid confusion. We also made a few other changes to clean up procedures and avoid potential confounds, detailed below.

Group Assignment. The child was asked about his/her preferred color between "blue" and "yellow" (51% participants chose blue). Instead of a storage board, the child got a non-transparent storage box with a color-matched lid so that s/he could not easily see how many blocks each one has, decreasing the possibility of comparing the number of tokens with each other.

Group Bias Measures in Manipulation Check. Instantly after *Group Assignment*, the child was shown 6 identical drawings of two side-by-side kids (one wearing blue markers and the other one wearing yellow markers) sequentially. We first checked that s/he knew which groups these kids belonged to and whether they were his/her ingroups. After s/he had successfully answered this experimenter-intervening question, we asked 3 "Explicit attitude"

questions ("*Who do you like better*?") as we showed s/he the first 3 drawings one by one. Then, we asked 3 "Behavior attribution" questions as we showed s/he the last 3 drawings (the order of the 3 questions were counter-balanced across participants): *1) "One of these kids helped her mum with housework at home today. Who did the housework?" 2) "One of these kids helped another kid at the preschool today. Who helped that kid?" 3) "One of these kids brought juice for her friends today. Who brought juice for her/his friends?"* These three questions were modeled after past research (Dunham, Baron, & Carey, 2011), which were shown to measure ingroup biases in young children (note that we included more questions than in Study 1 to better measure ingroup bias). Following prior work (Gonzalez, Blake, Dunham, & McAuliffe, 2020; Jordan et al., 2014; Wu & Gao, 2018), we added these questions before the main task to make sure the minimal group manipulation was successful before administering main behavioral measures.

Game Introduction. We introduced the child to the game which the other kids would come to play with s/he later. For ease of demonstration, E allocated the tokens unevenly, putting 1 token on one side and 3 tokens on the other side. The child was told that the tokens put on each side were given to the kid close by during the game (but during the demonstration there was no other "kid"), and one token could be exchanged for one sticker. We also told the child that s/he would get one page of stickers at the end of the game regardless of the outcome of the game (all participants said that they loved stickers, and we told them that the puppets loved stickers too). Comprehension questions were similar to those of Experiment 1 except for some minor changes in wording to refer to the two absent "kids". All but 1 participant passed the comprehension questions (who was then excluded from data analyses).

Puppets Introduction. The ingroup victim was always introduced first to the child; the order of introducing the other 3 puppets acting as the ingroup, outgroup and unaffiliated violators was counter-balanced across participants.

Experimental Phase. The ingroup victim, standing closest to the child, played one-on-one games with 3 puppets following the same order as in *Puppet Introduction*. In the partner-intervening trials, E gave each of the puppets 2 tokens. Then the ingroup victim left because it needed to go to the restroom, eat some food, or drink some water (the order of excuses was counterbalanced). After the ingroup victim left, the puppet partner took one of its tokens away and put it on its own side. Next, the victim returned and found that its token was taken away. E asked the child to help it push a car ("*He/She came back and found that one of his/her token was missing. But he/she couldn't reach the cars. Could you help him/her push a car?*"). This change aimed to make it clear that the partner stole a token from the ingroup victim against its will. In the experimenter-intervening trials, to match the outcome distribution, E gave 1 token to the ingroup victim and 3 tokens to the other puppet at the beginning of the trial. Unlike Experiment 1, here E did not change a 2:2 distribution to 1:3 by moving one of the tokens, so that we made the experimenter-intervening trials more distinctive in actions from the immoral trials but with the same outcome. Again, children could only punish the puppets but not the experimenter. Then E asked the child to push a car ("*Which car do you want to push?"*).

Explanation, Memory Check, Evaluation, and Final Manipulation Check. In the end, E1 asked the child the following questions:

1) Why s/he pushed certain car(s) in the games *[Explanation]*; As in Experiment 1, we treated this as an exploratory measure and only reported descriptive results in Supplemental Materials.

2) What the 3 puppets [point to the 3 partners] did in the games *[Memory Check]*; All participants were able to answer this question.
3) Evaluations of the partners (was good or bad) and emotion judgments of the victim (felt happy or sad). Each answer was followed with a question to measure the scale (e.g., was it a little bad or really bad?). We coded their answers on a 4-point Likert-like Scale [very bad = 1, very good = 4; very sad = 1, very happy = 4]. We always pointed to the certain puppets and called its name when we asked questions about it. Pictures of "thumb-up" and "thumb-down" (for evaluations), or "smiling face" and "crying face" (for emotions) were used to help children answer these questions *[Evaluation]*. We added this as an exploratory measure and report analyses of these questions in Supplemental Materials.
4) (Among the 3 partners) *Who do you like the most? [Final Manipulation Check]*.

Coding and reliability

E videotaped the experiment and manually coded children's behavioral data (0 = punishment, 1 = restoration) and answers to manipulation check questions during the experiment. A research assistant, who was unaware of the study design and hypothesis, independently coded 30% of the videos. Interrater agreement was excellent (for behavioral data, *Cohen's* $\kappa = 1.00$; for manipulation check questions, *Cohen's* $\kappa = 1.00$).

Results and Discussion

Data were analyzed in the similar models as specified in Experiment 1. Preliminary analyses revealed no effects of trial number, participant gender, or color group assignment, so we did not include these factors in the models (however, there were some unexpected order effects; they were detailed in Supplemental Materials). We also reported analyses of open-ended explanations and evaluation and emotion rating questions in Supplemental Materials.

Group bias in manipulation check

We first analyzed the manipulation check questions to confirm that intergroup bias was successfully induced. In the pre-game check (6 questions), children chose the ingroup member on 64.58% of the questions (forced-choice between an ingroup and an outgroup member; chance = .5), $t(71) = 7.18$, $p < .001$, Cohen's $d = .85$. In the post-game check (choosing among 3 puppets), 47 children said they liked the ingroup puppet the most, while only 17 favored the outgroup and 7 chose the unaffiliated, $\chi^2(2, N = 72) = 36.62$, $p < .001$. These results showed that our manipulation succeeded in inducing ingroup bias.

Primary results on restoration versus punishment choices

As in Experiment 1, we fit a binomial mixed effects model predicting children's choices (0 = punishment, 1 = restoration) as a function of group (ingroup, outgroup, and unaffiliated), type (partner-intervening and experimenter-intervening), age group (5yr and 6yr), and their interactions, with a random intercept for participant. There was no effect of age group ($\chi^2(1, N = 72) = .64$, $p = .42$), nor was there an effect of age if we included age as a continuous variable ($\chi^2(1, N = 72) = .93$, $p = .34$), so we collapsed both age groups.

As shown in Figure 4, the above model showed that children preferred restoration over punishment (ps < .01, intercepts comparing to chance). We also found a significant effect of trial type, $B = .63$, $SE = .15$, 95% CI [.34, .92]: children directed more restoration in the partner-intervening trials compared to the experimenter-intervening trials. In addition, the effect of group was not significant ($\chi^2(2, N = 72) = 4.94$, $p = .08$). Adding children's ingroup bias levels shown in manipulation check questions (answering in favor of the ingroup puppet on 0 to 6 questions on pre-game manipulation checks) to the model again revealed that the interaction between bias and group membership was not significant ($p = .28$). As in Experiment 1, we also classified children

into three categories: those who punished the ingroup less than the outgroup (N = 10), those who punished the ingroup more than the outgroup (N = 23), and those who punished the ingroup and the outgroup at the same degree (N = 39). The three categories of children did not differ in their levels of ingroup bias on the manipulation check (one-way ANOVA, p = 1.00). Taken together, children's ingroup biases on the manipulation check questions concerning ingroup preference did not seem to be related to their responses towards different groups in the restoration and punishment task.

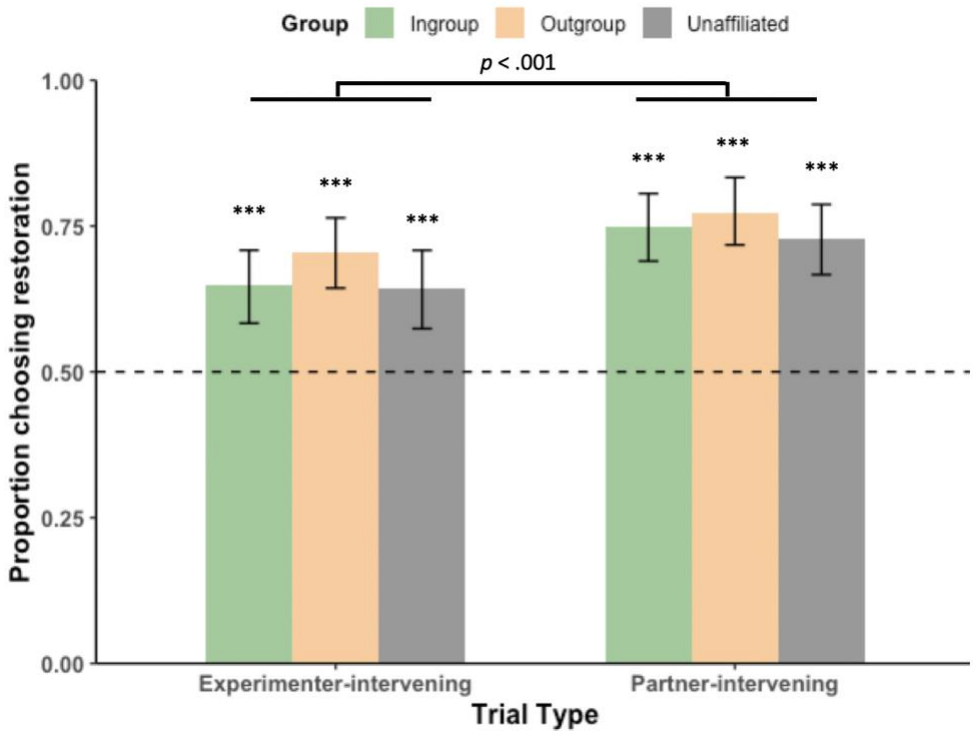

Figure 4. Experiment 2: Forced-choice between punishment and restoration (y-axis shows the proportion choosing restoration instead of punishment, possible range 0 to 1) across trial types (experimenter-intervening and partner-intervening trials), groups (ingroup, outgroup, and unaffiliated partners), with two age groups (5yr and 6yr) collapsed. Error bars are bootstrapped 95% confidence intervals.

Comparing Experiment 1 (second-party) and Experiment 2 (third-party) results (age-matched samples)

To compare 5-6-year-olds' restoration versus punishment choices across second-party and third-party contexts, we further analyzed data in two studies (only including 5-6y children in Experiment 1) predicting children's choices (0 = punishment, 1 = restoration) as a function of group (ingroup, outgroup, and unaffiliated), trial type (partner-intervening and experimenter-intervening), experiment (Experiment 1 and Experiment 2), and their interactions, with a random intercept for participant. The only significant effect was the interaction between trial type and experiment (see Figure 5), B = .83, SE = .24, 95% CI [.36, 1.30], which was driven by the effect of trial type being significant in Experiment 2 but not in Experiment 1. Children favored restoration over punishment more in the partner-intervening trials than in the experimenter-intervening trials in Experiment 2 (third-party), B = .62, SE = .15, 95% CI [.33, .92], but they did not choose these two options differently across trial types in Experiment 1 (second-party; but note that we found a significant group by trial type interaction there), B = -.20, SE = .19, 95% CI [-.57, .17].

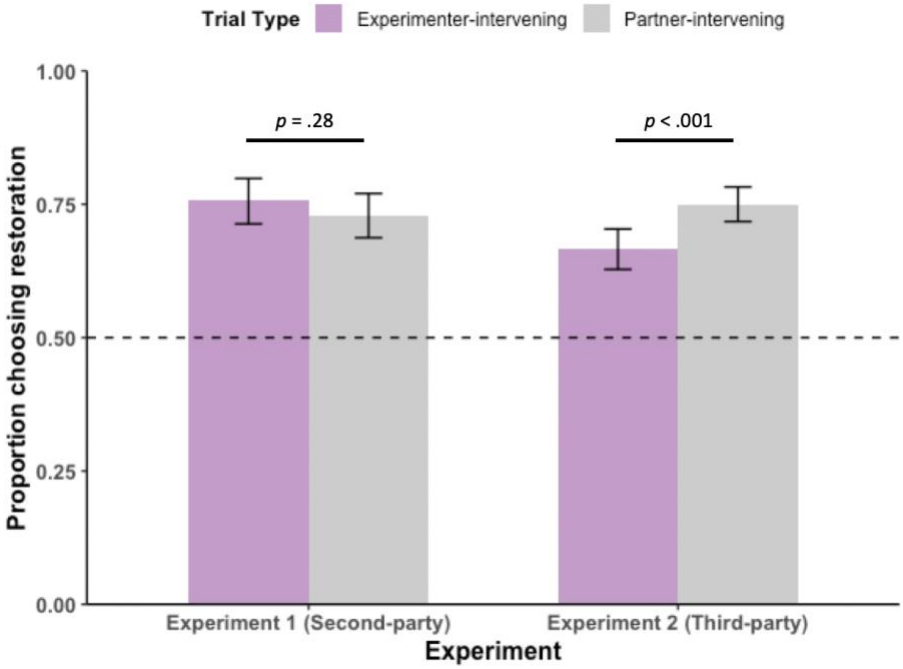

Figure 5. Experiment comparison (second-party Experiment 1 and third-party Experiment 2): Interaction between trial type and Experiment (y-axis shows the proportion choosing restoration instead of punishment, possible range 0 to 1), averaged across groups and participants (5-6-year-olds only). Error bars are bootstrapped 95% confidence intervals.

Discussion

In Experiment 2, we invited Chinese children to a third-party game when their own welfare was not involved. Similar to the finding from second-party game in Experiment 1, we found that 5- to 6-year-olds chose restoration more often than punishment. Importantly, children chose restoration more in the partner-intervening trials compared to the experimenter-intervening trials, suggesting strong preferences for restoration when witnessing a live transgression. In other words, children as young as 5 and 6 have a "victim-oriented" focus when they are not personally involved. Compared to Experiment 1, since children's own welfare was not affected here, these findings further imply a desire to return something to its rightful owner and a concern for

protecting the victim as opposed to more harshly punishing the transgressor, consistent with the broader claim made by Riedl et al. (2015). Moreover, we were the first to show that this pattern persisted regardless of the partner's group membership.

The relatively low levels of punishment as opposed to restoration again mirrored prior studies showing that children usually do not punish at above chance rates (e.g., Jordan et al., 2014; McAuliffe et al., 2015). These results also suggest that the punitive form of punishment might not be the preferred option in children. Future studies should consider alternative options like restoration when investigating punishment. Our findings are consistent with some adult work (Chavez & Bicchieri, 2013); we find that preschoolers favored restoration and that adult work adults compensated the victims more as compared to punishment in the third-party context (but see FeldmanHall et al., 2014 for somewhat different patterns).

General Discussion

Across two experiments, we investigated Chinese preschoolers' punishment and restoration behaviors in the face of transgressions both when their own welfare was involved (the second-party Experiment 1) and when it was not (the third-party Experiment 2). We found that children by ages 5 and 6 favored restoration over punishment, returning the object to its rightful owner without harshly punishing the transgressor across second- and third-party contexts. Interestingly, such a preference was more pronounced in the third-party context: children even increased restoration in partner-intervening trials (where they could punish the transgressor) compared to the experimenter-intervening trials (where they could not punish the experimenter who caused the unfair outcome and thus restoration was the expected choice, as per literature on children's sense of fairness, e.g. McAuliffe et al., 2017). We are the first to document such a

preference for restoration in a collectivistic culture as well as a clear age effect: older children (5-6yr) favored restoration over punishment more than younger children (3-4yr), suggesting that children develop this preference during preschool years. Our finding is consistent with recent adult and developmental studies using different paradigms in Western individualistic cultures (e.g., Chavez & Bicchieri, 2013; FeldmanHall et al., 2014; Heffner & FeldmanHall, 2019; Riedl et al., 2015), and with literature on theories of law and restorative measures (e.g., Saulnier & Sivasubramaniam, 2015; Weitekamp, 1993). These findings dovetail with research in children's developing sense of fairness, as in this age range children increasingly favor fairness between individuals and generally show a preference for fair outcomes (Killen & Smetana, 2005). The age-related changes revealed in these studies also parallel theory of mind work (Wellman et al., 2006; Wimmer & Perner, 1983). We propose that the understanding of others' mental states might be a potential mechanism underlying the developmental changes here (for related work on how mental state understanding is related to fairness preferences and sharing behaviors: e.g., Takagishi et al., 2010; Tsoi & McAuliffe, 2019; Wu & Su, 2014). We discuss other potential motivations underlying children's restorative and punishment behaviors in the General Discussion.

Across two experiments, we assigned Chinese children to novel social groups using a color-preference/participant-choice assignment procedure and measured resulting ingroup bias, following past work (Elashi & Mills, 2014; Gonzalez et al., 2020; Jordan et al., 2014; as well as a study that successfully used this same procedure in the same population, Wu & Gao, 2018). While not our main focus, our study extends the minimal group effect in children to a non-WEIRD (Western, educated, individualistic, rich, and democratic) culture, where developmental work on this effect is relatively rare (though see Wu & Gao, 2018). We showed that children

aged 3 to 6 favor their minimal ingroup members and evaluated them more positively, replicating past work on children's minimal group bias in WEIRD societies (e.g., Dunham et al., 2011). We acknowledge that other assignment procedures (e.g., pure random assignment and assignment allegedly based on personality) might lead to different levels of bias, and future work could further explore these differences (but see Yang & Dunham, 2019 for a recent study that found the random assignment procedure led to equally strong ingroup bias as more meaningful, personality-based groups).

Group membership significantly influenced choices in the second-party intervention context, but not in the third-party context. In the second-party context, compared to the experimenter-intervening trials, when puppets stole things children were more likely to punish outgroup and unaffiliated puppets (i.e., they chose restoration less), but not ingroup puppets. This result might imply ingroup love but not outgroup hate is driving the effect, which accords with and extends past research by adding a neutral, unaffiliated member (e.g., Bernhard. et al., 2006; Baumgartner et al., 2012). By contrast, in the experimenter-intervening trials when the puppet did not intentionally harm the child participant's welfare but the experimenter created a disadvantage for the child and an advantage for the puppet, children were less likely to choose restoration (instead favoring punishment) with an ingroup partner as compared to an outgroup partner. While we cannot definitively explain this effect, we suggest that the social comparison literature might provide some insight. In particular, children may have chosen the punishment option when they were motivated to get more than their partners (for similar findings in children with different paradigms, see Blake & McAuliffe, 2011; Sheskin et al., 2014). In other words, they might compare themselves more with ingroup members than outgroup members because social comparisons happen more with closer and more similar others (Chafel, 1985; Festinger,

1954; Pettigrew, 2016; Smith, Pettigrew, Pippin, & Bialosiewicz, 2012; Turner, 1975). This could motivate the desire to pursue a relative advantage in this case. Further, people in collectivistic cultures, e.g. China, have been shown to anticipate more threat and competition from ingroups (as compared to people in individualistic cultures like the US; Liu et al., 2019), which might motivate their enhanced comparison motives with ingroups in our task. In addition to social comparison, other motivations such as fairness concerns (see McAuliffe et al., 2017 for a review), self- or joint-outcome maximization, and spite (McAuliffe et al., 2014) might also facilitate children's behaviors in the current study. Future studies could look at different motives behind punishment and restoration behavior and explore what role each of these motivations plays.

We note that the restoration option in our design (following Riedl et al., 2015) is different from the compensation option in other studies. In those studies (e.g., FeldmanHall et al., 2014; Heffner & FeldmanHall, 2019), the compensation option increases the victim's payoff by introducing extra resources; whereas in our study, the restoration option restores the distribution to the original state ("returning the object to its rightful owner") and holds the total amount of resources constant. In both cases, however, the punishment option always reduces the perpetrator's payoff. Conceptually speaking, both designs are legitimate and commonplace: punishment options always focus on sanctioning the perpetrator, while the compensation/restoration options always focus on protecting the victim. We believe both designs reveal interesting insights into norm enforcement and future studies will be needed to explore people's preferences among various restorative/compensatory options. That said, we note that our design has some advantages in terms of everyday realism, at least in the most common childhood contexts, because returning stolen property to a victim is likely more frequent than

providing some additional compensation to victims while leaving the ill-gotten gains in the hands of a perpetrator. Future studies could compare these different designs as well as paradigms in the punishment literature (e.g., McAuliffe et al., 2017) to explore the shared and unique aspects of human cognition revealed by them.

It is also important to note that the punishment option in our study was designed to be a relatively severe form of punishment, in that the transgressor loses both of their own tokens in addition to the stolen one, thereby departing from Riedl and colleagues (2015), in which the transgressor only loses the one object that they stole). As we discussed in the Introduction, we operationalized the punishment option this way to better convey the idea of punishment as deterrent—though the victim does not get the lost token back, the transgressor has to pay a bigger price for the wrongdoing, thus potentially teaching them a lesson. However, this more severe form of punishment might have moved some children away from choosing punishment. Future studies could compare punishment options that differ in severity, e.g. the transgressor losing one token instead of both tokens, and explore if children's relative preference between restoration and punishment changes as the punishment option become more or less severe.

Another contribution of the current work is the inclusion of a group-neutral/unaffiliated perpetrator in addition to ingroup and outgroup transgressors, something that past research usually lacks. The value of the neutral group is that it allows us to better pinpoint the direction of this bias if it exists—i.e., is it driven by ingroup love (i.e., preferential treatment of ingroups), outgroup hate (i.e., discriminatory treatment of outgroups), or both? As far as we know, there is only one adult study that includes an unaffiliated violator (Schiller et al., 2014), and it finds that bias in third-party punishment is driven by both ingroup love and outgroup hate. In the current research, we explored this issue in children by including a group-unaffiliated individual

(following Schiller et al., 2014, we introduced it as someone who had not decided about which group to join and thus it had not joined either group). It thus allows us to further understand how intergroup bias affects norm enforcement, and whether it mirrors the developmental trajectory of intergroup attitudes (Buttelmann & Böhm, 2014; Nesdale, 2004). Unexpectedly, results for the neutral target did not always fall in between the ingroup and the outgroup. It is possible that children saw this undecided puppet as a complete outlier or outsider and therefore treated them even more harshly (i.e., perhaps via wondering why everyone belonged to group except this puppet). More research using neutral targets with children (as well as with different ways of introducing the neutral target) is needed to explore how children reason about the neutral/unaffiliated members and to more comprehensively investigate the directionality of intergroup bias.

We also acknowledge that across manipulation check questions and main behavioral trials we always pitted two options, e.g. ingroup vs. outgroup or restoration vs. punishment, against each other without providing other options like "both" or "neither". This use of dichotomous, forced-choice measures is common in the developmental literature, but we acknowledge that it might restrict children's responses and it will be important to examine how children might respond if other options are available. Additionally, our main behavioral trials involved decisions around unnecessary resources (stickers) as compared to necessary ones (such as medical or school supplies, which might create stronger concerns for equal distributions and thus reduce the possibility of finding ingroup bias, see Elenbaas & Killen, 2016; Rizzo, Elenbaas, Cooley, & Killen, 2016). A fruitful avenue of future work might be the investigation of children's restoration and punishment behavior across different types of resources.

In summary, we showed that preschoolers preferred restoration to punishment, and that they emphasized restorative justice not only for themselves, but also for others (i.e., an ingroup member) when their self-interest was not involved. There was an age-related increase in favoring restoration between age 3 and 6 (in second-party games). Children's responses were also shaped by intergroup concerns in complex ways. Our research challenges the view that punishment is the standard response to transgressions, and suggests that alternative options like restoration might instead be preferred by even young children. Future research is needed to explore the preference for restoration and how it is influenced by intergroup bias across development and cultures.

Reference

Barret, M. (2007). *Children's knowledge, beliefs and feelings about nations and national groups*. Psychology Press.

Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, *33*(6), 1452–1469. https://doi.org/10.1002/hbm.21298

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*, 912–915. https://doi.org/10.1038/nature04981

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences*, *18*(11), 559–561. https://doi.org/10.1016/j.tics.2014.08.003

Buttelmann, D., & Böhm, R. (2014). The ontogeny of the motivation that underlies in-group bias. *Psychological Science*, *25*(4), 921–927. https://doi.org/10.1177/0956797613516802

Chafel, J. A. (1985). Social comparison and the young child: Current research issues. *Early Child Development and Care*, *21*(1–3), 35–59. https://doi.org/10.1080/0300443850210104

Chavez, A. K., & Bicchieri, C. (2013). Third-party sanctioning and compensation behavior: Findings from the ultimatum game. *Journal of Economic Psychology*, *39*, 268–277. https://doi.org/10.1016/j.joep.2013.09.004

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780–793. https://doi.org/10.1016/j.tics.2018.06.004

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, *82*(3), 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Elashi, F. B., & Mills, C. M. (2014). Do children trust based on group membership or prior accuracy? The role of novel group membership in children's trust decisions. *Journal of Experimental Child Psychology, 128*, 88–104. https://doi.org/10.1016/j.jecp.2014.07.003

Elenbaas, L., & Killen, M. (2016). Children rectify inequalities for disadvantaged groups. *Developmental Psychology, 52*(8), 1318–1329. https://doi.org/10.1037/dev0000154

Falk, C. F., Heine, S. J., & Takemura, K. (2014). Cultural variation in the minimal group effect. *Journal of Cross-Cultural Psychology, 45*(2), 265–281. https://doi.org/10.1177/0022022113492892

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J., & Phelps, E. A. (2014). Fairness violations elicit greater punishment on behalf of another than for oneself. *Nature Communications*, *5*, 5306. https://doi.org/10.1038/ncomms6306

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. https://doi.org/10.1177/001872675400700202

Gonzalez, G., Blake, P. R., Dunham, Y., & McAuliffe, K. (2020). Ingroup bias does not influence inequity aversion in children. *Developmental Psychology*. https://doi.org/10.1037/dev0000924

Heffner, J., & FeldmanHall, O. (2019). Why we don't always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-49680-2

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences, 111*(35), 12710–12715. https://doi.org/10.1073/pnas.1402280111

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the Ultimatum Game. *Psychological Science*, *24*(12), 2498–2504. https://doi.org/10.1177/0956797613496435

Liu, S. S., Morris, M. W., Talhelm, T., & Yang, Q. (2019). Ingroup vigilance in collectivistic cultures. *Proceedings of the National Academy of Sciences*, *116*(29), 14538–14546. https://doi.org/10.1073/pnas.1817588116

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*, 0042. https://doi.org/10.1038/s41562-016-0042

McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Phil. Trans. R. Soc. B*, *371*(1686), 20150073. https://doi.org/10.1098/rstb.2015.0073

McAuliffe, K., & Dunham, Y. (2017). Fairness overrides group bias in children's second-party punishment. *Journal of Experimental Psychology: General*.

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*, 1–10. https://doi.org/10.1016/j.cognition.2014.08.013

McLeish, K. N., & Oxoby, R. J. (2011). Social interactions and the salience of social identity. *Journal of Economic Psychology*, *32*(1), 172–178. https://doi.org/10.1016/j.joep.2010.11.003

Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For members only: Ingroup punishment of fairness norm violations in the ultimatum game. *Social Psychological and Personality Science*, *5*(6), 662–670. https://doi.org/10.1177/1948550614527115

Nesdale, D. (2004). Social identity and ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The Development of the Social Self*. Psychology Press.

Pettigrew, T. F. (2016). In pursuit of three theories: Authoritarianism, relative deprivation, and intergroup contact. *Annual Review of Psychology*, *67*(1), 1–21. https://doi.org/10.1146/annurev-psych-122414-033327

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*(6), 999–1006. https://doi.org/10.1177/0956797612466267

Richter, N., Over, H., & Dunham, Y. (2016). The effects of minimal group membership on young preschoolers' social preferences, estimates of similarity, and behavioral attribution. *Collabra: Psychology*, *2*(1). https://doi.org/10.1525/collabra.44

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology*, *25*(13), 1731–1735. https://doi.org/10.1016/j.cub.2015.05.014

Rizzo, M. T., Elenbaas, L., Cooley, S., & Killen, M. (2016). Children's Recognition of Fairness and Others' Welfare in a Resource Allocation Task: Age Related Changes. *Developmental Psychology*, *52*(8), 1307–1317. https://doi.org/10.1037/dev0000134

Saulnier, A., & Sivasubramaniam, D. (2015). Restorative justice: Underlying mechanisms and future directions. *New Criminal Law Review: In International and Interdisciplinary Journal*, nclr.2015.150014. https://doi.org/10.1525/nclr.2015.150014

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior*, *35*(3), 169–175. https://doi.org/10.1016/j.evolhumbehav.2013.12.006

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, *124*(3), 325–333. https://doi.org/10.1016/j.cognition.2012.06.004

Smith, H. J., Pettigrew, T. F., Pippin, G. M., & Bialosiewicz, S. (2012). *Relative deprivation: A theoretical and meta-analytic review*. http://journals.sagepub.com/doi/abs/10.1177/1088868311430825

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102.

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology*, *5*(1), 1–34. https://doi.org/10.1002/ejsp.2420050102

Valenzuela, A., & Srivastava, J. (2012). Role of information asymmetry and situational salience in reducing intergroup bias: The case of ultimatum games. *Personality and Social Psychology Bulletin*, *38*(12), 1671–1683. https://doi.org/10.1177/0146167212458327

Weitekamp, E. (1993). Reparative justice: Towards a victim oriented system. *European Journal on Criminal Policy and Research*, *1*(1), 70–93. https://doi.org/10.1007/BF02249525

Wu, Z., & Gao, X. (2018). Preschoolers' group bias in punishing selfishness in the Ultimatum Game. *Journal of Experimental Child Psychology, 166*, 280–292. https://doi.org/10.1016/j.jecp.2017.08.015

Yang, X., & Dunham, Y. (2019). Minimal but meaningful: Probing the limits of randomly assigned social identities. *Journal of Experimental Child Psychology, 185*, 19–34. https://doi.org/10.1016/j.jecp.2019.04.013

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology*, *33*(2), 183–196. https://doi.org/10.1111/j.2044-8309.1994.tb01017.x

Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (in press). Young children police in-group members at personal cost. *Journal of Experimental Psychology. General*. https://doi.org/10.31234/osf.io/rc4ta

Data and analysis code that replicate all results in the manuscript and supplemental materials are also openly shared and are available on an anonymous link at: https://osf.io/87hme/?view_only=782b53ad77e94ea1a2b6dd2caee936c8.