Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 3, 608–620
https://doi.org/10.1037/xge0001515

Children's Third-Party Punishment Does Not Change Depending on the Prospect of Future Interaction

Young-eun Lee[1, 2], Susan He[1], and Felix Warneken[1]
[1] Department of Psychology, University of Michigan
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Children pay a cost to punish third parties for unfairness. However, theoretical debates highlight that such behaviors could reflect a strategic attempt to manipulate others in future interactions. The personal deterrence hypothesis claims that punishment is motivated to deter future unfairness toward punishers. Here we tested this hypothesis with a total of $n = 248$ five- to 10-year-olds. In two experiments, participants witnessed that a divider shared resources either fairly or selfishly with a third party. Participants learned that the same divider (same divider condition) or a new divider (different divider condition) would subsequently decide how to share resources with the participant. If children's punishment is motivated by personal deterrence, they should punish unfairness more often in the same divider condition (vs. different divider). Conversely, if children fear retaliation from dividers, they should punish dividers less often in the same divider condition (vs. different divider). Children intervened by taking resources away from the divider (Experiment 1) or by sending a disapproving or an approving verbal message (Experiment 2). Children were more likely to punish unfair than fair allocations through material punishment and disapproving messages, while being more likely to reward fair than unfair allocations by sending approving messages. However, children did so at the same level regardless of their future divider's identity. We discuss how these results speak to a children's emerging concern with fairness and how it challenges the notion that children punish for self-oriented reasons as suggested by the personal deterrence hypothesis.

Public Significance Statement
Adults often punish transgressions for self-oriented reasons even when they are an unaffected third party. For example, adults punish a transgressor to get a better bargain for themselves in the future by letting the transgressor know that they would not tolerate such selfish acts if those acts were directed at punishers themselves. This work, however, shows that 5- to 10-year-old children do not use punishment for this self-oriented reason. This finding suggests that at least in middle childhood, punishment is unlikely to be driven by the self-oriented motive.

Keywords: deterrence, development, fairness, punishment, reward

Supplemental materials: https://doi.org/10.1037/xge0001515.supp

To maintain cooperation within a group, it is important to intervene when individuals violate norms by acting in a selfish manner. One such intervention is costly third-party punishment, that is, punishment of selfish behaviors at a personal cost even if the punisher is not directly affected by the current selfish acts (Fehr & Fischbacher, 2004; Henrich et al., 2006; Nelissen & Zeelenberg, 2009; Yamagishi et al., 2017). This so-called costly third-party punishment is a striking phenomenon because individuals pay a personal cost to intervene against norm violations that occurred among third parties and the punishing individual was not directly affected by the transgression. This behavior has been demonstrated in anonymous interactions among strangers and one-shot interactions with no obvious short- or long-term benefit to punishers

This article was published Online First December 7, 2023.

This work was funded by a National Science Foundation CAREER Grant (1760238) awarded to Felix Warneken and a Rackham Research Grant awarded to Young-eun Lee. The authors thank the research assistants and families for their help with data collection. All data, analysis code, and research materials are available at https://osf.io/q3nhc/. The authors preregistered all experiments (see https://aspredicted.org/ep9zg.pdf for Experiment 1 and https://aspredicted.org/m98pk.pdf for Experiment 2). Data from Experiment 1 were presented as a poster at the meeting of the Society for Research in Child Development in 2019.

Young-eun Lee served as lead for conceptualization, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, writing–original draft, and writing–review and editing. Susan He served as lead for conceptualization, data curation, investigation, methodology, project administration, and writing–review and editing. Felix Warneken served as lead for conceptualization, funding acquisition, methodology, supervision, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Young-eun Lee, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, United States. Email: yelee@mit.edu

themselves (e.g., Fehr & Fischbacher, 2004; Henrich et al., 2006; Nelissen & Zeelenberg, 2009). Therefore, this phenomenon is difficult to explain in terms of a selfish agent model with a motive to maximize their own payoff who would not have enacted third-party punishment at all in these contexts. For these reasons, the motivations underlying third-party punishment remain debated.

Potential Motivations Underlying Third-Party Punishment

There are several accounts that could potentially explain proximate mechanisms of third-party punishment. Here we focus on two specific hypotheses relevant to the current research. Specifically, some theorists explain motives underlying third-party punishment in terms of the prospect of future interactions. That is, even though the benefit to the punisher is not immediately apparent, third-party punishment is driven by a motive to gain or protect some personal benefit in potential interactions in the future. For example, the personal deterrence hypothesis argues that people punish perpetrators to signal that they themselves would not accept being treated unfavorably, and thus try to get a better bargain for themselves in the future (Petersen et al., 2010). In line with this claim, studies have shown that adults are more likely to punish a person for being selfish toward another individual when they infer that this person would treat them poorly in a future interaction than when this person would treat a third-party individual poorly (Delton & Krasnow, 2017; Krasnow et al., 2016). Thus, the more adults thought they could fall victim to a selfish individual, the more they punished the individual to preemptively change his or her behavior. By contrast, their prediction of how the perpetrator would treat a third party was not correlated with their willingness to punish (Krasnow et al., 2016). These findings suggest that at least in adults, punishers might punish to prevent personal mistreatment by considering how a perpetrator might behave toward them in the future. Therefore, the personal deterrence hypothesis would predict that people would show increased third-party punishment when they expect to encounter the same perpetrator again than when they do not.

Although it is possible that people punish to prevent personal unfair treatment, other studies suggest a different possibility. In fact, people might refrain from punishing selfish others to protect their personal benefits. For instance, adults who receive punishment often retaliate against their punishers (Denant-Boemont et al., 2007; Fehl et al., 2012; Zheng & Nie, 2013). Also, they are less willing to punish transgressors when transgressors could retaliate (Balafoutas et al., 2014; Nikiforakis, 2008). Therefore, it is possible that people might be more hesitant to punish when they expect negative reciprocity from transgressors who receive punishment. We call this the "appeasement hypothesis." While the appeasement hypothesis predicts that people are less likely to punish when they have to face the perpetrator in the future (vs. when they do not have to), it is also built upon the idea that third-party punishment decisions change depending on the prospect of future interactions as the personal deterrence hypothesis. The present work examined if/how children's third-party punishment changes based on anticipation of future interactions.

Third-Party Punishment in Children and Their Underlying Motivations

Developmental research can contribute to this ongoing debate by looking at the ontogenetic emergence of third-party punishment relative to other cognitive and social capacities. For example, if third-party punishment aims at increasing or protecting one's prospective personal benefit, we would not expect third-party punishment to emerge before children acquire prospective thinking skills. Then what is the developmental trajectory of children's third-party punishment? Prior research has shown that starting at around 3 years of age, children protest against norm violations such as not playing a game by the rules and show third-party punishment for destroying other people's belongings or for inflicting physical harm on others (e.g., Kenward & Östh, 2012, 2015; Rakoczy et al., 2008; Rossano et al., 2011; Vaish et al., 2011; Yudkin et al., 2020).

However, it is not until age 6 that U.S. children punish unequal over equal sharing of resources systematically (Arini et al., 2021; Gummerum & Chu, 2014; House et al., 2020; Lee & Warneken, 2022a; McAuliffe et al., 2015). One study (Smith et al., 2013) has shown that children as young as 3 years of age know that resources should be distributed equally, endorsing the fairness norm. However, it is not until ages 7–8 that children actually share resources equally, showing that their behavior becomes consistent with their knowledge around this age. Furthermore, 5- to 6-year-olds often incur a cost to get more resources than their peers (Sheskin et al., 2014), whereas children ages 7 and older bear a cost to avoid getting more resources than their peers (Blake & McAuliffe, 2011; Shaw & Olson, 2012). Overall, these findings suggest that at least in the United States, children adhere to fairness norms in a principled manner over development.

For the purposes of the current research, third-party punishment of unfairness is of particular importance because it represents children's commitment to fairness norms, and therefore it is considered as one of the important milestones in fairness development (McAuliffe et al., 2017). In addition, adult literature has shown third-party punishment driven by personal deterrence in unfair resource allocations (e.g., Delton & Krasnow, 2017; Krasnow et al., 2016), which would allow us to compare developmental changes in motives between children and adults more precisely by using the same type of norm violation.

To the best of our knowledge, no prior study has directly tested whether children's third-party punishment is motivated by personal deterrence or appeasement when anticipating a future interaction with the same transgressor they previously encountered. Children's third-party punishment has been claimed to be an index of their fairness concerns (McAuliffe et al., 2017). However, this might not have been necessarily true if children had expected some sort of future interactions with other players either explicitly or implicitly even though experimenters did not provide the relevant information to children (e.g., Delton et al., 2011; Krasnow et al., 2016). Therefore, prior research demonstrating third-party punishment against unfair sharing cannot address clearly whether children's punishment reflects their genuine fairness concern or whether it reflects more self-oriented motives (i.e., strategic attempt to manipulate peers for potential future interactions). Hence, the current work manipulated directly whether children would interact with the same player in the future or not. Addressing this question is important because this would allow us to examine the degree to which children's fairness concerns and the prospect of future interactions, respectively, account for their third-party punishment.

Moreover, the present research could inform us of the motivations and cognitive mechanisms underlying punishment. To be specific, if third-party punishment is motivated by personal deterrence in humans, it is likely that children would show punishment around the same age they acquire cognitive skills necessary for deterrence-driven

punishment. For example, from age 6, children start to punish unfair sharing more often than fair sharing (McAuliffe et al., 2015). It is around the same age that children show the ability to plan for the future (Atance & Meltzoff, 2005), share resources with another person strategically depending on whether the person could reciprocate them in the future (Rosati et al., 2019), and display prosocial or antisocial behaviors selectively depending on whether they are being watched by others (Engelmann et al., 2012). Together, the findings suggest that the development of third-party punishment and the emergence of strategic future-thinking skills and reputation management that might support deterrence-based punishment occur around the same age in early childhood.

Recent research has shown some indirect evidence that prospective utility might underlie children's motive for third-party punishment. In the context of ownership violation (i.e., ripping apart another person's artwork), 5- to 7-year-old children were motivated by both retribution and a motive to teach a lesson. Specifically, children punished a transgressor who destroyed another person's belongings not only when they could not communicate the reason for punishment to the transgressor but also when they could communicate it (Marshall et al., 2021). Also, there were similar findings with 9- to 12-year-olds in an unfair distribution context (Twardawski et al., 2020). In sum, the prior studies suggest that children's third-party punishment is guided by both deservingness and a communicative motive.

While these previous studies provided important insight into the motivational basis in children's punishment, it is still possible that children's communicative motive was to deter personal mistreatment directed toward "punishers themselves" specifically rather than to deter the transgressor's mistreatment of "general others" broadly. This is important to distinguish because the identical act could have a completely different underlying motivation depending on what deterrence aims for. Specifically, the act of bearing a personal cost to intervene against a third-party transgression is more likely to reflect one's altruistic, justice-based motive if it aims to deter unfair treatment toward other people in general. In contrast, if the act aims to deter unfairness toward punishers themselves, then it is more likely to indicate the person's self-oriented motivation (i.e., getting a better bargain for themselves in future interactions). The present work examined whether children use punishment to deter personal mistreatment directed toward themselves.

Current Research

The current research tested whether children's third-party punishment is influenced by the prospect of encountering a selfish person in the future. To assess these possibilities, we presented children with a third-party punishment game, in which they observed one child (hereafter divider) share six valuable resources either equally (3:3) or unfairly (6:0) with another child (hereafter recipient). Our child participants acted as an unaffected third party who could decide whether to punish the divider by taking resources away from the person. Our main experimental manipulation was the subsequent future interactions children would be exposed to: Children knew that upon the completion of the third-party punishment game, they would play a different sharing game where they had the chance to receive resources from another individual who could decide whether to share with the child. We manipulated whether the game partner would be the same, selfish divider from the third-party punishment game (same divider condition) or a new, different divider (different divider condition). In this way, we were able to assess whether anticipating an interaction with the selfish divider in the future would influence how children intervene against unfair sharing with another person in the present.

We tried to adjudicate between two hypotheses: First, the personal deterrence hypothesis predicts that children would punish unfair allocations more often during the third-party punishment game when they have to encounter the same unfair person later, resulting in more punishment in the same divider over the different divider condition. This possibility is plausible because, at least in some contexts, children use third-party punishment to communicate to a transgressor (Marshall et al., 2021; Twardawski et al., 2020), suggesting that children are motivated to deter future transgressions although it was unclear whether their punishment was to deter transgressions directed to punishers themselves specifically or those directed to other people more broadly.

Second, the appeasement hypothesis predicts that children would punish unfairness less often during the third-party punishment game in the same divider condition than in the different divider condition. Such a finding would suggest that children's third-party punishment is likely to be affected by potential negative reciprocation from the divider. That is, in the same divider condition, children punish the divider less often to be treated nicer by the same divider later. Prior work supported this possibility. For instance, 4- to 6-year-old children are more likely to share resources with a partner who could reciprocate in the future than those who could not (Rosati et al., 2019; Sebastián-Enesco & Warneken, 2015), showing that children invest their resources in a strategic manner depending on the possibility that their partner would reciprocate. Additionally, adults are less likely to punish when there is a threat of retaliation (Balafoutas et al., 2014; Nikiforakis, 2008).

We tested these hypotheses with children between the ages of 5 and 9. We chose this age range because prior research has shown that children around this age begin to engage in third-party punishment in the fairness violation context (Arini et al., 2021; Gummerum & Chu, 2014; House et al., 2020; Lee & Warneken, 2022a; McAuliffe et al., 2015). Furthermore, from age 5, children are able to think about and plan for the future (Atance & Meltzoff, 2005), which is a critical cognitive skill required for the deterrence-based motive.

Experiment 1

Method

Participants

Our final sample included $n = 120$ five- to 9-year-old children ($M = 89.2$ months, range = 61–120 months; $n = 60$ per condition; $n = 24$ in each age group; 60 male, 60 female). Children were tested at a museum or public parks in the Midwest of the United States between August 2018 and June 2019. Demographic information such as race, education, and income was not obtained. Thirteen additional children were excluded because of failure to correctly answer at least one of the comprehension check questions (10), parental interference (1), being unable to understand English (1), or participation in a similar study before (1).

We determined our sample size based on the effect size ($d = 0.53$) observed from a pilot study ($\alpha = .05$, power = .80; see our preregistration at https://aspredicted.org/ep9zg.pdf). A post hoc sensitivity

power analysis also confirmed that our sample size was large enough to detect a medium-sized effect ($d = 0.52$) of condition ($\alpha = .05$, power = .80).

When preregistering, we determined our sample size based on the idea to run two between-subject conditions: same divider and different divider condition. However, due to Young-eun Lee's failure to catch a mistake in the preregistration document, the document indicated that we planned to run three conditions, which was not our actual plan. The additional condition we preregistered by mistake was the no-future condition (baseline) in which children will not be informed of a future sharing game in advance. In reality, we never assigned children to the no-future condition. Also, because we found a nonsignificant condition effect, it was not necessary to run the no-future condition. Not running the no-future condition did not influence the way we interpret our results.

Experimental Design and Procedure

After parents gave written consent, children sat at a table with the study apparatus while the parent was watching passively from a few steps away. A female experimenter introduced the computer game referred to as the "coin game" and explained that players could collect virtual coins to later exchange for prizes. During a prize introduction, children learned that the more coins they have during the coin game, the more and the better prizes they would be able to choose afterward. A vast majority (98%) of the children understood the exchange value of coins. The experimenter provided corrective feedback to a minority of children who answered incorrectly.

In the practice phase, the experimenter introduced the two other players in the game by stating that they were children of the same age and gender at another museum (or another park for those who were tested at a park), who were currently connected online. The experimenter introduced children to a third-party punishment game, in which they can decide whether to punish as an observer (see Figure 1). Specifically, children learned that the divider could divide coins in one of two ways: (a) three for the self and three for the recipient (3:3) or (b) six for the self and 0 for the recipient (6:0). The recipient was a passive player who had to accept the divider's allocation. After introducing the roles, children saw the allocations made by the divider enacted on the screen and practiced their role as a third-party punisher.

Children learned that after the divider makes an allocation to the recipient, their job was to press one of the two buttons. When they pushed the green button (no punishment), the six coins went into each player's basket just the way the divider allocated the coins (e.g., when the divider split it up 3:3, each player's basket received three coins), and the child's own coin went back into their own basket. That is, the green button incurred no cost to the child. In contrast, when children pushed the red button (punishment), a vacuum appeared at the top of the screen and the six coins were sucked up and disappeared into the vacuum. To enact the punishment, however, children had to pay one of their coins into the vacuum first. Therefore, pressing the red button served as a costly third-party punishment.

There were four practice trials in total. Children practiced four possible outcomes of each button (no punishment vs. punishment) in each allocation type (fair vs. unfair). The experimenter asked comprehension check questions about the consequence of each button and whether each button required the payment of the participant's coin or not. When the child answered incorrectly, the experimenter provided corrective feedback and asked the question(s) again in the next practice trial. All participants included in the data analysis answered the comprehension check questions correctly by the end of the practice phase.

After the four practice trials, to make children believe that the other players were real, the experimenter pretended to make a phone call to the other players on the speakerphone and checked if they were ready to play the game. In reality, another experimenter answered the phone call, and there were no other players. Children received 20 coins as their initial endowment (coins that dropped into their basket on the screen).

In the subsequent manipulation phase, children in both conditions were told that, after playing the third-party punishment game, they will play a sharing game in which a divider allocates six coins between themself and the child participant. Critically, children were assigned to one of two conditions (between-subjects). In the same divider condition, the experimenter told children that the same divider from the third-party punishment game will decide how to share coins with the child in the subsequent sharing game. In the different divider condition, the experimenter told children that a new divider will decide how to share coins with the child in the sharing game. In the different divider condition, the experimenter did not provide children with information about whether the divider from the third-party punishment would play a sharing game with another child or not. This was to reduce complexity in the experimental design and to avoid potential confusion in children.

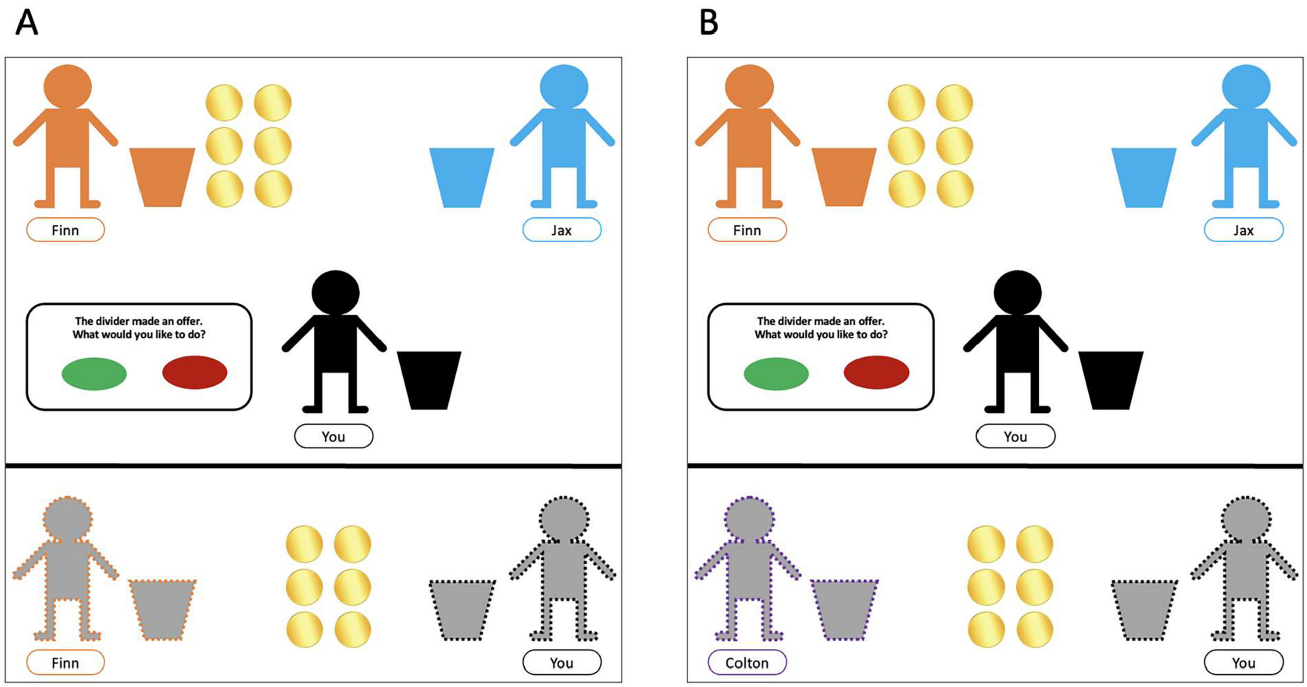
At no point in the study did children hear the word "third-party punishment game" (framed as a "three-person game" to children) or "sharing game" (framed as a "two-person game" to children). Before the third-party punishment game (test phase) began, most children (88%) correctly answered to the questions asking them to identify the divider and the recipient in the subsequent sharing game (e.g., "Who is going to be the divider in the two-person game?"). The experimenter provided corrective feedback to a minority of children who answered incorrectly. In both conditions, to visualize future interactions with the same or different divider, the computer game display indicated the future divider of the sharing game at the bottom of the screen. The display of players in the future sharing game was available throughout the test phase to remind children of their divider in the following sharing game (see Figure 1).

In both conditions, the experimenter told children that both the divider and the recipient in the third-party punishment game could see the child participant's decision to press green or red button. Children in the different divider condition were additionally told that the new divider could not see their decision in the third-party punishment game. The experimenter confirmed that most children (90%) identified players who could see or could not see their punishment decisions correctly. The experimenter provided corrective feedback to a minority of children who answered incorrectly before the test phase began.

During the following test phase, children in both conditions played the identical third-party punishment game, in which they could decide whether to punish an allocation by pressing either green (no punishment) or red button (punishment). Our dependent measure was children's rate of pressing the red button during the test phase. Children were presented with eight test trials in total. The divider made an unfair allocation (6:0) in six trials, while he or she made a fair allocation (3:3) in two trials. This manipulation was intended to make children perceive the divider as a selfish

Figure 1

Computer Screen Displays During the Test Phase in the Same Divider and Different Divider Condition



Note. (A) The screen display in the same divider condition with a green (left) and a red button (right). In this example, Finn kept six coins for the self and gave 0 coins to Jax who was the recipient. Children were told that Finn will be the divider in the subsequent sharing game. (B) The screen display in the different divider condition with a green (left) and a red button (right). In this example, children were told that Colton, a new player, will be the divider in the subsequent sharing game. See the online article for the color version of this figure.

individual, and thus induce the belief in children that they might be treated unfairly by the divider later. Across eight test trials, children saw fair trials twice: (a) in the first or second trial and (b) in the fifth or sixth trial. This order was counterbalanced across participants. All players were gender-matched to the participant.

At the end of the third-party punishment game, a memory check confirmed that most children (83%) correctly recalled that the divider kept coins for himself/herself most of the time. Furthermore, the experimenter asked children to predict whether the divider in the sharing game would share three coins or zero coins with the child themselves. We found that most children (77% in each condition) expected to receive three out of six coins from a divider (i.e., fair sharing) rather than to receive zero coins. Finally, the experimenter left and a secondary experimenter asked children a forced choice question whether they thought the players were real or pretend, with overall 80% of children saying the players were real. The pattern of results did not change even after excluding 20% of children who said the players were pretend players.

In this and subsequent experiment, we counterbalanced the order of practice trials and test trials, the position of green and red buttons, and the other player's roles. We video-recorded testing sessions (unless parents disagreed with video recording). For consistency, protocols were administered verbally by the same main experimenter across participants within an experiment, although the main experimenter in each experiment was different. Experimenters were not blind to conditions or research questions. However, they were trained to remain neutral as much as possible throughout the testing session.

Data Coding and Analyses

Children's responses were recorded by GameMaker Studio (https://www.yoyogames.com) and later entered into a spreadsheet by independent coders. All statistical analyses were conducted with R statistical software (R Version 4.1.2; R Core Team, 2021). We analyzed children's punishment rate with generalized linear mixed models (GLMMs) using the package glmmTMB (Brooks et al., 2017). We compared a full GLMM, which included preregistered predictors of interest (e.g., condition, age, allocation type) and interactions among these predictors as fixed effects and subject ID as a random effect with a null model, which included only subject ID as a random intercept (see Results section for the full model in each experiment). If the full model provided a significantly better fit to the data compared to the null model, we assessed changes in model fit using using likelihood ratio tests (LRTs). Specifically, we created a minimal model by sequentially dropping single terms from the full model, and finalized our minimal model when dropping single terms no longer provided a better fit to the data.

Although not preregistered, as a post hoc analysis, we employed Bayesian statistics to provide more information about the robustness of our findings (see also Lee & Warneken, 2022b for similar approach). Specifically, traditional null hypothesis significance testing (NHST) cannot distinguish between true negative and false negative (Aczel et al., 2018). On the contrary, Bayesian analysis can provide more information on whether the nonsignificant effect is likely to be a true negative or false negative. Specifically, a Bayes

factor (BF) quantifies the degree to which the data favor the null hypothesis over the alternative hypothesis, and vice versa (Aczel et al., 2018; Wagenmakers et al., 2016). Conventionally, a BF between 1 and 3 indicates anecdotal evidence, a BF >3 suggests moderate evidence, and a BF >10 provides strong evidence in favor of one hypothesis over the other (Jeffreys, 1961). BF_{10} refers to the BF in favor of the alternative hypothesis over the null hypothesis, whereas BF_{01}, which is the inverse number of BF_{10}, refers to the BF in favor of the null hypothesis over the alternative hypothesis.

We computed BFs by comparing a GLMM in which a predictor of interest was included with a GLMM in which the predictor was not included, using the package brms (Bürkner, 2017). As in the main analyses, we included subject ID as a random intercept in these models. The population-level regression coefficients had a weakly informative Student's *t* distribution prior which was zero-centered with three degrees of freedom and a scale of 2.5 (Gelman et al., 2008). All models were run with 10,000 iterations with the first half as burn-in. Rhat was <1.01 for all parameters, indicating convergence (Vehtari et al., 2021).

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and research materials are available at https://osf.io/q3nhc. Experimental designs and data analyses were preregistered.

Results

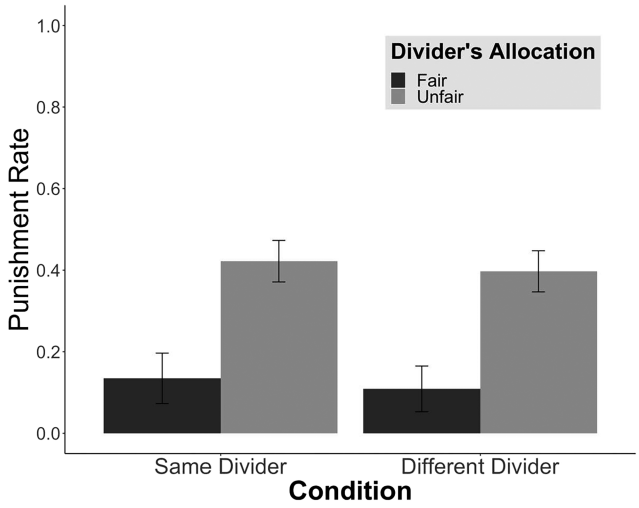
As preregistered, we ran a full GLMM on children's punishment (0 = *no punishment*, 1 = *punishment*) with main effects of allocation type (fair vs. unfair), condition (same divider vs. different divider), and age in months, an interaction between allocation type and condition, an interaction between allocation type and age, and an interaction among allocation type, condition and age as fixed effects and subject ID as a random effect. The full model provided a significantly better fit to the data than the null model with only random intercepts, LRT, $\chi^2(7) = 86.44$, $p < .001$.

Further analysis revealed that there was a significant main effect of allocation type, LRT, $\chi^2(1) = 82.627$, $p < .001$, showing that children punished unfair offers ($M_{same} = 0.40$, $SD_{same} = 0.49$; $M_{different} = 0.42$, $SD_{different} = 0.49$) more often than fair offers ($M_{same} = 0.11$, $SD_{same} = 0.31$; $M_{different} = 0.13$, $SD_{different} = 0.34$; see Figure 2) as shown in other studies with children (e.g., House et al., 2020; Lee & Warneken, 2022b; McAuliffe et al., 2015). However, other main or interaction effects did not reach significance, all *p*s > .20. This result indicates that children punished unfair dividers to a similar degree in the same and different divider condition.

Furthermore, the results from the Bayesian analysis confirmed the nonsignificant difference involving condition. We computed a BF for a main effect of condition as a post hoc analysis. This revealed strong evidence in favor of an absence of the main effect of condition ($BF_{01} = 11.05$), suggesting that the data were about 11 times more likely to be observed under the hypothesis that children's punishment rate is not affected by condition (same vs. different). Similarly, a BF for an interaction between condition and allocation type ($BF_{01} = 6.30$) and a three-way interaction among condition, allocation type, and age ($BF_{01} = 5.74$) were in favor of an absence of the interaction effects.

Figure 2

Average Punishment Rate in Third-Party Punishment Game by Condition and Allocation Type in Experiment 1



Note. Error bars represent 95% confidence intervals.

Discussion of Experiment 1

Experiment 1 examined if 5- to 9-year-old children's third-party punishment is driven by the anticipation of future interactions with the same unfair individual. Our results suggest that children overall punish unfair allocations more often than fair allocations. Importantly, however, children's rate of punishing unfair allocations did not differ significantly depending on the possibility of encountering the same divider in the future. This finding challenges the personal deterrence hypothesis, in which children should increase their punishment of unfair allocations when the same divider would decide how to split resources with the child participant in the following game. Also, our findings do not support the appeasement hypothesis, in which children would fear retaliation from the divider and would punish less often in the same divider condition (vs. different divider). As our comprehension checks showed, this finding cannot be due to a general lack of understanding of whether they would interact with the same or a different partner. Overall, these findings suggest that children's third-party punishment is not driven by the prospect of future interactions.

However, there are two alternative explanations for children's insensitivity to the same versus different divider conditions. One possibility is that while children understood who they would be interacting with in the future, they did not infer unfair treatment of themselves from observing how the selfish divider treated a third party. This possibility is supported by children's optimistic expectations about fair sharing. Concretely, in Experiment 1, a majority of children (77%) expected that they would receive three out of six coins from a divider in a sharing game regardless of conditions. Interestingly, even in the same divider condition in which the divider made a selfish allocation more frequently (i.e., six out of eight trials) than a fair allocation, most children expected that they would still receive a fair allocation from the same divider in the following game. Thus, children's general optimism about how they would be treated might have led to a nonsignificant difference in

punishment rate between conditions. Alternatively, it is also possible that a binary measure of children's expectations (i.e., receiving zero vs. three coins) might have inflated children's optimistic expectations. To address this possibility, in Experiment 2, children chose from zero to six coins to predict how a divider would share coins with themselves, providing a more fine-grained, continuous measure for children.

Another possibility is that it was difficult for children to infer the communicative function of punishment in this context. Specifically, children may not conceptualize the punitive behavior of taking coins per se as a disapproval signal sent to a transgressor. Thus, in Experiment 2, we allowed children to send a verbal message—the clearest communicative form—to a divider, so that they would not have difficulties in inferring the connection between their punishment choice and the disapproving message they signal to the divider.

Experiment 2

Experiment 1 showed that children are not sensitive to the prospect of encountering the same divider in the future. Yet, the use of monetary punishment (i.e., taking coins) in Experiment 1 might have underestimated children's sensitivity to this information. The use of monetary punishment in Experiment 1 allowed us to build on the existing literature and to compare the current finding with prior studies that used similar tasks. However, in everyday life, people use different punishment strategies (e.g., confrontation, gossip, social exclusion) to enforce social norms (Molho et al., 2020). In fact, in the existing literature, punishment is not limited to monetary punishment and can take many forms such as a time-out, verbal admonishment, gossip, and incarceration (Bregant et al., 2016; Dunlea & Heiphetz, 2021; Guala, 2012; Vaish et al., 2016). In particular, adults often communicate their disapproval by sending a negative verbal message to those who shared unfairly with them (Ellingsen & Johannesson, 2008). Furthermore, they are more likely to share resources fairly when they anticipate receiving verbal feedback from the recipient (Ellingsen & Johannesson, 2008). This result shows that people are not only sensitive to verbal feedback from others but actively use it to express their disapproval of unfairness, suggesting that verbal communication could be one way to enforce norms other than monetary punishment.

Moreover, results from a meta-analysis of adult studies have shown that rewards and punishments exhibited a statistically equivalent positive effect on cooperation (Balliet et al., 2011). The finding demonstrates that rewards should be studied alongside punishment as another major way to enforce norms. Additionally, punishment often involves aggressive and destructive behaviors, which could induce retaliation and destroy social relationships as well as the destruction of overall resources available in a group. Therefore, it would be especially important to study if/how children use rewards to enforce social norms, which would have a more positive influence on social relationships compared to punishment.

In Experiment 2, instead of measuring monetary punishment, we introduced verbal messages as a way for third parties to enforce fairness norms and to explicitly express their approval or disapproval of another person's sharing behavior. To be concrete, there were two between-subject conditions: (a) Negative message condition in which children could decide whether to send "Boo!" to a divider or not, and (b) positive message condition in which children could decide whether to send "Yay!" to a divider or not.

We aimed to achieve two goals with these changes. First, by making the communication of disapproval clearer with a verbal message, we could determine whether children's insensitivity to the same versus different divider condition is specific to monetary punishment or whether it is a general insensitivity children possess regardless of punishment type (monetary vs. verbal). Second, by introducing a verbal reward, we could examine the extent to which children use rewards to enforce fairness norms by comparing it with the use of verbal punishment. Also, this would allow us to test whether children's insensitivity to the prospect of future interactions is specific to punishment contexts or whether it is generalized to reward contexts as well. For example, children might reward fair dividers more often in the same (vs. different) divider condition to encourage fair sharing with children themselves in the future.

In Experiment 2, we tested a slightly older age group of 8- to 10-year-olds because prior work has shown that children become better at future thinking with increasing age (Atance & Meltzoff, 2005) and manage their reputation across different cooperative contexts over development (see Engelmann & Rapp, 2018 for review). Therefore, on the one hand, children have all the cognitive and social-cognitive abilities at their fingertips to use them for influencing others in a third-party context. On the other hand, if this older age group of children are still insensitive to the prospect of future interactions, then it would provide further evidence against the personal deterrence or appeasement hypothesis that children punish with the motive to change the divider's future behavior toward themselves. In sum, Experiment 2 tested the personal deterrence and the appeasement hypothesis more strictly by introducing verbal messaging and by focusing on older age groups.

Method

Participants

Our final sample included $n = 128$ eight- to 10-year-old children ($M = 113.9$ months, range = 96–131 months; $n = 64$ per condition; forty-three 8-year-olds, thirty-eight 9-year-olds, forty-seven 10-year-olds; 65 male, 63 female). Children were tested via an online meeting platform (Zoom) between September 2021 and January 2022 because in-person research was not possible due to the COVID-19 pandemic. Online research has been widely used by developmental researchers since the pandemic (Chuey et al., 2021; Sheskin et al., 2020). Demographic information such as race, education, and income was not obtained. Six additional children were excluded because there was an experimental error (2), a parent reported that their child had a developmental delay (2), the child wanted to stop the study before the test phase (1), or the child participated in a similar study before (1).

A priori power analysis established that our sample size ($n = 64$ per condition) would be large enough to detect a medium-sized condition effect ($d = 0.5$) with sufficient power (0.80) at α .05 (see our preregistration at https://aspredicted.org/m98pk.pdf).

Experimental Design and Procedure

After parents provided consent online, researchers scheduled an online testing session, in which children participated with a digital device (e.g., computer, tablet). At the beginning of the session, the experimenter confirmed that the child could see our experimental stimuli from their device. Then, a female experimenter introduced the

computer game referred to as the "coin game" and explained that players could collect virtual coins to later exchange for watching fun videos from National Geographic Kids (https://kids.nationalgeographic.com/videos). During a prize introduction, children learned that the more coins they have during the coin game, the more videos they would be able to watch afterward. All children (100%) understood that more coins could be exchanged for more videos. A vast majority of the children reported that these videos are interesting (93%) and that they would like to watch more of these videos than just one video (87%), establishing that videos could be an effective incentive for children in online research.

The practice phase was similar to Experiment 1, except that children learned how to send a verbal message to another player. Specifically, children were assigned to either the positive message condition or the negative message condition (between-subjects). In the negative message condition, children had two buttons: a gray "no sound" button and a blue "thumbs down" button (see Figure 3A). Children learned that pressing the blue button would result in the child sending a prerecorded voice message saying "Boo!" to a divider in the third-party punishment game. Critically, to press the blue button, children had to pay one of their coins into the vacuum. In contrast, pressing the gray button incurred no cost to the child and did not send any voice messages to the divider. In the positive message condition, children saw a blue "thumbs up" button that they could press after paying a coin, which then sent a "Yay!" sound to a divider (see Figure 3B).

Children practiced pressing each button on the experimenter's screen, using the remote control function in Zoom. Their responses were recorded automatically by Qualtrics. Nine out of 128 children were unable to use the remote control function over Zoom potentially due to their computer or Zoom setting. In this case, the child told the experimenter which button he or she would like to press and the experimenter pressed the button for the child. The pattern of results did not change even after excluding these nine children.

As in Experiment 1, there were four practice trials in total. Children practiced four possible outcomes of each button (sending message vs. no message) in each allocation type (fair vs. unfair). All participants included in the data analysis answered the comprehension check questions correctly by the end of the practice phase.

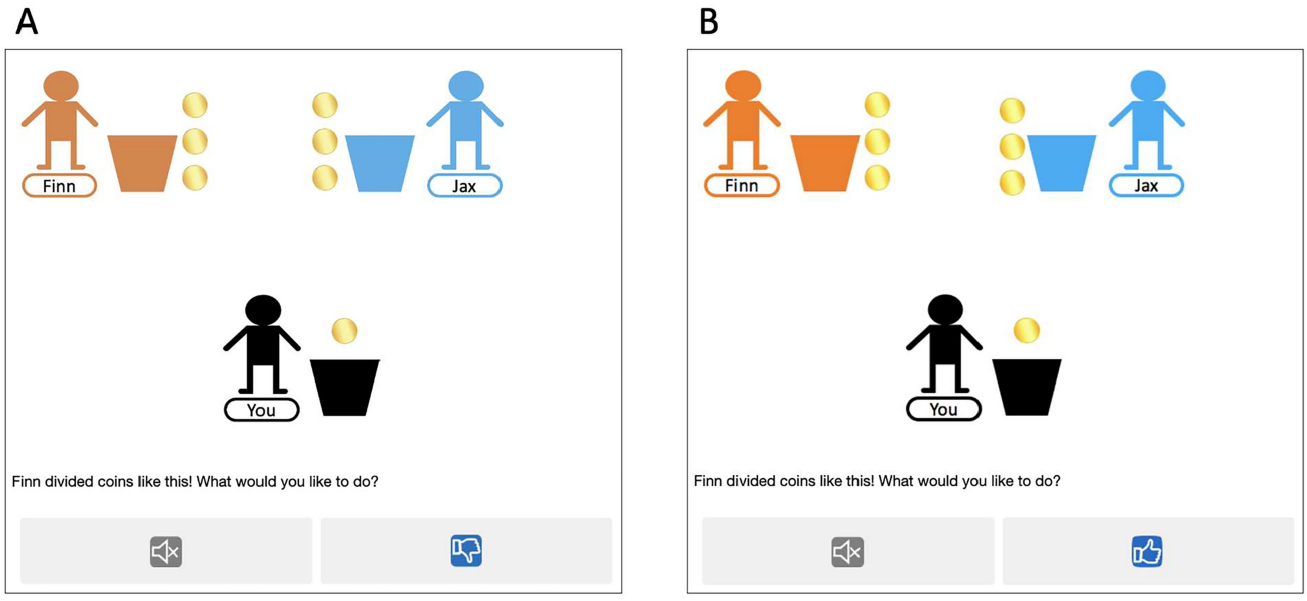
As in Experiment 1, the experimenter pretended to make a phone call to the other players on speakerphone and checked if they were ready to play the game. Also, as in Experiment 1, children received 20 coins as their initial endowment before the test phase.

The test phase consisted of eight trials of the third-party punishment game in total. During the test phase, children were presented with two blocks in counterbalanced order: The same divider block and a different divider block (within-subjects). We used different players in each block to prevent potential carryover between blocks. However, the players and their roles remained the same within each block. In each block, there were four test trials, either all fair (3:3) or all unfair (6:0) allocations. That is, within a block, a divider in the third-party punishment game allocated coins either always fairly or always unfairly. Half of the children were randomly assigned to interact with the same divider dividing coins fairly and a different divider dividing coins unfairly. Vice versa, the other half interacted with the same divider dividing coins unfairly and a different divider dividing coins fairly.

Before the third-party punishment began, most children (97%) correctly identified the divider and the recipient of the sharing game. Also, virtually all children (96%) understood that only they themselves and the divider in the third-party punishment game

Figure 3

Computer Screen Displays During the Test Phase in the Negative Message Condition and Positive Message Condition



Note. (A) The screen display in the negative message condition. In this example, Finn kept three coins for himself and gave three coins to Jax who was the recipient. Children learned that when they press the blue thumps down button (right), a voice message saying "Boo!" will be sent to Finn. (B) The screen display in the positive message condition. Children learned that when they press the blue thumbs up button (right), a voice message saying "Yay!" will be sent to Finn. See the online article for the color version of this figure.

could hear their voice message, while the recipient in the third-party punishment game and the new/different divider could not hear it. The experimenter provided corrective feedback to the few children who answered incorrectly before the test phase began.

At the end of each test block, the experimenter confirmed that children recalled how the divider shared coins during the third-party punishment game (95% of times children answered correctly). In addition, the experimenter asked children to predict how many coins the divider in the sharing game would share with the child themselves from zero to six coins. Results revealed that when asked to make a prediction about the same divider, children expected that the divider would share fewer coins when the person allocated unfairly in the third-party punishment game ($M = 1.39$, $SD = 1.78$) than when the person allocated fairly in the game ($M = 3.34$, $SD = 1.27$), LRT, $\chi^2(1) = 55.35$, $p < .001$, indicating that children make a prediction about future sharing based on sharing they observed during the third-party punishment game and adjust their expectations about the divider accordingly. In contrast, when asked to make a prediction about a new, different divider who did not have a prior sharing history, children expected that the new divider would share a similar number of coins regardless of whether they observed fair ($M = 3.01$, $SD = 1.19$) or unfair allocations ($M = 2.74$, $SD = 1.39$) during the punishment game, LRT, $\chi^2(1) = 2.08$, $p = .15$ (see the online supplemental materials for more detailed results). Hence, unlike Experiment 1 in which most children showed optimistic expectations about fair sharing regardless of same versus different divider, this result shows that when assessed with a continuous measure, at least 8- to 10-year-olds appropriately adjusted their expectations about how a divider would share coins with them based on their observations of the divider's sharing in the third-party punishment game.

At the end of the study, children reported that they would feel happy (100%) if they received a message saying "Yay!," while they would feel sad (96%) if they received a message saying "Boo!" from another person. This result shows that children understood how each message would influence their own feelings.

Finally, we found that 75% of children reported the players were real. The pattern of results did not change even after excluding 25% of children who reported the players were pretend players.

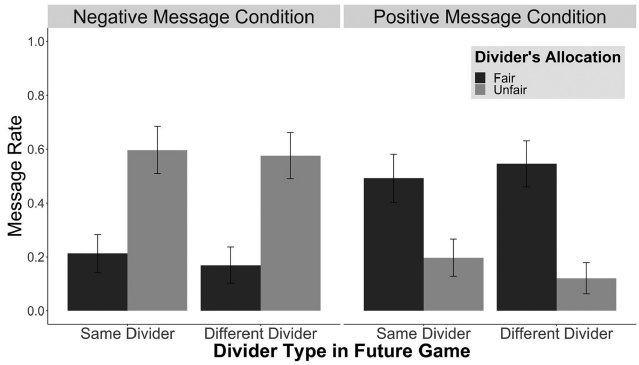
Results

As preregistered, we ran a full GLMM on children's decision to send a voice message to the divider in the third-party punishment game (0 = *no message*, 1 = *sent message*) with condition (negative message vs. positive message), divider type (same vs. different), allocation type (fair vs. unfair), age in months, and all interactions among these variables as fixed effects and subject ID as a random effect. The full model provided a better fit to the data than the null model with only random intercepts, LRT, $\chi^2(15) = 191.48$, $p < .001$.

Further analyses revealed that there was a significant interaction between condition and allocation type, LRT, $\chi^2(1) = 171.75$, $p < .001$ (see Figure 4). To unpack the interaction, we followed up by analyzing the negative and positive message conditions, respectively. In the negative message condition, children were more likely to send a negative message to unfair dividers than to fair dividers, LRT, $\chi^2(1) = 91.24$, $b = 1.92$, $SE = 0.22$, $p < .001$. The opposite pattern was observed in the positive message condition: Children were more likely to send a positive message to fair dividers than to unfair dividers, LRT, $\chi^2(1) = 80.78$, $b = -1.87$, $SE = 0.23$, $p < .001$. In summary, children sent the "Boo!" message more often to unfair dividers than fair dividers, and the "Yay!" message to fair dividers more often than to unfair dividers.

Figure 4

Average Rate of Sending a Message by Condition, Divider Type, and Allocation Type in Experiment 2



Note. Error bars represent 95% confidence intervals.

Additionally, we found a significant main effect of age, LRT, $\chi^2(1) = 6.54$, $b = 0.02$, $SE = 0.01$, $p = .011$, suggesting that with increasing age, children were more likely to send a message. However, we found a nonsignificant main effect of divider type (same vs. different), LRT, $\chi^2(1) = 1.08$, $p = .300$. Also, there were no significant interactions involving divider type, all $ps > .065$ (see the online supplemental materials for the marginally significant effect).

Overall, these findings suggest that children send positive or negative messages depending on the (un)fairness of the divider, independently of whether they would encounter the same or different divider in the future.

Furthermore, the results from the Bayesian analysis confirmed this finding. We computed BFs for effects involving divider type (same vs. different) to examine if there is evidence that children's rate of sending a message differs by divider type. This revealed strong evidence in favor of an absence of the main effect of divider type ($BF_{01} = 20.60$) and moderate evidence in favor of an absence of the interaction effect between divider type and allocation type ($BF_{01} = 8.02$), suggesting that the data were about 20 times and eight times, respectively, more likely to be observed under the hypothesis that children's rate of sending a message is not affected by divider type. All other BFs for interaction effects involving divider type were in favor of an absence of the effects or at least not in favor of the presence of these effects ($BF_{01} = 1.90$ for a four-way interaction among divider type, age, allocation type, and condition, $BF_{01} = 7.57$ for a three-way interaction among divider type, age, and allocation type, $BF_{01} = 9.55$ for a three-way interaction among divider type, age, and condition, $BF_{01} = 1.20$ for a three-way interaction among divider type, allocation type, and condition, $BF_{01} = 21.25$ for a two-way interaction between divider type and age, $BF_{01} = 9.63$ for a two-way interaction between divider type and condition).

Discussion of Experiment 2

In Experiment 2, children sent negative messages more often to unfair dividers than to fair dividers, whereas they sent positive

messages more often to fair than unfair dividers. Importantly, children's rate of sending messages did not change depending on whether the divider in the future will be the same or different. The results demonstrate that (a) children are indifferent to same versus different divider in the verbal punishment context as well as monetary punishment context and (b) that their insensitivity to the prospect of future interactions is not limited to the punishment context but extends to the reward context as well. Overall, the findings provide converging evidence against the personal deterrence or the appeasement hypothesis.

General Discussion

Across two experiments, we examined if children punish differently depending on whether they would encounter an unfair individual in the future. Neither experiment provided support for the personal deterrence hypothesis, according to which children use third-party punishment to deter uncooperative behavior directed at themselves. Also, we did not find evidence consistent with the appeasement hypothesis, according to which children would fear retaliation from a divider, and thus would show reduced third-party punishment against the divider who would interact with the child again in the future. The Bayesian analysis also confirmed this nonsignificant difference between divider types (same vs. different) by showing that the data were in favor of the absence of these effects in both experiments. Together, it would be reasonable to conclude that children's third-party punishment is unlikely to be driven by a personal motive to strategically manipulate peers in a third-party context to be treated better by them in the future.

If neither personal deterrence nor appeasement plays a major role in children's third-party punishment, then which motive might underlie their punishment? We speculate that children's sense of fairness and their inclination to rectify inequality might be their major motivation for punishment at least in the current context. Specifically, the pattern of results observed across two experiments is in line with the fairness-based account in which children would punish unfair over fair allocations regardless of whether they would encounter the same divider in the future or not. According to this account, children's primary motive for punishment is to restore existing inequality between individuals rather than to manipulate others to get better treatment for themselves. Existing literature supports this possibility. With age, children enforce fairness norms in an increasingly more impartial manner (McAuliffe et al., 2017; Shaw, 2013). Converging evidence comes from studies showing that from around ages 7 or 8, children avoid receiving more resources than their peers at a cost to adhere to the norm of equality (Blake & McAuliffe, 2011; Shaw & Olson, 2012). Furthermore, between ages 5 and 9, children's third-party punishment aims at restoring equality between individuals with increasing age and prefers a punisher who increases equality between two other individuals over another punisher who decreases equality (Lee & Warneken, 2020, 2022a). Additionally, third-party punishment in children is influenced by whether a peer is treated fairly or unfairly, not by whether children themselves is treated fairly or unfairly by others in previous interactions (Lee & Warneken, 2022b). Together, our study provides converging evidence for the notion that over development, children stick to and enforce the fairness norms in a principled way and are not acting out of self-oriented motives.

One noteworthy result is that we found no significant effects involving age in Experiment 1, suggesting that 5-year-olds' punishment patterns ($M_{unfair} = 0.42$, $SD_{unfair} = 0.50$; $M_{fair} = 0.15$, $SD_{fair} = 0.36$) were similar to those from older children. This finding contrasts with earlier work in which 6-year-olds, but not 5-year-olds, punish unfair (vs. fair) allocations more selectively (Lee & Warneken, 2022b; McAuliffe et al., 2015). Why did 5-year-olds in Experiment 1 punished unfair allocations more often than fair allocations as did older age groups? One possible reason is that unlike prior work, children in the current research were reminded of future sharing (i.e., how they might be treated by another person in the future) before third-party punishment game across conditions. Hence, this experimental procedure might have functioned as a priming, making young children think about the future and leading to the increased sensitivity to unfairness.

Furthermore, the current findings contrast with adult literature in which adults use third-party punishment to deter personal mistreatment (Delton & Krasnow, 2017; Krasnow et al., 2016). This finding suggests a developmental discontinuity in motivations for punishment. Why did 5- to 10-year-olds not show deterrence-based motive as adults? It is possible that the deterrence motivation for punishment develops in adolescence. During adolescence, peer relationships become increasingly important for adolescents, potentially contributing to increased concern for reputation such as how they would be perceived by their peers (Cage et al., 2016; Jankowski et al., 2014; Sebastian et al., 2008). Therefore, we speculate that deterrence motive might emerge during adolescence when reputation management among peers becomes a major part of their social life.

To the best of our knowledge, this is one of the first studies to investigate how children use verbal praise versus admonishment in response to (un)fair treatment of third parties. While past work has shown that 16-month-old infants are more likely to touch a screen that produces a positive verbal comment (vs. a negative comment) when a previously fair divider is present on the screen, they did not show a selective response to the presence of an unfair divider (Ziv et al., 2021). The present work found that at least 8- to 10-year-old children were more likely to send an approving message to fair dividers than to unfair ones, whereas they were more likely to send a disapproving message to unfair dividers than to fair ones. These results show that by age 8, children consider verbal messaging as not only an appropriate response to fair sharing but also a valid response against unfair sharing.

We would like to address potential concerns about our findings. One concern is that children did not show a significant difference between the same and different divider conditions because they did not infer that they would be treated unfairly by a selfish divider. However, this possibility is unlikely given that children in Experiment 2 predicted that they themselves would receive fewer coins from someone who treated others unfairly than someone who treated others fairly. Despite the fact that children predicted potential unfair treatment directed at themselves, they still did not increase or decrease their rate of sending a disapproving message to the selfish divider. It is possible that children would have shown a sensitivity to same versus different dividers if they had had direct interaction experience with the divider before the third-party punishment game. Specifically, if children had received an unfair allocation directly from the same divider, and thus they know for sure that they would receive unfair treatment from the same person in the future, they might have punished unfairness more often in the same divider (vs. different divider) condition later, although this experimental design could not disentangle whether

children's third-party punishment is driven by deterrence or personal retribution.

Another concern is that children might not understand whether they would be reencountering the divider or not. However, before the test phase, our comprehension checks established that children understood the conditions of the game and who they would play with. Overall, a majority of children (88% in Experiment 1, 97% in Experiment 2) correctly identified the dividers and recipients of the third-party punishment game and a subsequent sharing game.

Additionally, children might have been insensitive to the same versus different divider because over trials, children in the same divider condition learned that punishment was not effective in changing the unfair divider's behavior and thus decreased their punishment rate over trials. However, visual inspection of the data did not support this possibility. That is, children who chose to punish in a previous unfair trial did not seem to withdraw from punishing unfair dividers over trials (see the online supplemental materials for more details), ruling out the possibility that children's perception about punishment's ineffectiveness might have reduced their punishment rate.

Limitations and Future Directions

While the current research contributes to scientific understanding of proximate mechanisms underlying punishment in early childhood, like all research programs, it cannot possibly answer all questions. One potential direction for future research is to examine how children's punitive responses in a computer-based task generalize to face-to-face social interactions. Our computer-based task provides ultimate control over the stimuli presentation and systematic manipulation of key variables, including quantitative analyses of the degree of punishments and rewards. However, it does not involve in-person social interactions, which might limit the generalizability of the current findings. In face-to-face interactions, appeasement might play a larger role in children's decision to punish because a perceived threat of retaliation might be bigger (e.g., Kenward & Östh, 2015). Future research can test this possibility by examining the correlation in children's punishment between computer-based and face-to-face settings.

Second, future research should probe the message that children try to communicate to transgressors. The kind of punishment children could use in our experiments may differ from naturalistic responses that children might show against unfair sharing. Specifically, children may not take resources away from unfair sharer in their everyday life to punish the person. While Experiment 2 introduced a more ecologically valid response (i.e., making a verbal comment), it is still limited by the experimental design of predesignated verbal responses. We used the prerecorded voice messages to have a better control over our experimental manipulation and systematic measurement of the key dependent variable. However, we may have restricted children's ability to freely communicate with the divider. For instance, children often tattle or protest against transgressors by saying, for example, "You are not supposed to do that" (Vaish et al., 2011; Yucel & Vaish, 2018). Future research could consider allowing children to record and send their own words to transgressors.

Third, it would be important to test punishment motivations for different norm violations. The present work focused on the fairness norm because third-party punishment against unfair sharing has been well documented in children and it has been considered an important milestone in fairness development (Gummerum & Chu, 2014; House et al., 2020; Lee & Warneken, 2022a; McAuliffe et al., 2015). However, other forms of norm violation may present different motivations for imposing third-party punishment. For example, children's third-party punishment might be motivated by personal deterrence in a more serious moral transgression (e.g., physical harm) because they might be more sensitive to preventing harm directed at themselves in these contexts. Future work should explore this possibility by presenting more serious transgressions.

Fourth, another future direction is to study how punishment motivations might vary between or within participants. One might argue that we found a nonsignificant difference between same and different divider because the personal deterrence and appeasement motives might have canceled each other out. For example, about half of the children were motivated by personal deterrence, while the other half by appeasement. Although the current research was not designed to exclude this possibility completely, it seems that neither personal deterrence nor appeasement is dominant enough to override the other. If one of these motives had played a more dominant role in children's punishment either within or between participants, we should have found a significant difference between the same versus different divider, which was not the case in either experiment. In future studies, one approach could be to measure children's responses in other tasks and examine correlations between those tasks and the punishment task. For example, if children's punishment is motivated by their strategic manipulation of peers for future interactions, those who exploit these strategies more in nonpunitive tasks might be more likely to show third-party punishment. Conversely, if their punishment is motivated by genuine fairness concerns, then those who value equality more than self-interest in other tasks might be more likely to show third-party punishment against unfair sharing.

Lastly, the current research could not test all existing theories in punishment motivations. Instead, we focused on testing two plausible hypotheses—the personal deterrence hypothesis and the appeasement hypothesis—which have received relatively little attention in developmental research. Yet, our results do not rule out other possible motivations for punishment. For instance, even though children did not punish to deter unfair treatment directed at themselves, it is still possible that children's punishment is motivated to deter potential unfairness directed at other people more broadly. Also, children's punishment might be subject to other selfish motives such as increasing one's reputational benefit by advertising them as a trustworthy partner. Specifically, in the current work, the experimenter was present during the testing sessions. Thus, children might have been concerned more about their reputation from the experimenter rather than that from dividers, potentially leading them to be less sensitive to the divider manipulation. Although the current experimental design could not examine all of these possibilities, future research could adapt our task to focus on testing these accounts by manipulating whether children's intervention decisions are available to observers and/or experimenters.

Constraints on Generality

Our findings show that children are more likely to punish unfair over fair allocations as an uninvolved third party. This finding is consistent with prior research on children's third-party punishment (e.g.,

Arini et al., 2021; Gummerum & Chu, 2014; House et al., 2020; Lee & Warneken, 2022a, 2022b; McAuliffe et al., 2015). Thus, we believe that this result will be replicated in a similar setup with child participants from similar subject pools (i.e., those from Western, Educated, Industrialized, Rich, and Democratic [WEIRD] societies; e.g., Henrich et al., 2010). However, our findings may not be reproducible in children from non-WEIRD societies. There are cross-cultural variations in the degree to which adults punish unfairness (Henrich et al., 2006). Moreover, some societies even punish behaviors that are often considered as prosocial in WEIRD societies (Herrmann et al., 2008). We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

In conclusion, 5- to 10-year-olds were more likely to punish unfair over fair allocations as a third party. However, their third-party punishment was not affected by the prospect of future interactions with the same unfair person. We found the insensitivity to future partners when children used verbal punishment as well as monetary punishment. The present research shows that at least in unfair allocation contexts, third-party punishment in early childhood is unlikely to be shaped by the possibility of future interactions with the same partner.

References

Aczel, B., Palfi, B., Szollosi, A., Kovacs, M., Szaszi, B., Szecsi, P., Zrubka, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2018). Quantifying support for the null hypothesis in psychology: An empirical investigation. *Advances in Methods and Practices in Psychological Science*, *1*(3), 357–366. https://doi.org/10.1177/2515245918773742

Arini, R. L., Wiggs, L., & Kenward, B. (2021). Moral duty and equalization concerns motivate children's third-party punishment. *Developmental Psychology*, *57*(8), 1325–1341. https://doi.org/10.1037/dev0001191

Atance, C. M., & Meltzoff, A. N. (2005). My future self: Young children's ability to anticipate and explain future states. *Cognitive Development*, *20*(3), 341–361. https://doi.org/10.1016/j.cogdev.2005.05.001

Balafoutas, L., Grechenig, K., & Nikiforakis, N. (2014). Third-party punishment and counter-punishment in one-shot interactions. *Economics Letters*, *122*(2), 308–310. https://doi.org/10.1016/j.econlet.2013.11.028

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*(4), 594–615. https://doi.org/10.1037/a0023489

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair" eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. https://doi.org/10.1016/j.cognition.2011.04.006

Bregant, J., Shaw, A., & Kinzler, K. D. (2016). Intuitive jurisprudence: Early reasoning about the functions of punishment. *Journal of Empirical Legal Studies*, *13*(4), 693–717. https://doi.org/10.1111/jels.12130

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cage, E., Bird, G., & Pellicano, L. (2016). 'I am who I am': Reputation concerns in adolescents on the autism spectrum. *Research in Autism Spectrum Disorders*, *25*, 12–23. https://doi.org/10.1016/j.rasd.2016.01.010

Chuey, A., Asaba, M., Bridgers, S., Carrillo, B., Dietz, G., Garcia, T., Leonard, J. A., Liu, S., Merrick, M., Radwan, S., Stegall, J., Velez, N., Woo, B., Wu, Y., Zhou, X. J., Frank, M. C., & Gweon, H. (2021). Moderated online data-collection for developmental research: Methods and replications. *Frontiers in Psychology*, *12*, Article 734398. https://doi.org/10.3389/fpsyg.2021.734398

Delton, A. W., & Krasnow, M. M. (2017). The psychology of deterrence explains why group membership matters for third-party punishment. *Evolution and Human Behavior*, *38*(6), 734–743. https://doi.org/10.1016/j.evolhumbehav.2017.07.003

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13335–13340. https://doi.org/10.1073/pnas.1102131108

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, *33*(1), 145–167. https://doi.org/10.1007/s00199-007-0212-0

Dunlea, J. P., & Heiphetz, L. (2021). Children's and adults' views of punishment as a path to redemption. *Child Development*, *92*(4), e398–e415. https://doi.org/10.1111/cdev.13475

Ellingsen, T., & Johannesson, M. (2008). Anticipated verbal feedback induces altruistic behavior. *Evolution and Human Behavior*, *29*(2), 100–105. https://doi.org/10.1016/j.evolhumbehav.2007.11.001

Engelmann, J. M., Herrmann, E., Tomasello, M., & Zalla, T. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One*, *7*(10), Article e48433. https://doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, *20*, 92–95. https://doi.org/10.1016/j.copsyc.2017.08.024

Fehl, K., Sommerfeld, R. D., Semmann, D., Krambeck, H. J., & Milinski, M. (2012). I dare you to punish me—Vendettas in games of cooperation. *PLoS One*, *7*(9), Article e45093. https://doi.org/10.1371/journal.pone.0045093

Fehr, E., & Fischbacher, U. (2004). Third party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, *35*(1), 1–15. https://doi.org/10.1017/S0140525X11000069

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second- and third-party punishment behavior. *Cognition*, *133*(1), 97–103. https://doi.org/10.1016/j.cognition.2014.06.001

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. https://doi.org/10.1126/science.1127333

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808

House, B. R., Kanngiesser, P., Barrett, H. C., Yilmaz, S., Smith, A. M., Sebastian-Enesco, C., Erut, A., & Silk, J. B. (2020). Social norms and cultural diversity in the development of third-party punishment. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1925), Article 20192794. https://doi.org/10.1098/rspb.2019.2794

Jankowski, K. F., Moore, W. E., Merchant, J. S., Kahn, L. E., & Pfeifer, J. H. (2014). But do you think I'm cool? Developmental differences in striatal recruitment during direct and reflected social self-evaluations. *Developmental Cognitive Neuroscience*, *8*, 40–54. https://doi.org/10.1016/j.dcn.2014.01.003

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.
Kenward, B., & Östh, T. (2012). Enactment of third-party punishment by 4-year-olds. *Frontiers in Psychology*, *3*, Article 373. https://doi.org/10.3389/fpsyg.2012.00373
Kenward, B., & Östh, T. (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior*, *41*(5), 413–420. https://doi.org/10.1002/ab.21568
Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science*, *27*(3), 405–418. https://doi.org/10.1177/0956797615624469
Lee, Y., & Warneken, F. (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. *Cognition*, *205*, Article 104374. https://doi.org/10.1016/j.cognition.2020.104374
Lee, Y., & Warneken, F. (2022a). Does third-party punishment in children aim at equality? *Developmental Psychology*, *58*(5), 866–873. https://doi.org/10.1037/dev0001331
Lee, Y., & Warneken, F. (2022b). The influence of age and experience of (un) fairness on third-party punishment in children. *Social Development*, *31*(4), 1176–1193. https://doi.org/10.1111/sode.12604
Marshall, J., Yudkin, D. A., & Crockett, M. J. (2021). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*, *5*(3), 361–368. https://doi.org/10.1038/s41562-020-00975-9
McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), Article 0042. https://doi.org/10.1038/s41562-016-0042
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
Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment in daily life. *Nature Communications*, *11*(1), Article 3432. https://doi.org/10.1038/s41467-020-17286-2
Nelissen, R. M. A., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making*, *4*(7), 543–553. https://doi.org/10.1017/S1930297500001121
Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1–2), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008
Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2010). Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. *Human Morality and Sociality Evolutionary and Comparative Perspectives*, *2*, 72–131. https://doi.org/10.1007/978-1-137-05001-4_5
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881. https://doi.org/10.1037/0012-1649.44.3.875
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Rosati, A. G., Benjamin, N., Pieloch, K., & Warneken, F. (2019). Economic trust in young children. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907), Article 20190822. https://doi.org/10.1098/rspb.2019.0822
Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition*, *121*(2), 219–227. https://doi.org/10.1016/j.cognition.2011.06.007
Sebastian, C., Burnett, S., & Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences*, *12*(11), 441–446. https://doi.org/10.1016/j.tics.2008.07.008
Sebastián-Enesco, C., & Warneken, F. (2015). The shadow of the future: 5-year-olds, but not 3-year-olds, adjust their sharing in anticipation of reciprocation. *Journal of Experimental Child Psychology*, *129*, 40–54. https://doi.org/10.1016/j.jecp.2014.08.007
Shaw, A. (2013). Beyond "to share or not to share": The impartiality account of fairness. *Current Directions in Psychological Science*, *22*(5), 413–417. https://doi.org/10.1177/0963721413484467
Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, *141*(2), 382–395. https://doi.org/10.1037/a0025907
Sheskin, M., Bloom, P., & Wynn, K. (2014). Anti-equality: Social comparison in young children. *Cognition*, *130*(2), 152–156. https://doi.org/10.1016/j.cognition.2013.10.008
Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004
Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS One*, *8*(3), Article e59510. https://doi.org/10.1371/journal.pone.0059510
Twardawski, M., Hilbig, B. E., & Capraro, V. (2020). The motivational basis of third-party punishment in children. *PLoS One*, *15*(11), Article e0241919. https://doi.org/10.1371/journal.pone.0241919
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. *British Journal of Developmental Psychology*, *29*(1), 124–130. https://doi.org/10.1348/026151010X532888
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221
Wagenmakers, E. J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176. https://doi.org/10.1177/0963721416643289
Yamagishi, T., Li, Y., Fermin, A. S., Kanai, R., Takagishi, H., Matsumoto, Y., Kiyonari, T., & Sakagami, M. (2017). Behavioural differences and neural substrates of altruistic and spiteful punishment. *Scientific Reports*, *7*(1), Article 14654. https://doi.org/10.1038/s41598-017-15188-w
Yucel, M., & Vaish, A. (2018). Young children tattle to enforce moral norms. *Social Development*, *27*(4), 924–936. https://doi.org/10.1111/sode.12290
Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (2020). Young children police in-group members at personal cost. *Journal of Experimental Psychology: General*, *149*(1), 182–191. https://doi.org/10.1037/xge0000613
Zheng, X., & Nie, P. (2013). Effective punishment needs legitimacy. *Economic Record*, *89*(287), 522–544. https://doi.org/10.1111/1475-4932.12073
Ziv, T., Whiteman, J. D., & Sommerville, J. A. (2021). Toddlers' interventions toward fair and unfair individuals. *Cognition*, *214*, Article 104781. https://doi.org/10.1016/j.cognition.2021.104781

Received May 16, 2023
Revision received September 18, 2023
Accepted September 30, 2023 ■