

# Children's Group Bias in Evaluating other's Third-party Punishment

A Thesis

Presented to the Faculty of the Graduate School  
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of  
Master of Arts

by

Yunru Ma

August 2020



## ABSTRACT

Past studies have shown that people favor their own group members when they themselves are the third-party punisher. But in real life, most people do not have adequate power to enforce justice by themselves. Instead, they are most likely to enforce moral norms by choosing who to grant the power to punish (through election, petition, public discussion, etc.), and such choice relies crucially on their evaluation of the third-party punishment and the punisher. Do children have group bias when evaluating other's moral punishment at its emergence? In this study ( $N = 197$ , age 3-7), we found that in evaluating other's third-party punishment, children showed in-group favoritism toward in-group perpetrators and out-group discrimination toward out-group perpetrators, evaluating a third-party punishment more positively when it is enforced on an out-group perpetrator than an in-group perpetrator. However, the group affiliation of the victim did not induce biased evaluation for a third-party punishment. In this Chinese sample, we also found that children are biased in judging the deserved punishment of a transgression and in their third-party punishment behavior. They favored in-group perpetrators, which was in accordance with previous study conducted on western samples. We also found that over development, children judged a transgression as more serious and deserve more punishment, and gave increasingly higher evaluation toward other's third-party punishment. Taken together, this study added to the evidence that group bias is rooted deeply in ontology, and punishment toward injustice is biased even at its emergence.

*Keywords:* group bias, moral judgment, third-party punishment, moral development

### **BIOGRAPHICAL SKETCH**

Yunru Ma is a master's student at Department of Human Development. She holds a BSc in Psychology from Beijing Normal University, China. During her study at Cornell, she works with Dr. Qi Wang in the Culture & Social Cognition Lab. She is interested in the mechanisms underlying children's prosocial behaviors.

## **ACKNOWLEDGEMENTS**

I would like to thank Dr. Qi Wang for accepting me as her master student, giving me opportunities for doing this research, guiding me on how to become a qualified researcher and all the assistance about this study. Thanks to Huixian Yu and Xiaoqian Zhang for helping me collect data in China. Thanks to all the children who participated in our study. Thanks to my lab mates Li Guan and Tong Suo for helping me develop my idea for this study. Thanks to all the friends with whom I spent happy time during my master's study at Cornell.

## TABLE OF CONTENTS

<b>1 Introduction.....</b>	<b>7</b>
1.1 Third-party punishment in children .....	7
1.2 Children’s evaluation of punisher .....	8
1.3 Children’s in-group favoritism in third-party punishment .....	10
1.4 The present study .....	13
<b>2 Methods .....</b>	<b>14</b>
2.1 Participants .....	14
2.2 Procedure .....	15
2.3 Coding .....	16
<b>3 Results .....</b>	<b>17</b>
<b>4 Discussion .....</b>	<b>23</b>
<b>5 Reference .....</b>	<b>31</b>

## **Introduction**

### **1.1 Third-party punishment in children**

Conforming to moral rules and maintaining justice is important to guarantee reliable social interactions and cooperation (Sober & Wilson, 1998). But immoral behaviors (physical violence, stealing, robbery of property, etc.) often bring quicker and easier personal gain, which gives moral perpetrators more benefit than those who stick faithfully to moral rules. This makes the punishment on moral transgressions especially crucial (Boyd and Richerson, 1985, Nowak, 2006, Nowak and Sigmund, 2005). By punishing those who break the moral rules, we both deter future injustice and encourage people to act more prosocially in future interactions (Gummerum & Chu, 2014; Jordan, McAuliffe, & Warneken, 2014; McAuliffe, Jordan, & Warneken, 2015). Especially in large-scale social interaction which involves unfamiliar individuals, repeated social interaction between two same parties is less likely, making second-party punishment (i.e., when the punisher is the victim suffering from the transgression) less effective (Marlowe et al., 2008; Pedersen, Kurzban, & McCullough, 2013). And therefore the maintenance of social order in large-scale societies (which is more and more common) relies even more on third party punishment.

Past studies have proved that people's punishment on uncooperative, unfair, antisocial behaviors is not limited to the context when their own interest is directly conflicted. The third-party punishment is commonly found across human societies (Buckholtz et al., 2008; Henrich et al., 2010; Raihani et al., 2012). People are willing to pay to intervene in an unfair resource allocation, even when their economic payoff is not harmed (Fehr & Fischbacher, 2004; Henrich et al., 2006).

Same as adults, starting from a very young age children systematically punish people who violate moral norms (Furman & Masters, 1980). Current developmental studies about children's third-party punishment focus mainly on the age it emerges. Researchers found that 10-month-old infants already have expectations for how others would react as a third party, expecting punishment to be placed on an unfair distributor rather than a fair distributor (Meristo & Surian, 2014). At the age of 2, as an uninvolved observer, children voluntarily punish the hinderer by taking away the treatment (Hamlin, Wynn, Bloom, & Mahajan, 2011). And after observing one puppet destroying another puppet's belongings, 3-year-old children would intervene by protesting verbally during the transgression, punishing the perpetrator, and behaving prosocially toward the victim (Vaish, Missana, & Tomasello, 2011). By age 6, children start to robustly punish unfair individuals as an unaffected third party at a personal cost, selectively punishing the unfair distributor rather than the fair distributor (Jordan, McAuliffe, & Warneken, 2014; McAuliffe, Jordan, & Warneken, 2015). Aside from the binary option of punish vs. not punish, researchers have also found that 6-year-old children, but not 4- to 5-year-olds, have developed a deserve-based reasoning of punishment. When judging the deserved punishment of a hypothetical character, 6- to 10-year-old children assigned punishment in proportion to the blameworthiness of the bad deed (Smith & Warneken, 2016). To sum up, the ability to sanction injustice as a third-party emerges quite early in ontology, providing the cognitive basis for higher order moral evaluation of punishers.

## **1.2 Children's evaluation of punisher**

As is discussed in the previous section, to maintain reliable interaction in large-scaled society, it requires enforcing punishment toward moral violations. However, moral punishment often comes with risk and price (e.g., cost of resource, revenge), which results in a greater loss



for the punishers than for free riders (who benefit from the outcomes of the guaranteed cooperation without sacrificing any cost) (Vaish, Herrmann, Markmann, & Tomasello, 2016). This brings a problem: how can the moral punishment toward injustice be sustained? A possible mechanism that solves this puzzle is to positively evaluate and praise the third-party punishers. Studies on adults have found that compared to non-punisher, people give more reputational and material benefit to third-party punishers (Barclay, 2006; Nelissen, 2008; Horita, 2010).

Compared to the massive study on children's direct third-party punishment (the first-order norm enforcement), very few studies have looked at how children think about the punisher (the second-order norm enforcement) (Vaish, Herrmann, Markmann & Tomasello, 2016). However, this topic is worth exploring because first, in the crucial stage of social learning, knowing about children's judgment of the moral punishment helps us understand what kind of moral behavior they imitate and internalize; and second, with limited power children seldom have the chance to enforce moral punishment in real life.

Among the limited number of studies on children's evaluation of third-party-punishers, some empirical evidence suggested that the ability of understanding and admiring altruistic third-party punishment emerges very early in life. In Hamlin and colleague's study (2011), they found that 8-month-old infants preferred the agent who took away treatment from a hinderer than the agent who offered treatment to the hinderer. Similarly, after watching physical aggressions between two characters (e.g., a perpetrator hitting a victim), 6-month-old infants preferred the third-party agent who intervened than the agent who didn't (Kanakogi et al., 2017). However, since these two studies were conducted on preverbal infants, and used the rather vague clue of touching to reason children's preference, there could be alternative interpretation aside from infant's positive evaluation to the third-party punisher (Hamlin et al., 2011). Some studies on

older children provided stronger evidence. They found that in the scenario of unfair allocation, starting from age 5 children gave positive evaluation to third-party punishers (although they gave even higher evaluation to third-party helpers than third-party punishers) ( Lee & Warneken, 2020). Also, older children make positive evaluations toward third-party punishers based on more subtle clues: after watching a perpetrator broke the moral norm (e.g., destroying the victim's property), 5-year-old children more positively evaluated the third-party who scolded the perpetrator than the third-party who made neutral comments (Vaish, Herrmann, Markmann & Tomasello, 2016).

However, these existing studies on children's evaluation of third-party punishers focused only on comparing between punishers vs. non-punishers. They seldom focused on what factors could affect children's evaluation. To my knowledge, there is no study that looks at whether group membership affects children's evaluation on third-party punishment. Although moral rules inherently value impartiality, does that mean no group bias in children's evaluation on punishers?

### **1.3 Children's in-group favoritism in third-party punishment**

There is a rapid increase of developmental research on children's in-group favoritism and out-group bias. As a byproduct of social categories, group bias is deeply rooted in ontology. A wide range of categories can serve as the basis for children's in-group favoritism, like biological categories (e.g., gender, race), language, shared beliefs (e.g., religion), shared affiliation (e.g., supporting the same sports team), and shared origin (e.g., nationality) (Yee & brown, 1994; Kinzler, Dupoux, & Spelke, 2007; Heiphetz, Spelke, & Banaji, 2013; Abell, 2008). Moreover, many minimal group experiments found that even when the group is not meaningful and created based on highly superficial criterias (e.g., the color of T-shirt, coin flip, etc.), children's in-group favoritism is still robust (Tajfel, 1970; Dunham, 2018).

From the evolutionary angle, most altruistic behaviors (e.g., food sharing, cooperative hunting, self-protection against another tribe) happen among people who share the same group identity. Therefore to understand altruistic behaviors like third-party punishment, it is important to study the effect of group membership. Some researchers even proposed a mutual reinforcement among in-group favoritism, cooperation, and norm enforcement (Choi & Bowles, 2007). In Hetherington, Hendrickson, & Koenigchildren's study (2014), they set a scenario where group membership and moral behavior conflict. They found that when comparing an antisocial in-group member and a neutral out-group member, their preference for the antisocial in-group members was attenuated; however, when comparing a neutral in-group member and a prosocial out-group member, their preference for the ingroup member was less affected. Therefore, it is likely that group membership plays an important role on the altruistic punishment from the third-party, as well as how people evaluate the punisher (Schiller, Baumgartner, & Knoch, 2014).

Extensive studies on adults have discussed the discriminatory third-party punishment influenced by the group affiliation of both the perpetrator and the victim. After observing the same transgression, people judge the out-group perpetrator more harshly than the in-group perpetrator (Graham et al., 1997; Sommers and Ellsworth, 2000). Bernhard, Fischbacher and Fehr's study (2006) found in-group favoritism was only induced by the victim's group affiliation: third-party punishers protected ingroup victims much more than they did outgroup victims, punishing the perpetrator much harsher in the previous situation. In contrast, the group identity of the perpetrator doesn't induce discriminatory treatment in their study. Interestingly, in the same experiment they found that perpetrators would expect third-party punishers to be more lenient if they share the same group affiliation. However, another study suggested the group

affiliation of the perpetrator might also drive in-group favoritism: in the situation when an in-group member is the victim, third-party punishers were more likely to punish when the perpetrator was an out-group member than an in-group member (Schiller, Baumgartner, & Knoch, 2014).

With age children develop both the willingness to enforce justice and in-group favoritism (Jordan, McAuliffe, & Warneken, 2014). However, there are strikingly few developmental studies that explored the relationship between group membership and third-party punishment. In Jordan, McAuliffe, and Warneken's study (2014) on unfair allocation, they found by age 6 children already showed in-group favoritism in third-party punishment in a minimal group setting. Children punished more harshly on an unfair out-group allocator than an in-group one, and were more protective to an disadvantaged in-group recipient than an out-group one. Another developmental study conducted by Chapman and colleagues (2020) on 3- to 6-year-old children focused on the effect of the perpetrator's group affiliation. They found that group membership robustly affected the judgments of both the seriousness of transgressions and the quality of perpetrators. After listening to the story about physical aggression (e.g., hitting and kicking), children consistently judged a transgression as more serious and deserve more punishment when it was conducted by an out-group perpetrator than the one conducted by an in-group or neutral perpetrator; Out-group perpetrators are also judged more negatively than neutral and in-group perpetrators (Chapman et al., 2020). This result allowed us to look at group bias in a more detailed way: there was out-group discrimination in moral evaluation, but the in-group favoritism might not be salient. To sum up, these findings suggested that as a third-party, preschool age children were already biased in judging the seriousness, judging the deserved punishment, and assigning third-party punishment on a moral transgression. And the group affiliation of both the

perpetrator and the victim could induce this biased norm enforcement, even when the group is not meaningful (assigned based on the color of T-shirt).

#### **1.4 The present study**

Based on the above empirical evidence, we ask this question: will children also have in-group favoritism in their evaluation of a third-party punishment? This hypothesis, although hasn't been empirically tested, can be observed in some real life cases. In many controversial legal cases (which is a common type of third-party punishment), we often find the evaluations widely splitted, and people often stand on the side of their own social group members. In the *People v. Turner's* case, women like Michele Dauber were the first to stand out to argue against the sentence and raise a petition for recalling the unfair judge Aaron Persky from office. However, there also existed some voices that claim that the 6-month-sentence is not that unacceptable, mostly from people as privileged as the criminal. Therefore, we may infer that people have group bias in evaluating a third-party punishment.

We make this hypothesis because: by comparing previously mentioned studies on children's own choice of third-party response with studies on their evaluation of other's third-party response, we found that children's evaluation of other's third-party response aligns with their own choice of third party response. They punish hinderer rather than the helper (Hamlin, Wynn, Bloom, & Mahajan, 2011), and they prefer agents who punish the hinderer rather than the helper (Hamlin et al., 2011); they choose to intervene as a third-party to deter harmful behavior (Vaish, Missana, & Tomasello, 2011), and they also prefer third-party agent who intervened than the agent who didn't (Kanakogi et al., 2017); they voluntarily engage in costly third-party punishment (Jordan, McAuliffe, & Warneken, 2014; McAuliffe, Jordan, & Warneken, 2015), and similarly they have more positive evaluations toward third-party punishers than the non-

punishers (Vaish, Herrmann, Markmann & Tomasello, 2016; Lee & Warneken, 2020). Given that children show in-group favoritism in their own third-party punishing behavior, in the current study we aim to testify if this pattern is also valid in their evaluation of other's third-party punishment.

## **Methods**

This study used a 4 (group affiliation: in-group perpetrator & out-group victim, out-group perpetrator & out-group victim, in-group perpetrator & in-group victim, in-group perpetrator & out-group victim)  $\times$  3 (age: 3-4, 5, 6-7) between subject design. The in-group and out-group relation referred to the characters' group relationship with the participants, created using minimal group paradigm.

### **2.1 Participants**

Participants comprised 197 preschool children of 3-7 years old (88 girls and 109 boys) recruited from a kindergarten in Nanyang, China. Of these participants, 23 were 3-year-old, 58 were 4-year-old, 49 were 5-year-old, 44 were 6-year-old, 23 were 7-year-old. 6 other children were tested but excluded from data analysis because of refusing to answer questions or unable to pass check questions. The average household income in the tested kindergarten was 75,048 RMB (about 10,735 USD). 34% of parents in the sample had at least a bachelor's degree, and 79% of parents had at least a high school degree. Participants were randomly assigned to contexts of different group affiliations with the factor of age balanced: in-group perpetrator & out-group victim ( $N = 51$ ), out-group perpetrator & out-group victim ( $N = 48$ ), in-group perpetrator & in-group victim ( $N = 46$ ), in-group perpetrator & out-group victim ( $N = 52$ ). This study was approved by Cornell University's Institutional Review Board, protocol number: 2001009334.

## 2.2 Procedure

Children were first informed that there were two teams in one kindergarten, the “yellow team” and the “blue team”. The experimenter then told children that they were assigned to the “yellow team,” and showed them a yellow T-shirt and helped them to put it on. As a manipulation check question, participants were asked to identify their team member in a picture. Only children who passed the check question could proceed to the next stage of the experiment.

The experimenter used pictures to tell children a story, and participants responded to a series of questions. The gender of the characters was matched with the participant’s own gender.

The stories about moral transgression were as follows:

In this kindergarten, there are many kids. Some of them belong to the yellow team as you do, others belong to the blue group.

*Physical aggression story:* Look! This is Mingming, and this is Qingqing.

Mingming is in the *yellow* team, and Qingqing is in the *blue* team. Look what happened!

Mingming and Qingqing had some disagreement. Mingming kicked Qingqing to the ground! (As is an example of in-group perpetrator and out-group victim.)

After telling the story, ask children 2 questions:

*Seriousness of the transgression:* Do you think Mingming is right or wrong? How much do you think he/she is right/wrong? Like really really right/wrong, right/wrong, or just a little right/wrong? The Experimenter let participants choose from 3 circles, whose areas indicated the extent of rightness/wrongness.

*Deserved punishment:* Do you think Mingming should be punished? How much do you think Mingming should be punished? Like really really much, some, or just a little? The

Experimenter let participants choose from 3 circles, whose areas indicated the harshness of punishment.

Then the experimenter goes on to tell the story of the agent's punishment: "Look! Here's the teacher in Mingming's class, and she saw this. All kids have five cookies as treatment. Let's see what happened. She decided to take away all the five cookies from Mingming."

Following the punisher's story, the experimenter asked children the following two questions:

*Evaluation of the third-party punishment:* "Do you think the teacher is being right or wrong to take away all five cookies from Mingming?" Like really really right/wrong, right/wrong, or just a little right/wrong? The Experimenter let participants choose from 3 circles, whose areas indicated the extent of rightness/wrongness.

*Assignment of third-party punishment:* "How many cookies would you think you should take away from Mingming? None, one, two, three, four, or five?"

Considering preschool age children's limited verbal and imaginary ability, in the above questions we showed participants real cookies and asked them to assign cookies with their own hands.

### **2.3 Coding**

In the current study we used 4 questions to test children's judgement about the seriousness of the transgression, the deserved punishment, the evaluation of the third-party punishment, and their own assignment of third-party punishment after observing how another agent enforced punish.

In the first question measuring children's judgement about the seriousness of the transgression, each participant's answer was scored 3 ("really really wrong"), 2 ("wrong"), 1



(“just a little wrong”), 0 (“not sure”), -1 (“just a little right”), -2 (“right”), -3 (“really really right”).

In the second question measuring children’s judgement about the deserved punishment of the perpetrator, each participant’s answer was scored 3 (“deserve really really much punishment”), 2 (“deserve some punishment”), 1 (“deserve just a little punishment”), 0 (“not sure”), -1 (“a little don’t deserve punishment”), -2 (“don’t deserve punishment”), -3 (“really really don’t deserve punishment”).

In the third question measuring children’s evaluation of the punisher’s third-party punishment, each participant’s answer was scored 3 (“did really really right”), 2 (“did right”), 1 (“did a little right”), 0 (“not sure”), -1 (“did a little wrong”), -2 (“did wrong”), -3 (“did really really wrong”).

In the fourth question measuring children’s own assignment of third-party punishment, each participant’s answer was scored from 0-5 according to the number of cookies they chose to take away from the perpetrator (out of 5).

## **Results**

The primary goal of this study was to test whether children have group bias in judging other’s third-party punishment; The secondary goal was to explore the developmental trajectory of moral evaluations and its intersection with group bias in such moral evaluation; Aside from these, we also aimed to replicate in Chinese samples the previous findings about children’s in-group favoritism in their judgment of the seriousness of a transgression, judgment of deserved punishment as the third-party, and their own choice of third-party punishment.

**Table 1.** *Children's moral judgment by group affiliation (Standard Deviations in parentheses)*

	Out-group perpetrator & In-group victim	In-group perpetrator & Out-group victim	In-group perpetrator & In-group victim	Out-group perpetrator & Out-group victim
Seriousness of transgression	2.42(0.87)	2.29(0.86)	2.23(0.81)	2.50(0.77)
Deserved punishment	2.28(1.03)	1.66(1.67)	1.45(1.52)	2.00(1.32)
Evaluation of punishment	2.28(1.36)	1.56(2.02)	1.57(2.17)	2.28(1.46)
Children's own punishment	4.09(1.03)	3.84(1.29)	4.07(1.04)	4.33(0.88)

**Table 2.** *Children's moral judgment by age (Standard Deviations in parentheses)*

	3-year-old	4-year-old	5-year-old	6-year-old	7-year-old
Seriousness of transgression	1.96(0.98)	2.34(0.91)	2.31(0.85)	2.52(0.66)	2.74(0.54)
Deserved punishment	0.30(2.06)	1.33(2.04)	2.08(1.39)	2.14(0.77)	2.48(0.51)
Evaluation of punishment	0.78(2.45)	1.67(2.01)	2.10(1.62)	2.3(1.44)	1.91(1.73)
Children's own punishment	3.35(1.23)	4.19(0.98)	4.07(1.24)	3.85(1.47)	4.24(0.67)

*Did children have in-group favoritism in judging the seriousness of a transgression?*

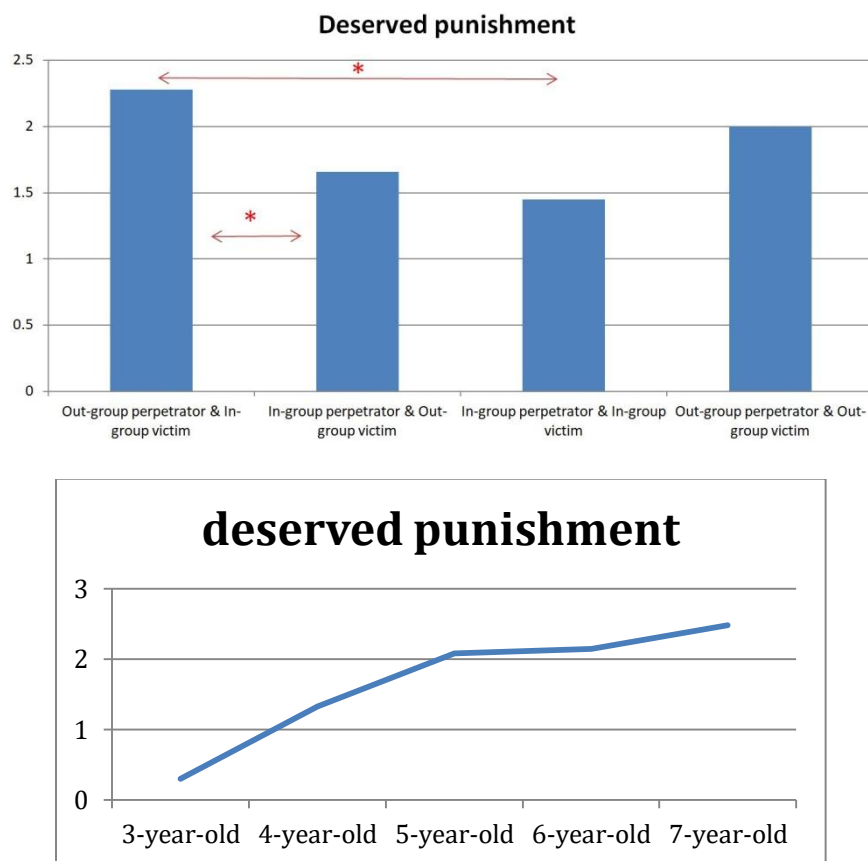
For judging the seriousness of a transgression, we didn't detect a significant main effect of group affiliation. The main effect of age was significant ( $F[4.192] = 3.12, p = .016, \eta^2 = .06$ ). Children at age 7 significantly judged transgressions as more serious ( $M = 2.74, SD = .54$ ) than children at age 3 ( $M = 1.96, SD = .98$ ) ( $p = .019$ ). The difference among other age groups is not significant. The intersections between age and group affiliation was not significant.



*Did children have in-group favoritism in judging the deserved punishment?*

For judging the deserved punishment of a transgression, as expected there was a significant main effect of group affiliation ( $F[3,159] = 6.47, p < .01$ ). Children evaluated a transgression conducted by an out-group perpetrator toward an in-group victim as more deserving of punishment ( $M = 2.28, SD = 1.03$ ) than the one conducted by an in-group perpetrator toward an in-group victim ( $M = 1.45, SD = 1.52, p < .01$ ) and the one conducted by an in-group perpetrator toward an out-group victim ( $M = 1.66, SD = 1.67, p < .01$ ). The main effect of age was significant ( $F[4,192] = 8.73, p < .001, \eta^2 = .15$ ). 5-year-old ( $M = 2.08, SD = 1.38$ ) ( $p = .007$ ), 6-year-old ( $M = 2.14, SD = .77$ ) ( $p = .003$ ) and 7-year-old ( $M = 2.48, SD = .51$ ) ( $p$

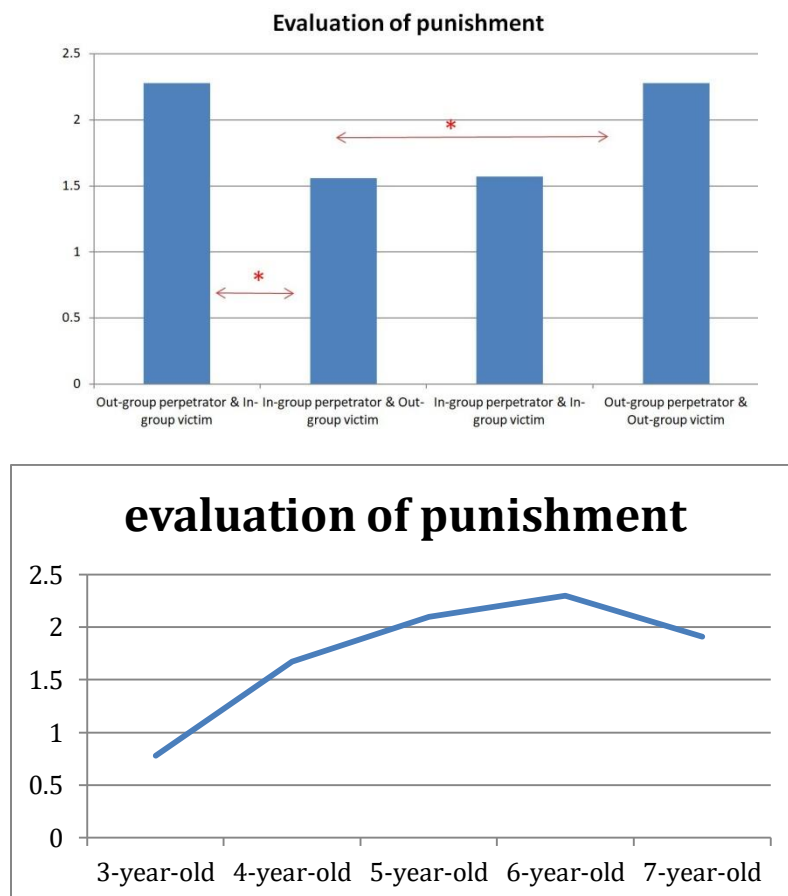
< .001) children significantly judged transgressions as more deserving of punishment than children at age 3 ( $M = .30$ ,  $SD = 2.06$ ). 7-year-old ( $M = 2.48$ ,  $SD = .51$ ) ( $p = .002$ ) children significantly judged transgressions as more deserving of punishment than children at age 3 ( $M = 1.33$ ,  $SD = 2.04$ ). The difference among other age groups is not significant. The intersection between age and group affiliations was not significant.



*Did children have in-group favoritism in evaluating other's third-party punishment?*

For children's evaluation of other's third-party punishment, as expected there was a significant main effect of group affiliation ( $F[3,159] = 13.41$ ,  $p < .01$ ). Children evaluated a third-party punishment more positively when it was enforced on an out-group perpetrator who hurt an out-group victim ( $M = 2.28$ ,  $SD = 1.46$ ) than an in-group perpetrator who hurt an out-group victim ( $M = 1.56$ ,  $SD = 2.02$ ,  $p < .05$ ). Children also evaluated a punishment more positively when it was enforced on an out-group perpetrator who hurt an in-group victim ( $M = 2.28$ ,  $SD = 1.36$ ) than an in-group perpetrator who hurt an out-group victim ( $M = 1.56$ ,  $SD = 2.02$ ,  $p < .01$ ).

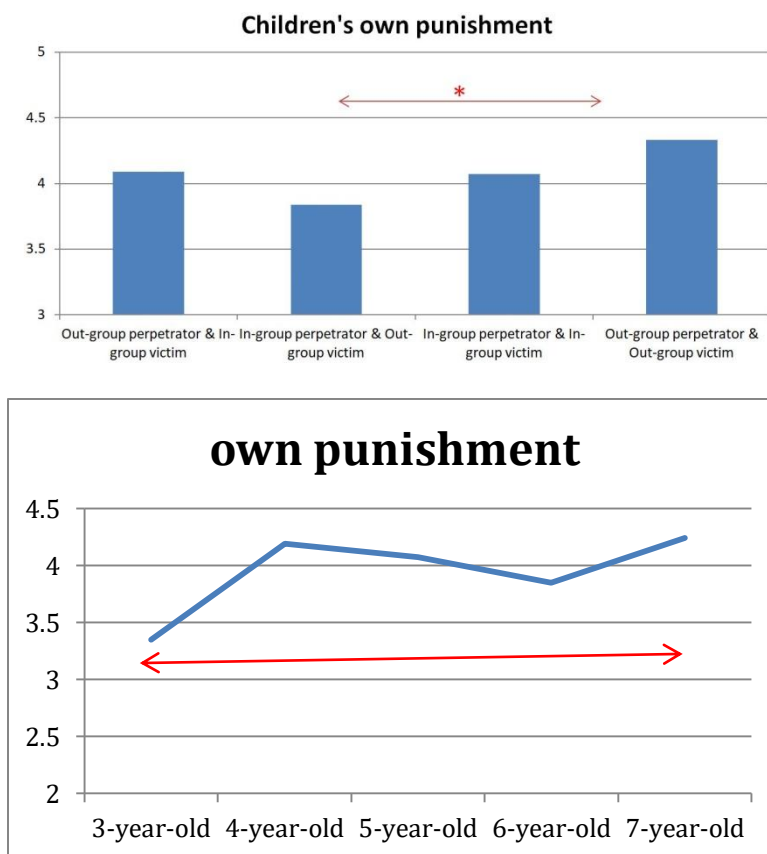
The main effect of age was significant ( $F[4,192] = 2.98, p = .02, \eta^2 = .06$ ). 4-year-old ( $M = 1.67, SD = 2.01$ ) ( $p = .05$ ), 5-year-old ( $M = 2.10, SD = 1.62$ ) ( $p = .005$ ), 6-year-old ( $M = 2.30, SD = 1.44$ ) ( $p = .002$ ) and 7-year-old ( $M = 1.91, SD = 1.73$ ) ( $p = .04$ ) children significantly evaluated punishment toward transgressions more positively than children at age 3 ( $M = .78, SD = 2.45$ ). The difference among other age groups is not significant. The intersection between age and group affiliations was not significant.



*Did children have in-group favoritism in enforcing third-party punishment?*

For children's choice of third-party punishment, as expected there was a significant main effect of group affiliation ( $F[3,159] = 4.36, p < .05$ ). Children as a third-party would choose to punish an out-group perpetrator who hurt an out-group victim more harshly ( $M = 4.33, SD = .88$ ) than an in-group perpetrator who hurt an out-group victim ( $M = 3.84, SD = 1.29, p < .05$ ). The main effect of age was significant ( $F[4,192] = 2.61, p = .037, \eta^2 = .052$ ). 7-year-

old ( $M = 14.24$ ,  $SD = .67$ ) ( $p = .04$ ) children significantly allocate more punishment toward transgressions than children at age 3 ( $M = 3.35$ ,  $SD = 1.23$ ). The difference among other age groups is not significant. The interaction were not significant.



## Discussion

In this study, children saw the same moral transgression happen between characters of different group affiliation, and observed a third-party punisher committed the same punishment. Of interest was whether children would judge the same punishment toward the same transgression differently when the perpetrator and victim's group membership varied. This study also aimed to shed some light on the broader question of children's "default morality" (Jordan,

McAuliffe, & Warneken, 2014), examining whether moral evaluation is biased from its emergence.

Our result showed that not only the punishment toward unfairness, but also the evaluation of the fairness of other's punishment, was already biased at a very young age. Aligns with results in many existing studies (Bigler, Jones, & Lobliner, 1997; Dunham et al., 2011; Hetherington, Hendrickson, & Koenig, 2014; Chapman et al., 2020), children in our study displayed a robust preference for in-group perpetrators in moral evaluation, judging transgressions conducted by in-group perpetrators as deserve more lenient punishment.

To be noted, our results showed that children's evaluation of other's third-party punishment were also biased. They had in-group favoritism for the in-group perpetrator and out-group bias toward the out-group perpetrator. This finding showed the similar bias pattern with previous findings on adult's group bias in third-party punishment (Graham et al., 1997; Sommers & Ellsworth, 2000) and children's third-party punishment (Chapman et al., 2020), which also found in-group favoritism induced by the perpetrator's group affiliation. However, we didn't detect in-group favoritism based on the victim's group affiliation, suggesting that children treated in-group victim and out-group victim equally in evaluating a third-party punishment. This result contrasted the pattern in Bernhard, Fischbacher and Fehr's study in adults (2006), who found in-group favoritism in third-party punishment was only induced by the victim's group affiliation, but not by the perpetrator's group affiliation. Our guess is that compared with adults, with limited cognitive capacity children may only focus on the direct recipient of the punishment (i.e., the perpetrator). Another possibility is that this disparity might be explained by the inherent difference between moral enforcement and moral evaluation: when directly conducting third-party punishment, people are more likely to be motivated by empathy toward the victim, and



thus make the group affiliation of the victim salient; but when evaluating a third-party punishment made by another person, we may focus on how much the perpetrator get suffered from the punishment, which may induce people to pay attention to the perpetrator's group affiliation. A third explanation is that the group bias that discriminates against the out-group perpetrators may be motivated by a "desert-based" punishment (Schiller, Baumgartner, & Knoch, 2014). This is supported by neural network findings that punishment toward out-group perpetrators and failure of rival group members can activate areas in the brain's reward circuit (Cikara, Botvinick, & Fiske, 2011; Baumgartner et al., 2012). Therefore, we infer that an affection of "sweet revenge" against the outgroup may inclicit higher evaluation on punishment toward the out-group perpetrators.

The finding that children judge the sanction on an out-group perpetrator more positively than the one on an in-group counterpart is not predicted by the group norm maintenance theory. The group maintenance theory describes a phenomena that group members punish in-group perpetrators more harshly than out-group or neutral perpetrators, so that the pro-cooperative norms in the group can be sustained (Shinada, Yamagishi, & Ohmura, 2004). Inconsistent with this theory, we only detected simple and straightforward group bias in our results. One explanation is that the group norm maintenance effect may be more salient in conventional norm transgressions than in moral norm transgressions (Jordan, McAuliffe, & Warneken, 2014). The moral norms are universally applicable, while conventional norms are more dependent on group context (Smetana, 1981; Turiel, 1983). By breaking the conventional norms in a group, perpetrators show disrespect to group regulations. Therefore people may punish in-group member's violation of conventional norms more harshly because it is more detrimental to group norm maintenance than the violation of moral norms. Another explanation to this inconsistency

is that with limited reasoning capacity, children at age 3-7 haven't yet obtained this complicated ability, especially considering that past studies supporting the group maintenance theory is mostly conducted on adults (Shinada, Yamagishi, & Ohmura, 2004; Bernhard, Fischbacher, & Fehr, 2006). In a developmental research conducted by Jordan, McAuliffe and Warneken (2014), similarly they only found straightforward group bias in 6- and 8-year-old children's third-party punishment. Therefore, we infer that the group maintenance theory in moral punishment may not be valid from the emergency of altruistic punishment and its evaluation.

One unsolved question in this study is that although we detected a discriminatory treatment between in-group and out-group perpetrator, it is not clear whether this bias was driven by an preference toward in-group perpetrators (in-group favoritism) or an aversion toward out-group perpetrators (out-group discrimination). To answer this question, future study should add a baseline group of unaffiliated perpetrators, and compare it with out-group and in-group perpetrators.

From a developmental perspective, our study found that through age 3 to 7, children judged the same transgression as more serious and deserve more punishment. This result was in line with previous study on children's moral behavior (Jordan, McAuliffe, & Warneken, 2014; Gummerum & Chu, 2014; Vaish, Herrmann, Markmann & Tomasello, 2016), showing children's increased sense of morality and desire for justice. Also we found that with age, older children evaluated a third-party punishment as more positive. This result showed a similar developmental pattern with existing studies on children's moral evaluation (Lee & Warneken, 2020). To be noted, we found group bias in evaluating other's third-party punishment started off at the youngest age group of our study. This result might suggest that starting from 3-year-old children already favor in-group members in evaluating other's third-party punishment. This was

ahead of the turning point at 6-year-old where children showed in-group favoritism in enforcing third-party punishment (Jordan, McAuliffe, & Warneken, 2014). But this emergence age was in accordance with previous studies about moral evaluations (Chapman et al., 2020), which found that children at age 3 already favor in-group transgressors in their evaluation of deserved punishment and seriousness of transgression. Therefore, we may infer that children's group bias in moral evaluation develops earlier than moral behaviors.

However, we didn't detect any intersection between age and group affiliation. Children were neither more biased nor more impartial over development. With age children gain group loyalty, the consciousness to monitor self image, and a moral desire for impartiality (Jordan, McAuliffe, & Warneken, 2014). These three factors could have an interacting effect that children have to integrate over development (Choi & Bowles, 2007; Abrams, Rutland, Pelletier, & Ferrell, 2009). Some studies even detected declined explicit in-group favoritism and out-group prejudice around mid-childhood (Dunham, Baron, & Banaji, 2008; Raabe & Beelmann, 2011). For example, an intersection effect between age and group affiliation was found in Jordan, McAuliffe and Warneken's study (2014). Although third-party punishment increased between ages 6 and 8, 8-year-old children's group bias is attenuated. Therefore, the insignificant intersection could be because the increased group loyalty was offset by the increased desire for impartiality and to maintain an impartial self-image. So to answer the question of how the group bias and desire for impartiality together affect children's evaluation of third-party-punishment, future studies may include more older children in their sample. To be noted, this study was conducted in Chinese sample. And possibilities are that children from Chinese culture both value group loyalty and self-representation more than children from Western cultures. It will be interesting to investigate how age and in-group favoritism intersect in the U.S. sample. Future

cross-cultural comparisons may give us a whole picture of the balance among aversion to injustice, self-image maintenance and group loyalty.

This study filled the following two theory gaps: First, to our knowledge this study was the first to examine the role of group membership on children's evaluation of third-party punishment. Although a large body of studies have looked at both adult's group bias in third-party punishment (Graham et al., 1997; Sommers and Ellsworth, 2000; Bernhard, Fischbacher & Fehr, 2006; Schiller, Baumgartner, & Knoch, 2014) and adult's evaluation on third-part punisher (Barclay, 2006; Nelissen, 2008; Horita, 2010), very few studies have talked about these two topics from the developmental angle, letting alone examining their relationship. Evolutionarily speaking, group activity brings both parochiality and altruism, and both of them are crucial to sustain group activity. By linking parochiality with altruistic moral enforcement, it can help us to understand the mechanism that maintains stable cooperation (Choi & Bowles, 2007; Schiller, Baumgartner, & Knoch, 2014). Our result had some implications for this integration: Although children showed robust group bias in moral behavior and moral judgment, they still chose to enforce morality toward an in-group perpetrator. We may infer from the result that although group bias exists, children weigh the moral rule more than group loyalty. Second, this study was one of the first to examine children's group bias in moral enforcement in a non-WEIRD culture. The exclusive focus on western culture in developmental work omitted one important culture factor that could influence in-group favoritism: the sense of belonging and the value of group loyalty. As a collectivism culture, Asian cultures value the sacrifice of self to a group, encouraging conformity of behaviors and opinions, regarding children's ability to fit well in a social group as an indicator of intelligence (Azuma, & Kashiwagi, 1987). While the U.S. as an individualism culture, praises self uniqueness and critical thinking (Corriveau & Harris, 2010;

Corriveau, Kim, & Song, 2013). By comparing U.S. samples with the current Chinese sample, we may expect to see differences in the developmental trajectory.

Our study builded on the widely adapted minimal group paradigm. In discussing the key motivations behind group bias, it was proposed that the minimal group effect in children may derive from a general positive affection towards in - group members (Bigler & Liben, 2006; Dunham et al., 2011). However, our study showed that despite evaluating in-group members quite negatively (e.g., judging them as doing wrong and deserve punishment), children still robustly favor in-group members both in moral judgment and moral behavior. This result echoed with Hetherington, Hendrickson, & Koenig's study (2014), which found an in-group learning preference even when children evaluate in-group members negatively. Therefore, the in-group favoritism may not be driven solely by the positive feelings. Another existing hypothesis is that the expectations of reciprocity from in-group members may drive the minimal group effect (Yamagishi & Kiyonari, 2000; Hetherington, Hendrickson, & Koenig, 2014). However, our study examined the topic of punishment, which is inherently against reciprocity (offer benefit to each other and expect to receive more in the future). However, it has been found that perpetrators would expect third-party punishers to be more lenient if they share the same group affiliation (Bernhard, Fischbacher, & Fehr, 2006). So future study can clarify if the in-group favoritism in punishment is motivated by a more complicated logic of reciprocity: children favoring in - group perpetrators because of a belief that in-group members are likely to be more lenient when positions switch in the future.

The current study examined children's evaluations of the punisher by directly asking their attitude with verbal questions, which only reflected their explicit attitude. As discussed in the previous part, children gain the concern to monitor their self image over development: they

become less willing to appear parochial. Therefore as a supplement of the development trajectory of explicit attitude found in the current study, future studies can look at children's implicit evaluations. For example, we can adapt the paradigm that examines the preference for a person by asking children about their willingness to trust, to share treatment, and to endorse treatment to the punisher (Hetherington, Hendrickson, & Koenig, 2014).

Another future direction that integrates loyalty and impartiality together is: will children still favor in-group when morality is in conflict with parochiality? In the current study, when the in-group perpetrator hurted an out-group victim (group loyalty and impartiality conflicts), children robustly judged the in-group perpetrator as wrong, deserving punishment, and gave positive evaluation to those who punished. This result suggested that children's moral behavior might prioritized group loyalty across ages 3 to 7, which is consistent with previous study which found antisocial in - group members could attenuate 4- to 5-year-old children's in-group favoritism (Hetherington, Hendrickson, & Koenig, 2014). Together, these evidence implicated that when morality and group membership are at conflict, preschoolers choose morality to guide their moral decisions.

In conclusion, we found that children have group bias in their evaluation of third-party punishment. This study has added to the evidence that in-group favoritism emerges early in ontology. And this study demonstrated that the tendency to favor in-group members is so strong that it exists in moral enforcement that is motivated by the aversion to injustice and unfairness (which inherently goes against bias).

## Reference

- Abell, J. . (2008). Children"s knowledge, beliefs and feelings about nations and national groups: essays in developmental psychology by martyn barrett. *Developmental Psychology*, 14(1), 193-194.
- Abrams, D., Rutland, A., Pelletier, J., Ferrell, J. M. (2009). Children’s group nous: Understanding and applying peer exclusion within and between groups. *Child Dev*, 80(1): 224–243
- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27(5), 325–344. <https://doi.org/10.1016/j.evolhumbehav.2006.01.003>
- Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. *Nature* 442(7105):912–915
- Bigler, R.S., Jones, L.C., & Lobliner, D.B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, 68, 530– 543.  
doi:10.1111/j.1467-8624.1997.tb01956.x
- Bigler, R.S., & Liben, L. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behavior*, 34, 39– 89.  
doi:10.1111/j.1467-8721.2007.00496.x
- Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, 60(5), 930–940.  
<https://doi.org/10.1016/j.neuron.2008.10.016>
- Chapman, M. S., May, K. E., Scofield, J., DeCoster, J., & Bui, C. (2020). Does group membership affect children's judgments of social transgressions?. *Journal of experimental child psychology*, 189, 104695. <https://doi.org/10.1016/j.jecp.2019.104695>

- Choi, J. K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science (New York, N.Y.)*, 318(5850), 636–640. <https://doi.org/10.1126/science.1144237>
- Dunham, Y., Baron A. S., Banaji, M. R., (2008) The development of implicit intergroup cognition. *Trends Cogn Sci* 12(7):248–253.
- Dunham, Y., Baron, A.S., & Carey, S. (2011). Consequences of ‘minimal’ group affiliations in children. *Child Development*, 82, 793– 811. doi:10.1111/j.1467-8624.2011.01577.x
- Dunham, Y., (2018). Mere membership. *Trends in cognitive sciences*.
- Furman, W., & Masters, J. C. (1980). Affective consequences of social reinforcement, punishment, and neutral behavior. *Developmental Psychology*, 16(2), 100–104. <https://doi.org/10.1037/0012-1649.16.2.100>
- Garmon, L., Basinger, K., Gregg, V., & Gibbs, J. (1996). Gender Differences in Stage and Expression of Moral Judgment. *Merrill-Palmer Quarterly*, 42(3), 418-437. Retrieved June 15, 2020, from [www.jstor.org/stable/23089870](http://www.jstor.org/stable/23089870)
- Götte L, Huffman D, Meier S (2006) The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am Econ Rev* 96(2):212–216.
- Gump, L. S., Baker, R. C., & Roll, S. (2000). Cultural and Gender Differences in Moral Judgment: A Study of Mexican Americans and Anglo-Americans. *Hispanic Journal of Behavioral Sciences*, 22(1), 78–93. <https://doi.org/10.1177/0739986300221004>
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26(1), 30–39. <https://doi.org/10.1016/j.cogdev.2010.09.001>



- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 19931–19936. <https://doi.org/10.1073/pnas.1110306108>
- Heiphetz, L., Spelke, E. S., & Banaji, M. R. (2013). Patterns of implicit and explicit attitudes in children and adults: Tests in the domain of religion. *Journal of Experimental Psychology: General*, 142(3), 864–879. <https://doi.org/10.1037/a0029714>
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science (New York, N.Y.)*, 327(5972), 1480–1484. <https://doi.org/10.1126/science.1182238>
- Hetherington, C., Hendrickson, C., & Koenig, M. (2014). Reducing an in-group bias in preschool children: the impact of moral behavior. *Developmental science*, 17(6), 1042–1049. <https://doi.org/10.1111/desc.12192>
- Horita, Y. . (2010). Punishers may be chosen as providers but not as recipients. *letters on evolutionary behavioral science*.
- 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 of the United States of America*, 111(35), 12710–12715. <https://doi.org/10.1073/pnas.1402280111>
- Kanakogi, Y., Inoue, Y., Matsuda, G. *et al.* Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat Hum Behav* 1, 0037 (2017). <https://doi.org/10.1038/s41562-016-0037>

- Kinzler, K. D. , Dupoux, E. , & Spelke, E. S. . (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), p.12577-12580.
- Lee, Y., & Warneken, F. (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. <https://doi.org/10.31234/osf.io/x8e7w>
- 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>
- Mysterud, I. Unto others: The evolution and psychology of unselfish behavior. *Popul Environ* 21, 581–588 (2000). <https://doi.org/10.1007/BF02436773>
- Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, 29(4), 242–248. <https://doi.org/10.1016/j.evolhumbehav.2008.01.001>
- Raabe, T., Beelmann, A., (2011) Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Dev* 82(6):1715–1737.
- Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in ecology & evolution*, 27(5), 288–295. <https://doi.org/10.1016/j.tree.2011.12.004>
- 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.
- Smith, C. E., & Warneken, F. (2016). Children's reasoning about distributive and retributive justice across development. *Developmental psychology*, 52(4), 613–628. <https://doi.org/10.1037/a0040069>

- Sripada, C. S.. (2005). Punishment and the Strategic Structure of Moral Systems. *Biol Philos* 20, 767–789 . <https://doi.org/10.1007/s10539-004-5155-2>
- Taijfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223 5, 96-102.
- Vaish, A., Herrmann, E., Markmann, C., & Tomasello, M. (2016). Preschoolers value those who sanction non-cooperators. *Cognition*, 153, 43–51.  
<https://doi.org/10.1016/j.cognition.2016.04.011>
- Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *The British journal of developmental psychology*, 29(Pt 1), 124–130. <https://doi.org/10.1348/026151010X532888>
- Yamagishi, T., & Kiyonari, T. (2000). The Group as the Container of Generalized Reciprocity. *Social Psychology Quarterly*, 63(2), 116-132. Retrieved July 23, 2020, from [www.jstor.org/stable/2695887](http://www.jstor.org/stable/2695887)