



HHS Public Access

Author manuscript

Soc Dev. Author manuscript; available in PMC 2024 July 11.

Published in final edited form as:

Soc Dev. 2024 May ; 33(2): . doi:10.1111/sode.12730.

No one is going to recess: How children evaluate collective and targeted punishment

Sarah Thomas¹, Caroline Kelsey^{1,2,3}, Amrisha Vaish¹

¹Department of Psychology, University of Virginia, Charlottesville, Virginia, USA

²Department of Pediatrics, Division of Developmental Medicine, Boston Children's Hospital, Boston, Massachusetts, USA

³Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

Abstract

This study examined children's responses to targeted and collective punishment. Thirty-six 4–5-year-olds and 36 6–7-year-olds (36 females; 54 White; data collected 2018–2019 in the United States) experienced three classroom punishment situations: Targeted (only transgressing student punished), Collective (one student transgressed, all students punished), and Baseline (all students transgressed, all punished). The older children evaluated collective punishment as less fair than targeted, whereas younger children evaluated both similarly. Across ages, children distributed fewer resources to teachers who administered collective than targeted punishment, and rated transgressors more negatively and distributed fewer resources to transgressors in Collective and Targeted than Baseline. These findings demonstrate children's increasing understanding of punishment and point to the potential impact of different forms of punishment on children's social lives.

Keywords

children; collective punishment; development; fairness; targeted punishment

1 | INTRODUCTION

Human societies, institutions, and cooperative activities are held together by social norms, that is, expectations or rules about how one ought to behave in a given situation (Boyd & Richerson, 2009; Chudek & Henrich, 2011; Schmidt & Tomasello, 2012; Smetana et al., 2014). Those who break social norms are often punished in some way, which

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Correspondence: Amrisha Vaish, Department of Psychology, P.O. Box 400400, University of Virginia, Charlottesville, VA 22904, USA. vaish@virginia.edu.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

typically induces the norm violators to follow the norms in the future (Boyd & Richerson, 1989; Martin et al., 2021). Such punishment is considered a hallmark of our moral psychology (Fehr & Fischbacher, 2003), with some theorizing that the motivation to punish noncooperators has deep biological and evolutionary roots borne out of the benefits of increasing cooperation (Kanakogi et al., 2022; Strobel et al., 2011). Though the specific forms of punishment that individuals engage in are likely learned, culturally transmitted, and stabilized through cultural transmission (e.g., Guala, 2012; House et al., 2020; Salali et al., 2015), and punishment is certainly not the only avenue to increasing cooperation (see Arini et al., 2023; Lee & Warneken, 2020; Riedl et al., 2015), there is general consensus that punishment is an important feature of human psychology that emerges early in development and across cultures (see Henrich et al., 2010; Marshall & McAuliffe, 2022).

Indeed, even infants and young toddlers demonstrate rudimentary forms of punitive behavior towards antisocial agents (e.g., Hamlin et al., 2011; Kanakogi et al., 2022; Ziv et al., 2021). This enforcement continues to strengthen with age, as 3- to 4-year-olds spontaneously enforce norms through acts such as protesting, tattling, actively punishing (e.g., by taking away resources), and reducing prosocial behavior towards norm violators (Masters & Furman, 1981; Riedl et al., 2015; Vaish et al., 2010; Vaish et al., 2011; Yucel & Vaish, 2018). By 5–6 years of age, children are even willing to pay a personal cost to punish transgressors (McAuliffe et al., 2015; Robbins & Rochat, 2011).

In addition to administering punishment themselves, young children also have expectations of normal enforcement. Infants expect that those who act fairly should receive a reward and those who act unfairly should receive punishment (Geraci & Surian, 2023; Meristo & Surian, 2013). Similarly, infants prefer those who intervene in an aggressive situation over those who do not intervene (Kanakogi et al., 2017). Toddlers also show surprise when an agent who defended a victim is punished and they expect nondefenders to receive punishment (Geraci, 2021; Geraci & Surian, 2021). By preschool age, children judge rule violators as deserving of punishment (Kenward & Östh, 2012; Tisak, 1993), positively judge norm enforcers (Lee & Warneken, 2020; Marshall & McAuliffe, 2022), and distribute more resources to a norm enforcer compared to a nonenforcer (Vaish et al., 2016). In addition, young children tend to believe everyone is obliged to intervene, to report the transgression to an authority figure, and to punish antisocial behavior, whereas older children and adults see this as the responsibility of authority figures rather than peers (Loke et al., 2011; Marshall, Mermin-Bunnell, & Bloom, 2020). Furthermore, children demonstrate multiple motivations for administering and endorsing punishment, including retribution, deterrence, and compensation (Arini et al., 2023; Marshall & McAuliffe, 2022; Marshall, Yudkin, & Crocket, 2020; Marshall et al., 2022). Taken together, these findings suggest that from remarkably early in human ontogeny, children begin to sanction norm violators themselves and expect others to do the same.

Importantly, however, children do not blindly endorse all punishment. School-age children begin to show a sensitivity to the fairness of punishment by recognizing that recipients of unfair punishments would experience negative feelings, and they can often cite a personal experience of being similarly treated unfairly (Evans et al., 2001). Children also consider the type of punishment administered when making evaluations. For example, third and fifth

graders reported that presentation punishment (i.e., delivering an unwanted stimulus) was more common and effective overall, but that removal punishment (i.e., taking away a wanted stimulus or privilege) was more effective in reducing the likelihood of moral transgressions than conventional transgressions (Brinker et al., 2003). Overall, children demonstrate an increasingly nuanced understanding of the forms and functions of punishment across childhood.

Though a lot of research has examined the factors that influence children's evaluations of punishments, one important aspect has received little attention in the literature, namely, whom the punishment targets. Consider the following hypothetical: A teacher, Ms. Brown, sees Sally misbehaving in her class and faces a choice about how to punish the misbehavior: Does she take away recess only from Sally or does she make the whole class stay behind? These represent two distinct classes of punishment. The first is a *targeted* punishment, where discipline is administered solely to the individual who committed the transgression. The second is a *collective* punishment, where the entire group is disciplined for the actions of an individual or a smaller subgroup (Smith & Warneken, 2016). Although our common moral understanding (at least in many Western societies) as well as most existing psychological theories of moral judgment call for targeted punishment, wherein only those causally responsible for the transgression are punished (Cushman, 2015), collective punishment nonetheless occurs in a variety of contexts, including crime (Nakao & Chai, 2011), war (Darcy, 2010) and sports (Cushman et al., 2012). As such, it is important to study how targeted versus collective punishment is perceived from early in development.

This is all the more relevant because children themselves encounter collective punishment in school settings (Harber, 2021; Selman & Dray, 2006), where they spend a great deal of time learning and interacting with peers (Hofferth & Sandberg, 2001). Schools are also where children are heavily socialized with regard to rules and their enforcement, and where they likely have sufficient opportunity to consider the appropriateness and fairness of punishment, particularly by authority figures such as teachers. There are several reasons why a teacher might implement collective punishment, such as that the teacher views the whole group or class as an entity, or that the teacher hopes to increase peer pressure and children's willingness to enforce norms among their peers, to promote a sense of cooperation and cohesion among classmates, or to increase efficiency by punishing the entire group rather than working out who is responsible and how they ought to be punished (e.g., Gao et al., 2015; Pereira & van Prooijen, 2018). However, if children perceive collective punishment to be unreasonable and unfair, this may lead them to feel angry or resentful, undermine their relationship with their teacher and peers, and thus ultimately work against the teacher's goals of managing behavior, maintaining order, or fostering cohesion (Chapkovski, 2021). We must thus understand how children perceive and evaluate targeted versus collective punishment and how these forms of punishment may impact children's relationships with their teachers and peers.

Yet little work to date has examined this question. Some interview studies have found that school-aged children preferred interventions for misbehavior to be directed at the misbehaving individual rather than the group (Elliott, 1988), and criticized collective punishment when interviewed about school rules and classroom discipline (Lewis &

Lovegrove, 1987; Thornberg, 2008). In a more recent study (Smith & Warneken, 2016), 4- to 10-year-old children and adults heard vignettes in which one child in a classroom performed a norm violation (e.g., throwing an object in the classroom) and the teacher punished either the entire class or only the transgressor. The study revealed that the younger (4- and 5-year-old) children were more likely to rate the collective punishment as fair and less likely to rate the targeted punishment as fair than were the older children. These findings align with the developmental shift from a tendency to prefer equality (everyone gets the same) to an increasing focus on equity (each person gets what they deserve) (Baumard et al., 2012; Huppert et al., 2019; Hook & Cook, 1979; McGillicuddy-de Lisi et al., 1994; Shaw & Olson, 2012). In other words, younger children may be more likely than older children to view collective punishment as fair because such punishment entails an equal distribution of punishment to everyone, whereas older children may view targeted punishment as fair because only the individual deserving of punishment—the transgressor—is punished. In line with this, Smith and Warneken found that children’s reasoning about the punishments also varied with age: Younger children generally stated that they preferred collective punishment because it was an equal administration of punishment across all children, but with age, participants tended to focus on the deservingness of the punishment.

Interestingly, Smith and Warneken (2016) also found that the 4- to 5-year-olds did not rate collective punishment as fairer than targeted punishment. Thus, although there was a developmental shift in fairness judgments of collective and targeted punishments, the 4–5-year-olds did not clearly view either type of punishment as fairer. A clear preference emerged by 6 years, by which age children judged the targeted punishment as fairer than the collective punishment. Nonetheless, this study was one of the first to demonstrate that from a fairly young age, children do differentiate between collective and targeted punishment, and furthermore, that there is a developmental shift between the preschool and early school years in children’s evaluations of targeted versus collective punishments that is tied to their developing understanding of fairness.

Importantly, however, Smith and Warneken (2016) focused on children’s evaluations of punishments in third-party contexts. It remains unknown, therefore, how children view collective punishment in first-person contexts—when they are themselves involved in and affected by the punishment. Yet children do experience collective punishment first-hand in their classrooms and often criticize these punishments (Harber, 2021; Lewis & Lovegrove, 1987; Thornberg, 2008), so it is imperative to systematically examine children’s first-person responses. For instance, though Smith and Warneken found that the 4- to 5-year-olds did not judge targeted punishment as fairer than collective punishment, it is possible that children of this age would do so if they were themselves affected by the collective punishment¹. The current study aims to address this important gap by examining how children view collective punishment in first-person contexts.

¹Note that Smith and Warneken did conduct an initial study in which they attempted to create a first-person context by asking children to think of themselves in each scenario. However, children in that study still did not stand to lose anything as a result of the punishments, and the authors themselves acknowledged that the task did not create a sufficiently meaningful first-person situation. They added that an important step for future work will be to examine children’s responses in scenarios in which they are truly impacted, as we have done here.

Our study had additional important and novel aspects. First, in addition to the targeted and collective punishment scenarios, we added a third, baseline scenario in which everyone in the classroom (including the participant) was said to have committed the transgression and everyone was punished. This baseline scenario allowed us to assess whether children evaluate punishment of groups similarly regardless of context or whether they are sensitive to *why* the entire group was punished: based on the actions of an individual (collective punishment) or due to the actions of the entire group. Additionally, if we found that children viewed collective punishment as unfair, the baseline scenario could provide insight into why they did so. Specifically, if children in first-person contexts view collective punishment negatively simply because they themselves are negatively affected by it, then they should view the baseline scenario similarly negatively because they are similarly negatively affected by it (i.e., they are punished equally in both cases). Alternatively, if children view collective punishment negatively out of equity concerns (i.e., they think only those who transgressed should be punished), then they should view collective punishment as more unfair than the baseline scenario. More generally, children's responses to the collective versus baseline scenarios can help shed light on some of the mechanisms underlying children's responses to group punishment.

Second, we investigated for the first time how targeted versus collective punishment may affect children's relationships with others in the classroom: the peers and teacher. Children identify peer-peer relationships as an impactful aspect of their experience at school and as increasing their feeling of belongingness (Gowing, 2019). But the type of punishment used by authorities may undermine these relationships, particularly when the punishment impacts children directly (Chapkovski, 2021). For instance, prior work has shown that children like and act more prosocially towards those whose actions cause no harm (e.g., breaking conventional norms, or hitting but not hurting someone) versus those whose actions do cause harm (e.g., Smetana & Ball, 2017; Zelazo et al., 1996). It is possible, therefore, that children may like and be more prosocial towards peer transgressors who receive targeted punishment than if the peer's transgression leads to collective punishment because in the targeted case, the transgressor's actions do not go on to negatively impact others, including the children themselves.

Teacher-child relationships are also important as they can moderate a child's socio-emotional adjustment and shyness, and students generally view teachers as a secure base for both socio-emotional and academic support (Arbeau et al., 2010). Prior research indicates that at least by middle childhood, perceptions of fairness, both about the world in general and about the school context specifically, are positively associated with trust in authority figures (Correia & Vala, 2004; Flanagan et al., 2007; Sallay, 2004; Thomas et al., 2019), and breaches of fairness can undermine interpersonal or social trust (Brugman et al., 2016). The trust children place in their teachers might in turn affect aspects of their academic success (Imber, 1973). Thus, it is important to understand how children's views of and trust in teachers may be impacted by children's perceptions of the fairness of the teachers' punishment decisions.

A final, exploratory question in our study was whether children might take into account the role of the other (nontransgressor) children in the class, and specifically, whether their

evaluations of punishment vary based on whether or not other children were present during the transgression and could have intervened to try and prevent it. As noted above, children not only punish norm violators themselves but also expect others to do so, such as by reporting to a teacher or otherwise intervening and enforcing norms (e.g., Loke et al., 2011; Marshall, Mermin-Bunnell, & Bloom, 2020). We thus reasoned that if others are present during the transgression, children may expect them to intervene and may view collective punishment as more justified when those others could have but did not intervene. Together, then, this study aims to expand our currently limited understanding of children's perceptions of targeted versus collective punishment strategies and of the interpersonal impacts of those perceptions.

1.1 | The current study

In the present, preregistered study, children between the ages of 4 and 7 years were presented with three hypothetical classroom scenarios involving either collective punishment, targeted punishment, or a baseline scenario². We predicted, following Smith and Warneken (2016), that children's evaluations of the fairness of targeted versus collective punishments would shift between 4 and 7 years such that the younger children (4- to 5-year-olds) may be more likely to judge collective punishment as fair due to their tendency to prefer equality, whereas older children (6- to 7-year-olds) may be more likely to judge targeted punishment as fair due to their increasing focus on equity and deservingness. Regarding our exploratory question about the impact of peer presence, based on prior findings that children expect others to and think others should intervene against transgressions (Loke et al., 2011; Marshall, Mermin-Bunnell, & Bloom, 2020), we hypothesized that children may view collective punishment as fairer when peers were present but did not intervene to prevent the transgression than when peers were absent and thus could not have prevented the transgression.

Our study also examined whether the punishment type affected how children viewed the transgressor and teacher. As this was the first investigation of these questions in the literature, we did not preregister specific hypotheses for these measures. However, based on prior findings that young children dis-prefer and act less prosocially towards those whose actions cause harm (e.g., Smetana et al., 2018; Zelazo et al., 1996), we considered that children may dis-prefer and show less prosocial behavior towards the peer transgressor in the collective than the targeted punishment scenario.

Regarding views of the teacher, we considered prior findings that children's perceptions of fairness are positively associated with trust in authority figures (Correia & Vala, 2004; Flanagan et al., 2007; Thomas et al., 2019) and breaches of fairness can undermine interpersonal or social trust (Brugman et al., 2016). Further, from a young age, children act more prosocially towards fair than unfair individuals (e.g., Ziv et al., 2021). We thus

²Although we used hypothetical scenarios, we took steps to enhance ecological validity and help children immerse themselves in the scenarios, including: using a game-like board to represent a school and classrooms, using real photos of children and adults to represent the students and teachers, asking participants to draw a picture to represent themselves on the board, and moving the visual representations of the students and teachers, the transgressions, and the objects to match each scenario. Nonetheless, we acknowledge that our scenarios and tasks may have somewhat low ecological validity, particularly given prior findings that the reality of the situation seems important for children's own punishment decisions (Kenward & Öst, 2015; see also Packer & Moreno-Dulcey, 2022).

reasoned that if children judge targeted punishment as more or less fair than collective punishment, then teachers' use of targeted versus collective punishment may also affect children's trust and prosocial behavior towards the teachers.

To measure trust in the teacher, we asked children whether they would share an important secret with the teacher. Secret sharing is a common measure used to assess children's interpersonal trust, including trust in teachers (Bernath & Feshbach, 1995; Crick et al., 1996; Rotenberg et al., 2005, 2010, 2014). To measure prosocial behavior, we used a resource distribution task in which children were asked to distribute three flowers between the teacher and a novel adult. (A similar task was also used to assess prosocial behavior towards the peer transgressor). This distribution task has been widely used in the developmental literature (e.g. Marlow et al., 2023; McElroy et al., 2023; Misch et al., 2014; Vaish et al., 2011) as it elicits children's preferences behaviorally, with minimal reliance on verbalizations, and presents simple, dichotomous response options (e.g., teacher vs. stranger). Though children in prior studies did not distribute resources to teachers specifically, they did distribute resources between adults. Further, although flowers may not hold particular meaning in this context, they represent a simple, familiar resource that children can straightforwardly understand as positive or rewarding; they are thus a practical resource for a distribution task designed to nonverbally assess children's prosocial preference. For the teacher trust and distribution tasks, following Smith and Warneken's (2016) findings that by age 6, children viewed collective punishment to be less fair than targeted punishment, we tentatively hypothesized that the older children in our study may show less trust and prosocial behavior towards a teacher who administered collective punishment versus targeted punishment.

2 | METHOD

The study design, procedure, hypotheses, sample size (based on power analyses using effect sizes from Smith & Warneken, 2016), and exclusion criteria were preregistered using the [AsPredicted.org](https://osf.io/w9prs/?view_only=ba189c1a07274e4cb0195f809fc037d1) template through the Open Science Framework (https://osf.io/w9prs/?view_only=ba189c1a07274e4cb0195f809fc037d1). Note that the preregistration was completed after approximately half the data had been collected, although no data had been examined or analyzed at that time.

2.1 | Participants

A total of 72 children participated in the present study. As preregistered based on Smith and Warneken (2016), participants were grouped according to age: The younger group consisted of 36 4- to 5-year-olds (*age range*: 48–71 months, $M_{age} = 59.78$ months, $SD = 6.28$; $n = 18$ females) and the older group consisted of 36 6- to 7-year-olds (*age range*: 75–95 months, $M_{age} = 85.72$ months, $SD = 5.20$; $n = 18$ females). An additional 13 children were tested but excluded due to experimenter error ($n = 1$), equipment failure ($n = 3$), or the participant not completing the study ($n = 9$). All children were fluent in English and were recruited from a mid-Atlantic university town. The demographics were representative of the city's community, with a predominantly White sample ($n = 56$ White; $n = 10$ Mixed race; $n = 2$ Black; $n = 1$ Asian; $n = 1$ Native Hawaiian or Pacific Islander; $n = 2$ did not disclose) and highly educated ($n = 48$ Postgraduate; $n = 22$ college educated; $n = 2$ did

not disclose). This study was approved by the authors' institutional IRB. All parents gave written informed consent and children over the age of 7 gave informed assent in accordance with the Declaration of Helsinki. All children received a small prize for their participation.

2.2 | Design and materials

There were two within-subjects factors: punishment type (Targeted, Collective, Baseline) and transgression situation (paint, crayon, poster), and two between-subjects factors: age (4–5-year-olds, 6–7-year-olds) and peer presence (Peers Present, Peers Absent). The order of punishment type and transgression situations were counterbalanced across children. The order of questions was randomized across children. Half of each age group received the Peers Present condition and the other half received Peers Absent (for practical reasons, most of the Peers Present conditions were conducted first)³.

During the experiment, children were seated at a table across from the experimenter. For the scale training, we used a bar graph with four bars that increased in height (see Figure S1 in OSM) as well as four photographs of jugs filled with different amounts of water (none, a little bit, a medium amount, and a lot). For the main study, we used a cardboard mat (3.75 × 1.5 ft), with five different colored sections (each around 1.5 × .75 ft) representing five different classrooms (see Figure S2). During each of the three phases of the experiment (one for each punishment type), photos (2 × 2 in) of children of a similar age were used to represent peers in the classroom, and photos of adults (2 × 2 in) were used to represent teachers. Photo stimuli of the children and adults were taken from standardized datasets and all photos featured individuals with a neutral expression (Ebner et al., 2010; LoBue & Thrasher, 2014). During each phase, the following items were placed onto a mat to represent the individuals in the class: (1) six photos of peers (with at least two White male and two White female peers per class), (2) one drawing to represent the participant, and (3) one photo of the teacher. In each phase, a picture of the transgression (e.g., a scribble on the wall) and a picture of the evidence (e.g., crayon) were paired with the transgressor(s) to further illustrate the transgression event (see Figure 1).

2.3 | Procedure

All children were tested by the same female experimenter (E). E first introduced and had the children practice with the scale they would later use to answer questions (McLoughlin et al., 2018). The scale had four points corresponding to 4 values (1 = not at all, 2 = a little bit, 3 = a medium amount, 4 = a lot), and could also be used to indicate a negative amount (e.g., how unfair; -1 = not at all, -2 = a little, -3 = a medium amount, -4 = a lot). Each child was then introduced to the “school” set up. The experimenter brought out a mat and told the participant that this was a school. Participants were told to pretend that they were a

³We pre-registered the between-subjects Peers Present/Absent conditions as two separate studies. This was for logistical reasons: We planned for the Peers Present condition to constitute an honor's thesis project and thus to be conducted first, and the Peers Absent condition to be conducted subsequently. Our pre-registered power analyses were therefore conducted to detect medium-to-large effect sizes for differences *within* the Peers Present or Peers Absent conditions, not *between* these conditions. However, as noted in our pre-registration (see “Comparing Study 1 and Study 2 results”), we always intended to analyze the Peers Present/Absent factor along with the other factors within the same analyses. Ultimately, to avoid issues of multiple comparisons, we chose to only conduct analyses with all factors included rather than first analyzing the Peers Present/Absent conditions separately and then again within the same analyses.

student at this school and that they would be “going to different classes, meeting different teachers, and meeting different students.” They were then informed about a school-wide rule that “everyone gets a sticker for coming to class, but these stickers can be taken away if someone breaks a rule.”

The experimenter brought out cards with photos of different “classmates” and placed them in one section of the mat representing the first classroom. In each class there were three male and three female images. Before starting the first story, participants were asked to draw a picture of themselves on a blank card. The drawing was then placed in the first classroom among the images of the peers to show the child that s/he was now a part of the class. This drawing was used to represent the child throughout the study.

Each child was presented with three classroom scenarios (Collective, Targeted, and Baseline; within-subjects). At the beginning of each scenario, the experimenter distributed a sticker to each student by placing one above every card, explaining that everyone got a sticker for coming to class. Next an image of an adult (always female) representing the teacher was introduced and her card was placed in the classroom. E now either told participants that the teacher and the entire class left the classroom (Peers Absent condition) or that the teacher left the classroom (Peers Present condition); this was a between-subjects factor.

Then, E described the transgression. In the Collective and Targeted scenarios, a single student (White and gender-matched to the participant) committed one of three transgressions (poured paint on the floor, drew on the wall with crayon, or tore down a poster). E placed pictures representing the transgression (e.g., image of a crayon and a scribble) next to the image of the student who had committed the transgression. E then explained that the student committed the transgression “even though it’s against the rules.” Next, E either explained that “No one sees him/her [commit the transgression] and no one has a chance to stop him/her, because everyone is in the other room” (Peers Absent condition) or “You and your classmates see him/her [commit the transgression] and no one tries to stop him/her, even though they could have because everyone is in the classroom” (Peers Present condition). In the Baseline scenario, E said that everyone in the classroom committed the transgression, and she placed a picture representing the transgression next to each student’s image, including the participant’s. E further stated that “Everyone is in the classroom, so everyone sees each other [committing the transgression].”

The teacher was then described as returning to the classroom, seeing the transgression, and finding evidence on the transgressor(s) (e.g., crayon in transgressor’s hand). Children were then reminded of the school’s rule that “stickers can be taken away if someone breaks a rule.” E then said that the teacher either took away everyone’s sticker including the participant’s sticker (Collective and Baseline conditions) or just the single transgressor’s sticker (Targeted condition). After announcing the teacher’s decision, E physically removed the respective sticker(s). (Scripts for all scenarios are provided in the online supplemental materials [OSM].)

After each scenario, E asked children comprehension questions to ensure they remembered who committed the transgression, who saw the transgression occur, and whose sticker(s) were taken away (see OSM for details). If children were confused about any of the events in the story, E re-explained what had occurred in the classroom and asked the comprehension questions again. Participants would be excluded if they failed to answer comprehension questions correctly three times in a row. However, no participant failed to answer correctly three times in a row and thus no participants were excluded for this reason.

Children were then asked test questions to assess how they evaluated the different types of punishment and how these punishments influenced their judgments of others in the classroom. Children were asked open-ended justification questions after each test question or behavioral response to gain additional insights into their reasoning. Before each set of questions, children were reminded about who committed the transgression, who was or was not in the classroom, and whose sticker was taken away. The question sets were as follows:

2.3.1 | Fairness of the punishment

1. “Do you think it is fair that this student’s/everyone’s sticker was taken away?”
2. “How fair/unfair?” (using the bar graph that was introduced at the beginning of the study)
 - a. If children answered “Fair”: “Is it not fair at all, a little fair, fair, or really fair?”
 - b. If children answered “Unfair”: “Is it okay, a little unfair, unfair, or really unfair?”
3. Justification question: “Why do you think it was [quantity selected] fair/unfair?”

2.3.2 | Transgressor

1. “How much do you like this student? Not at all, a little bit, a medium amount, or a lot? Point to the bar that shows how much you like her/him.” (using bar graph). (In Baseline, this question was asked about a White, gender-matched child chosen at random.)
2. Justification: “Why do you like this person [quantity chosen]?”

2.3.3 | Teacher

1. “If you had a really important secret you didn’t want people to know, would you tell [the teacher]?”
2. “Why would you (not) share your secret with [the teacher]?”

Finally, we conducted two resource distribution tasks to obtain a behavioral measure of children’s responses to the peer transgressor and the teacher. Specifically, after the Transgressor question set, children were given 3 cloth flowers to share with the transgressor and another classmate, and after the Teacher question set, children were given 3 cloth flowers to share with the teacher and a neutral adult (a stranger). In both cases, they were told that E would see the two individuals later and could bring them something from the

child; this was done to add value to the flowers and make them seem like rewards or gifts that children could distribute as they wanted. After children distributed the flowers, E asked, “Why did you give him/her more flowers?” (Detailed scripts for all questions are provided in the OSM).

The order of the three scenarios (Collective, Targeted and Baseline) and the order of the transgressions (paint, crayon, and poster) were counterbalanced across children. The order of the question sets (fairness of punishment, transgressor, and teacher) was also randomized across children, though questions within each set were presented in a fixed order.

2.4 | Coding and reliability

Children’s responses to all questions were coded offline from video recordings of the experimental session by the primary experimenter (the first author). In order to assess reliability, a random selection of 25% of the videos ($n = 18$) were coded by two secondary coders naive to research hypotheses (9 videos per coder).

Responses to test questions were coded as follows: whether children thought the punishment was fair (yes = 1, no = 0), how fair/unfair (from $-3 =$ really unfair, to $3 =$ really fair⁴), how much they liked the transgressor (from 1 = not at all, to 4 = a lot), and if they would share their secret with a teacher (yes = 1, no = 0). We also coded how many flowers children shared with the transgressor versus another peer and how many flowers they shared with the teacher versus another adult.

The coding schemes for children’s justifications (partially adapted from Smith & Warneken, 2016) varied based on the set of test questions. For questions about the fairness of the punishment, participants’ justifications were coded for referencing deservingness explicitly, referencing deservingness implicitly, sameness (desire for everyone to receive the same treatment), group responsibility, and emotion. Participants’ justifications for the transgressor ratings and flower distribution were coded for referencing the transgression or the punishment. Participants’ justifications for why students would (not) share their secret with the teacher were coded for referencing the transgression, the punishment, trust, conventions of a secret, and role of a teacher/adult. Justifications for teacher flower distribution were coded for referencing the transgression, punishment, or familiarity (being familiar with or “knowing” the adult; e.g., “She’s my teacher”). For all questions, the justification codes were not mutually exclusive. Detailed information about the justification coding schemes is provided in Tables S1–S4 in OSM.

Reliability for the test questions ($kappa = 1.00$), behavioral responses ($kappa = .99$) and justification coding ($kappa = .88$) were excellent.

⁴This 7-point scale was created by combining the “How fair?” scale (0 = ‘not fair at all’ to 3 = ‘really fair’) and “How unfair?” scale (0 = ‘okay’ to $-3 =$ ‘really unfair’). That is, the first point on both scales was scored 0 to have both scales begin at the same neutral point, and the combined scale then extended down to -3 for unfairness and up to 3 for fairness.

2.5 | Analysis plan

Analyses were conducted in line with our preregistration. One exception was that given the structure of the fairness and trust variables (yes/no), we updated our analysis plans and performed generalized mixed effects models. In addition, for all significant age x condition interaction effects, we conducted posthoc analyses (not preregistered) to compare younger and older children for each of the conditions. Bonferroni correction for multiple comparisons was applied for all posthoc analyses.

3 | RESULTS

3.1 | Fairness of the punishment

Analyses of children's responses to whether they thought the punishment was fair (yes/no) and children's fairness ratings ('how fair/unfair?') revealed very similar results. We thus only report results from children's fairness ratings here; detailed results for the binary (yes/no) question are reported in the OSM.

3.1.1 | Fairness ratings—A repeated-measures ANOVA was conducted with ratings of fairness on a 7-point scale (from -3 = really unfair, to 3 = really fair), with condition as a within-subjects factor and age group and peer presence as between-subjects factors. Significant effects were followed up with posthoc pairwise comparisons with Bonferroni adjusted p -values. This analysis revealed a main effect of condition, $F(2, 136) = 31.73, p < .001, \eta^2 = .32$, as well as a main effect of age, $F(1, 68) = 7.82, p = .007, \eta^2 = .10$. These main effects were qualified by a significant condition by age interaction, $F(2, 136) = 10.28, p < .001, \eta^2 = .13$. To follow up on this interaction, separate repeated-measures ANOVAs were conducted for each age group.

Among the older children, there was a significant difference between conditions, $F(2,70) = 50.82, p < .001, \eta^2 = .59$. Specifically, their average ratings of the collective punishment were in the 'a little unfair' to 'unfair' range ($M = -1.56, SD = 2.09$) whereas their average ratings of the targeted and baseline punishments were in the "fair" range (Targeted: $M = 2.08, SD = 1.63$; Baseline: $M = 2.08, SD = 1.75$). Pairwise comparisons revealed that, as predicted, the older children rated the punishment in the Collective condition as significantly less fair than in the Targeted ($p < .001, d = 1.29$) and Baseline conditions ($p < .001, d = 1.34$). Their fairness ratings of the Targeted and Baseline conditions were not significantly different, $p = 1.00$. Among the younger children, however, the ANOVA revealed no statistically significant difference in fairness ratings between the Collective ($M = -.83, SD = 2.44$), Targeted ($M = .11, SD = 2.83$), and Baseline ($M = .22, SD = 2.65$) conditions, $p = .10$, which all fell in the neutral to "a little unfair" range (see Figure 2). There were no other significant effects (all $ps > .27$).

Additional independent samples t-tests (not preregistered) revealed a significant difference in ratings between the younger and older children for the Targeted ($t(55.80) = -3.62, p < .001, d = .85$) and Baseline ($t(60.55) = -3.52, p < .001, d = .82$) conditions such that the older children rated the punishment in the Targeted and Baseline conditions as more fair

compared with the younger children. However, there was no difference in fairness ratings between the younger and older children for the Collective condition ($p = .182$).

3.1.2 | Justifications for fairness ratings—At both ages, the most common form of justification that children provided for their fairness ratings implicitly referenced deservingness, that is, mentioned actions of the transgressor or punishment related to deservingness (e.g., “Because he drew on the wall”). The older children’s justifications were especially likely to mention implicit deservingness, with 72% of the 6–7-year-olds and 44% of 4–5-year-olds doing so. A few children at each age also explicitly referenced deservingness, that is, specifically mentioned how the individual action did or did not deserve the punishment given (e.g., “...it was his fault so it’s kind of not fair, we didn’t do anything bad”; see Table S5 in OSM for details).

3.2 | Transgressor

3.2.1 | Ratings of liking—A repeated-measures ANOVA was conducted with ratings of liking the peer transgressor (from 1 = not at all, to 4 = a lot) in each condition as a within-subjects factor and age group and peer presence as between-subjects factors. This analysis revealed a main effect of condition, $F(2, 134) = 17.26, p < .001, \eta^2 = .21$, such that across ages, children reported liking the transgressor more in the Baseline condition ($M = 2.42, SD = 1.08$) than in both the Collective condition ($M = 1.75, SD = .89, p < .001, d = .93$), and the Targeted condition ($M = 1.83, SD = .96, p < .001, d = .88$). There was no statistically significant difference between the Collective and Targeted conditions, $p = 1.00$ (see Figure 3). There were no main effects or interaction effects in the omnibus model for age (all $ps > .20$) or peer presence (all $ps > .36$). Thus, overall, children reported liking the transgressor more in the Baseline compared to both the Collective and Targeted conditions.

3.2.2 | Justifications for transgressor liking—About three-quarters of the children at both ages in the Targeted and Collective conditions referenced the transgression when justifying their reported liking of the transgressor; the proportions were lower in the Baseline condition, where closer to half of the children at each age referenced the transgression. Few children in either age group referenced the punishment given by the teacher in any of the conditions (see Table S6 in OSM for details).

3.2.3 | Flower distribution—A repeated-measures ANOVA was conducted with conditions as a within-subjects factor and age group and peer presence as between-subjects factors. This revealed a main effect of condition on the number of flowers shared with the transgressor, $F(2, 136) = 20.32, p < .001, \eta^2 = .23$. Consistent with their ratings of liking, across ages, children shared more flowers with the transgressor in the Baseline condition ($M = 1.43, SD = .55$) compared to both the Collective condition ($M = 1.04, SD = .57, p < .001, d = .68$), and the Targeted condition ($M = .90, SD = .56, p < .001, d = .17$). There was no statistically significant difference between the Collective and Targeted conditions, $p = .165$ (see Figure 3). There were no other significant main effects or interactions (all $ps > .18$).

3.2.4 | Justifications for transgressor flower distribution—Most children at both ages in the Targeted and Collective conditions referenced the transgression (e.g., “She drew

on the wall”) when justifying their flower distribution between the transgressor and a neutral peer; the proportions were lower in the Baseline condition, where most children at both ages provided a justification categorized as “other.” Few children at either age in any of the conditions referenced the punishment that the transgressor had received (see Table S7 in OSM for details).

3.3 | Evaluations of teacher

3.3.1 | Secret sharing—A generalized mixed effects model with a binary logistic link function was conducted with condition as a within-subjects factor and age group and peer presence as between-subjects factors predicting children’s willingness to entrust the teacher with a secret (yes = 1, no = 0). This analysis revealed an interaction between peer presence and age, $F(1,204) = 4.64, p = .032$ (OR = 4.09 CI [.56–29.67]). However, in follow-up chi-square analyses, there were no significant differences between age groups in either the Peers Absent condition ($p = .712$) or the Peers Present condition ($p = .337$). There were no other significant main effects or interactions (all $ps > .38$).

3.3.2 | Justifications for secret sharing—At both ages and in all conditions, a substantial minority of children referenced conventions of secrets in their justifications, that is, they cited their understanding of how one handles secrets (e.g., “Because secrets are secrets; if you have a secret, you shouldn’t share it”). The rest of children’s justifications fell fairly evenly into the other four categories, including references to the transgression, the punishment, the role of the teacher (e.g., “Because the teacher needs to know”), and trust (e.g., “Because I don’t think I would trust her”) (see Table S8 in OSM for more details).

3.3.3 | Flower distribution—A repeated-measures ANOVA was conducted with the number of flowers shared with the teacher (vs. a neutral adult) in each condition as a within-subjects factor and age group and peer presence as between-subjects factors. This revealed a main effect of condition, $F(2,136) = 4.96, p = .008, \eta^2 = .07$: Children gave significantly fewer flowers to the teacher in the Collective condition ($M = 1.53, SD = .69$) compared to both the Baseline condition ($M = 1.81, SD = .62, p = .035, d = .36$), and the Targeted condition ($M = 1.79, SD = .56, p = .026, d = .38$). There was no significant difference between the Baseline and Targeted conditions, $p = 1.00$ (see Figure 4). There were no other significant main effects or interaction effects, including no significant effects of age (all $ps > .62$) or peer presence (all $ps > .09$). Thus, children at both ages distributed less generously to a teacher who meted out collective punishment than one who used targeted punishment and one who punished the whole group for the whole group’s transgression (Baseline).

3.3.4 | Justifications for teacher flower distribution—When justifying their flower distribution to the teacher, a substantial minority of older children referenced the punishment in the Targeted condition (e.g., “Because she didn’t take everyone’s sticker away”) and referenced familiarity with the teacher in the Collective and Baseline conditions (e.g., “Because she’s the teacher of our class”). Younger children generally provided justifications categorized as “other” (see Table S9 in OSM for details).

4 | DISCUSSION

Targeted and collective punishments are both widely used forms of punishment, including in the classroom (e.g., Cushman et al., 2012; Darcy, 2010; Selman & Dray, 2006). A large body of work has examined how collective punishment is used (Hechter, 1990), its effectiveness (Chapkovski, 2021), and how people view collective punishment (Heckathorn, 1990). However, little is known to date about how children evaluate collective punishment when they are themselves impacted by it, such as in a school setting. Consistent with prior work conducted in a third-party context (Smith & Warneken, 2016), our study using a first-person collective punishment context revealed that children's judgments of targeted versus collective punishments shifted between the preschool and early school ages. We also showed for the first time that the type of punishment affected children's evaluations of and affiliations with the peer transgressor as well as the adult who meted out the punishment.

4.1 | Evaluation of the punishment

In support of our central hypotheses, we found that children's evaluations of the fairness of different punishment types changed across early childhood. Specifically, as predicted and consistent with the developmental shift documented by Smith and Warneken (2016), the older children (6–7 years) in our study clearly differentiated between the fairness of targeted versus collective punishment, evaluating targeted punishment as fair but collective punishment as unfair. On the other hand, the younger children (4–5 years) did not clearly evaluate either collective or targeted punishment as fairer. Moreover, the older children rated targeted punishments as fairer compared to the younger children (though they did not rate collective punishment as less fair than the younger children). Importantly, using a novel, baseline scenario in which the entire group was said to have transgressed and to be punished, we found that the older children did not simply view all group punishment to be unfair but rather did so specifically in the collective case, when the group was punished for the transgression of one individual. Together, our findings and those of Smith and Warneken provide converging evidence that between the preschool and early school years, children (at least in the US) develop an increasingly sophisticated understanding of punishment and its fair application such that by 6–7 years of age, they begin to perceive punishment that targets the transgressor as substantially fairer than punishment delivered to the transgressor's entire group.

This shift is in line with a well-documented shift at these ages in children's reasoning about equality and equity (see also Smith & Warneken, 2016). Prior work demonstrates that by around 5–6 years of age, children shift from preferring equal distributions to equitable distributions that integrate more complex aspects such as need and merit (Damon, 1975; Fraser et al., 2007; Hook & Cook, 1979; McGillicuddy-de Lisi et al., 1994). Remarkably, this developmental shift from equality-based to equity-based reasoning emerges across a wide variety of contexts and cultures (Huppert et al., 2019). This shift may partially explain why the older children in our study viewed targeted punishment as fair, and significantly more fair than collective punishment: because targeted punishment is the equitable consequence for an individual's transgression.

In line with Smith and Warneken (2016), we found an age effect wherein the older children judged targeted punishment to be fairer than the younger children, consistent with the above-noted development shift towards prioritizing equity over equality. However, whereas Smith and Warneken reported a further age effect such that older children judged collective punishment to be less fair than did the younger children, the younger and older children in our study did not differ in this regard, with both age groups judging collective punishment to be somewhat unfair. This lack of age effect for the collective condition could be due in part to our first-person design, wherein the participating child's own sticker was taken away in the collective scenario (as opposed to Smith and Warneken's third-party design where the child was not personally affected). This may have led children in our study to have a somewhat negative response to the collective punishment, which may have counteracted their otherwise positive view of collective punishment (though children's fairness assessments likely did not rely solely on their personal negative impact given that they did not judge targeted punishment—which did not personally affect them—as fairer than collective punishment).

This proposal aligns well with prior findings that although young children understand that resources should be shared equally, they nonetheless continue to accept and actively engage in unequal distributions that are advantageous to themselves and they do not reject distributions that advantage them until a few years later (around 7–8 years; e.g., Fehr et al., 2008; Kogut, 2012; Smith et al., 2013). Similarly, although 4- to 6-year-olds prefer a procedurally just method for resource allocation in third-party contexts, they do not show this preference in first-party contexts, that is, when they stand to gain from distributive injustice; the latter preference is only evident by 7–8 years (Dunham et al., 2018). It seems plausible, therefore, that the younger children in our study were not able to objectively assess and view collective punishment as relatively fair due in part to their conflicting feelings about being punished themselves. Though this involvement of their own interests may seem problematic for assessing children's views on collective punishment, we believe that because children may experience collective punishment themselves, examining their perceptions of such punishment in first-person situations is important for a fuller understanding of the phenomenon.

Consistent with Smith and Warneken (2016), the younger children in our study did not differ in their fairness evaluations of the targeted versus collective punishments. There are a few possible reasons for this pattern among the younger children. One is that children at these younger ages simply do not approve of punishment of any kind. In line with this, prior work found that a majority of younger children (5–6 years) preferred to live in a world with no punishment for theft whereas older children preferred a world with punishment (Bregant et al., 2016). Preschool-age children may thus not view punishment in general as fair or appropriate.

A second possibility is that within the younger age group, children's responses to different forms of punishment were heterogeneous and thus, as a group, did not reveal significant differences. Indeed, a closer look at the younger children's fairness ratings data supports this idea. As Figure 2 shows, the younger (but not older) group's data appear to be largely bimodal, with a substantial proportion responding that each type of punishment was fair and

a similar proportion responding that each type of punishment was unfair. That is, the means derived from these data obscure the fact that the younger children were quite evenly split in their fairness evaluations of each type of punishment. An interesting explanation for this bimodal pattern is that this is a transitional stage wherein some of the younger children have started to shift from an equality-based to a more equity-based perspective (see also Baumard et al., 2012, for evidence of equity understanding by as early as 3–4 years), and thus, like the older age group, view targeted punishment as fair and collective punishment as unfair; other children in the younger group may still primarily hold an equality-based perspective and thus view targeted punishment as unfair and collective punishment as fair. This split at the younger age may help explain why as a group, the 4- to 5-year-olds in our study (and Smith & Warneken, 2016) did not rate either collective or targeted punishment as fairer. This will be an important direction to explore in future work.

More generally, our results contribute to the broader literature on children's reasoning about punishment. It is well documented that children are willing to administer punishment themselves (e.g., McAuliffe et al., 2015; Riedl et al., 2015), have a preference for punishment over no intervention (Killen et al., 1994), and expect that others will punish (Geraci, 2021; Geraci & Surian, 2021; Loke et al., 2011; Marshall et al., 2020), and these tendencies emerge in infancy (Geraci & Surian, 2023; Kanakogi et al., 2022). The present study adds to this work and joins other recent findings in demonstrating that by the preschool and early school years, children have an increasingly nuanced understanding of punishment, including considering the fairness and the type of punishment being administered (e.g., Brinker et al., 2003; Evans et al., 2001; Smith & Warneken, 2016).

4.2 | Responses to peer transgressor

Our study addressed several additional, novel questions related to the impact of different forms of punishment on children's social lives. First, we explored whether the type of punishment affected children's relationships with the peer who committed the transgression. We had tentatively predicted that children would like the transgressor less in the Collective scenario compared to the Targeted scenario because the transgressor's actions in the Collective case ultimately cause more harm: by resulting in the removal of everyone's (including the child's own) stickers. This prediction was based on prior findings that children dis-prefer and act less prosocially towards those whose actions cause harm than those whose actions cause no harm (e.g., Smetana & Ball, 2018; Zelazo et al., 1996). However, we found that at both ages, children in both the Collective and Targeted conditions showed a similarly low level of liking the peer transgressor (between 'not at all' and 'a little bit') and distributed fewer flowers (~1 out of 3 flowers) to the transgressor than in Baseline.

Although this appears to contradict prior findings, we note that unlike in prior work, in which the transgression directly caused (or did not cause) harm (e.g., Smetana & Ball, 2018; Zelazo et al., 1996), the harm in our study (taking away the child's and other classmates' stickers) was caused (or not caused) indirectly, by the teacher's decision to use collective or targeted punishment. In other words, the transgressor's original actions were similarly negative in both the Targeted and Collective conditions, and the children seemed to evaluate the transgressors on those terms, separate from the ensuing punishment chosen by the

teacher. This aligns with Bregant et al.'s (2016) findings that children liked a thief who was punished as much as a thief who was not punished, indicating that the punishment did not change children's regard for the transgressor. Children's justifications support this interpretation, as the majority of children at both ages mentioned the transgression to justify both their evaluations of and flower distribution to the transgressor, whereas very few children mentioned punishment.

Interestingly, children reported liking the peer significantly more in the Baseline condition—when the entire class (including the participant) was said to have transgressed. This may be because in this case, the entire group had transgressed and thus each individual was viewed as less blameworthy. Further, although children were evaluating a peer transgressor, children in the Baseline condition were themselves in the role of transgressors as well. Children may thus have viewed the transgressor as more similar to them or perhaps even as an in-group member, resulting in more positive ratings of the transgressor in the Baseline than the other conditions (Aboud, 2003; Baron & Dunham, 2015; Dunham et al., 2011; Kinzler et al., 2007).

4.3 | Responses to punisher (teacher)

We also investigated for the first time how the type of punishment may affect children's evaluations of the punisher (i.e., the teacher). Across ages, children gave more flowers to a teacher who administered a targeted punishment versus a collective punishment. The fact that this effect was also evident at the younger age suggests that even (some) 4–5-year-olds may have begun to differentiate (perhaps in an implicit way) between the targeted and collective punishments and to evaluate targeted punishment as preferable to collective punishment. Impressively, in contrast to the peer findings, the older children's justifications for their distribution to the teacher were punishment-focused, that is, children explained these distributions based on the kind of punishment the teacher had administered. Together, this set of findings suggests that teachers' choice of punishment may indeed influence children's assessments of their teachers, and specifically, a teacher's use of collective punishment may reduce children's motivation to affiliate with the teacher and could eventually erode a student-teacher relationship.

However, our second measure of teacher evaluation did not reveal a difference, as children did not report greater willingness to trust (by sharing a secret with) the teacher who administered targeted versus collective punishment. One reason may be that it takes a series of encounters with a particular teacher to trust or distrust them. Therefore, perhaps the one punishment decision was not enough to sway children's trust. An alternative is that our secret-sharing measure did not clearly capture children's trust. Though prior work has used secret sharing with teachers to measure interpersonal trust, it has done so by presenting hypothetical scenarios in which a student tells their teacher a secret and participants judge how likely the teacher is to keep the secret (e.g., Rotenberg et al., 2005), whereas we asked children whether they would tell the teacher an important secret. This may impact the results; for instance, asking children whether they would share a secret with the teacher may have led children to focus on the conventions surrounding secret keeping in general. We found some evidence for this in children's justifications: A substantial minority of children

across ages provided justifications focused on the conventions of secrets and the typical information one shares with adults. Overall, it is important to interpret our teacher evaluation findings with caution; though intriguing, these findings warrant further research, particularly using an applied approach in real classrooms.

Taken together, our findings regarding children's responses to the peer transgressor versus teacher compellingly show that although children were directly negatively impacted by the collective punishment, they did not respond to this punishment with generalized negative affect. Rather, the teacher's decision to impose a collective punishment seemed to impact children's perception of and responses to the teacher but (appropriately) did not generalize to their perception of and responses to the transgressor. In other words, at both ages, children's evaluations focused on what was most relevant to the individual being evaluated.

Along related lines, we included a key Baseline scenario in which the entire group (including the participant) had committed the transgression and the entire group was punished. This scenario helped further clarify potential reasons why children prefer particular forms of punishment. Specifically, older children rated the collective punishment as less fair compared to the targeted punishment, but there was no difference in their fairness evaluations of the Targeted versus Baseline conditions. Similarly, at both ages, children's evaluations of the transgressor and teacher revealed differences between the Collective and Baseline conditions. These findings suggest that children were evaluating the fairness of the punishments based on the specifics of the situation and not solely on whether they were themselves negatively impacted (i.e., their sticker was taken away). It also suggests that children do not always perceive group punishment to be unfair; rather, participants viewed group punishment as relatively fair if the entire group was responsible for the transgression, including the participant. Overall, our findings show that from the preschool to early school years, children's evaluations of transgressions and punishments rely on a fairly nuanced understanding of the transgressions, transgressors, punishments, and punishers involved.

4.4 | Additional exploratory questions, limitations, and future directions

A further, exploratory question was whether the presence of bystanders who were aware of the transgression (and could have intervened) increases the perceived fairness of collective punishment. We thus manipulated whether the entire class was present during the transgression but found virtually no effects of peer presence versus absence on any measure. Though it is possible that the presence or absence of peer observers truly did not matter for children's evaluations, there are several leaner interpretations. One is that there was simply too much information in the scenarios for children to keep track of. Moreover, although the comprehension questions after each scenario ensured that children knew who did (not) observe the transgression, they did not ensure that children comprehended that the observers could have intervened in the transgression; it is possible that children did not draw this key inference and their evaluations were thus not impacted by peer presence/absence. Our study may also have been underpowered to detect this effect. Our a priori (preregistered) power analyses were conducted to assess effects within the Peers Present and Peers Absent conditions, not between these conditions. We did subsequently conduct a sensitivity analysis for detecting between-condition differences, which indicated that given our design and

sample size, we would have the power to detect a between-conditions difference with a large effect size (see OSM for details). It is thus possible that peer presence had a small or medium effect that our study was not sufficiently powered to detect. Though our findings are inconclusive, we do think that the question of the impact of bystanders who could have intervened (as well as of other external factors) on children's views of punishment is worth exploring further. To do so, future work should include a larger sample, employ a more powerful manipulation, and clearly inform children about whether the observers could have intervened.

An important related question our study points to is whether children might view targeted and collective punishments as serving different functions, such as retribution or deterrence. Recent work has shown that children consider and use both of these motivations in their evaluations and administration of punishment (e.g., Arini et al., 2023; Marshall & McAuliffe, 2022; Marshall et al., 2021). To our knowledge, however, prior work has not considered whether children view different kinds of punishment as serving these functions to a greater or lesser degree (though both punishment types likely serve both functions to some extent). It seems plausible, for instance, that targeted punishment is perceived as better suited for retribution because it focuses on punishing the transgressor, whereas collective punishment might align better with deterrence as it may more effectively deter both the transgressor and others from future transgressions. Some suggestive evidence for this comes from Bregant et al. (2016), who found that 4- to 8-year-olds expected targeted punishment to serve as a deterrent for the individual transgressor but not for bystanders or victims. However, Bregant et al. only assessed targeted punishment, leaving unclear whether children hold different beliefs about the motivations or effects of collective punishment. Our study was not designed to tease apart this issue, but we do believe this is an exciting new direction for this research. For instance, future studies could ask children why they think the teacher used targeted versus collective punishment or what they believe the teacher *should* do when a child transgresses (administer targeted or collective punishment).

An important limitation of our study is that it was conducted in a predominantly White and educated population in the US. It therefore does not address the possible impact of culture on children's views of various forms of punishment. One useful distinction here may be between cultures that are more heavily individualistic (represented in the present study) versus collectivistic. Cultures that foster individualism generally focus on personal responsibility and freedom of choice (Waterman, 1984). In cultures that have a collectivist focus, however, priority is often given to maintaining harmonious interdependence with others (Markus & Kitayama, 1991). This latter emphasis fosters the belief that one must bear responsibility for others' actions. Adults and children in collectivistic cultures also have somewhat distinct moral evaluations, such as having broader and more stringent expectations of social responsibility (Miller et al., 1990). Further, Huppert et al. (2019) recently found that the levels of individualism versus collectivism in a culture impacted the age and extent to which children preferred equality or equity: Children in more individualistic cultures were found overall to endorse equitable distributions at a younger age than children in more collectivist cultures. And of direct relevance to the present work, collective punishment is widely used in collectivistic cultures such as Japan (Hechter & Kanazawa, 1993). Together, the prior work raises the very interesting possibility that

children raised in more collectivist environments may be less likely to judge collective punishment as unfair or less fair than targeted punishment. Additionally, teachers' use of collective punishment may not have a negative impact—and may even have a positive impact—on children's evaluations of and relationships with the teachers, who may be viewed as upholding cultural expectations. Thus, extending this study to other cultures, particularly those with more collectivistic values, will be a critical next step for this work.

Another limitation is the hypothetical nature of our scenarios. Though we did take several steps to increase the ecological validity of the scenarios, we acknowledge that using pretend situations can impact children's behaviors, including their willingness to administer punishment (Kenward & Östh, 2015; Packer & Moreno-Dulcey, 2022; but see Stengelin et al., 2023). Any age differences may also partially stem from differences in younger versus older children's ability to understand and accept the hypothetical situations presented in the study rather than from the experimental variables. Moreover, although our flower distribution and secret sharing tasks were adapted from prior work and included as an initial exploration of how punishment may impact children's relationships, these tasks also likely had low ecological validity (e.g., asking children to consider sharing a hypothetical secret with a hypothetical adult). Thus, although we believe that the current paradigm provides important insights into children's responses to targeted versus collective punishment, it will be important to increase the realism of the scenarios and tasks in future work to assess the validity and generalizability of our findings.

Future studies would also benefit from expanding the scope of this study, such as by including other transgression types (e.g., ones involving a direct victim, fairness transgressions, or moral transgressions; Smetana et al., 2014), varying the severity and intentionality of the transgression (Cushman et al., 2013; Loke et al., 2011), having the transgressor express remorse (Darby & Schlenker, 1989; Oostenbroek & Vaish, 2019), and varying the group membership of the transgressor (Dunham et al., 2011; Schmidt et al., 2012). It will also be important to take this research out of a controlled lab setting, such as by conducting naturalistic observations in actual classrooms, which offer a rich network of norms, existing relationships, and social dynamics.

In conclusion, this study shows that even quite young children are able to think critically about various punishment strategies and that their assessments of these strategies change across development. These different punishment strategies also impact how children view the transgressors and the punishers (or, in the context of a classroom, how they view their peers and teachers). As this understanding grows, educators and others working with children ought to critically consider the various psychological and social impacts of a given sanction on children of varying ages before implementing it in a classroom. More generally, this line of work sheds new light on children's developing perceptions of norm enforcement and enforcers, which is a key feature of human cooperation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We are deeply grateful to all the children and families for their participation. We would also like to thank Kate Haynes, Marie Grace Martinez, and Carolynn McElroy for their help with coding. This work was supported in part by the University of Virginia's Harrison Undergraduate Research Award to Sarah Thomas, and by the National Institute of Child Health and Human Development of the National Institutes of Health (R03HD101652) to Amrisha Vaish.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

- About FE (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental psychology*, 39, 48–60. 10.1037/0012-1649.39.1.48 [PubMed: 12518808]
- Arbeau KA, Coplan RJ, & Weeks M (2010). Shyness, teacher-child relationships, and socio-emotional adjustment in grade 1. *International Journal of Behavioral Development*, 34(3), 259–269. 10.1177/0165025409350959
- Arini RL, Mahmood M, Aljure JB, Ingram GPD, Wiggs L, & Kenward B (2023). Children endorse deterrence motivations for third-party punishment but derive higher enjoyment from compensating victims. *Journal of Experimental Child Psychology*, 230, 105630. 10.1016/j.jecp.2023.105630 [PubMed: 36731278]
- Baron AS, & Dunham Y (2015). Representing ‘Us’ and ‘Them’: Building blocks of intergroup cognition. *Journal of Cognition and Development*, 16(5), 780–801. 10.1080/15248372.2014.1000459
- Baumard N, Mascaro O, & Chevallier C (2012). Preschoolers are able to take merit into account when distributing goods. *Developmental Psychology*, 48, 492–498. 10.1037/a0026598 [PubMed: 22148948]
- Bernath MS, & Feshbach ND (1995). Children's trust: Theory, assessment, development, and research directions. *Applied & Preventive Psychology*, 4(1), 1–19. 10.1016/s0962-1849(05)80048-4
- Boyd R, & Richerson PJ (1989). The evolution of indirect reciprocity. *Social Networks*, 11(3), 213–236. 10.1016/0378-8733(89)90003-8
- Boyd R, & Richerson PJ (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), 3281–3288. 10.1098/rstb.2009.0134
- Bregant J, Shaw A, & Kinzler KD (2016). Intuitive jurisprudence: early reasoning about the functions of punishment. *Journal of Empirical Legal Studies*, 13(4), 693–717. 10.1111/jels.12130
- Brinker SR, Goldstein SE, & Tisak MS (2003). Children's judgements about common classroom punishments. *Educational Research*, 45, 189–198.
- Brugman D, Out C, & Gibbs JC (2016). Fairness and trust in developmental psychology. In Kury H, Redo S & Shea E (Eds.) *Women and Children as Victims and Offenders: Background, Prevention, Reintegration* (pp. 265–289). Springer, New York, NY. 10.1007/978-3-319-08398-8_10
- Chapkovski P. (2021). Strike one hundred to educate one: Measuring the efficacy of collective sanctions experimentally. *PLoS ONE*, 16(4), e0248599. 10.1371/journal.pone.0248599 [PubMed: 33831026]
- Chudek M, & Henrich J (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15(5), 218–226. 10.1016/j.tics.2011.03.003 [PubMed: 21482176]
- Correia I, & Vala J (2004). Belief in a just world, subjective well-being and trust of young adults. In Dalbert C & Sallay H (Eds.), *The justice motive in adolescence and young adulthood: Origins and consequences* (pp. 85–100). Routledge.

- Crick NR, Bigbee MA, & Howes C (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development*, 67(3), 1003–1014. 10.1111/j.1467-8624.1996.tb01779.x [PubMed: 8706506]
- Cushman F. (2015). Punishment in Humans: From Intuitions to Institutions. *Philosophy Compass*, 10(2), 117–133. 10.1111/phc3.12192
- Cushman F, Durwin AJ, & Lively C (2012). Revenge without responsibility? Judgments about collective punishment in baseball. *Journal of Experimental Social Psychology*, 48(5), 1106–1110. 10.1016/j.jesp.2012.03.011
- Cushman F, Sheketoff R, Wharton S, & Carey S (2013). The development of intent-based moral judgment. *Cognition*, 127(1), 6–21. [PubMed: 23318350]
- Damon W. (1975). Early conceptions of positive justice as related to the development of logical operations. *Child Development*, 46(2), 301–312. 10.2307/1128122
- Darby BW, & Schlenker BR (1989). Children's reactions to transgressions: Effects of the actor's apology, reputation and remorse. *British Journal of Social Psychology*, 28(4), 353–364. 10.1111/j.2044-8309.1989.tb00879.x [PubMed: 2611611]
- Darcy S. (2010). Prosecuting the war crime of collective punishment: Is it time to amend the Rome statute? *Journal of International Criminal Justice*, 8(1), 29–51. 10.1093/jicj/mqp083
- Dunham Y, Baron AS, & Carey S (2011). Consequences of “minimal” group affiliations in children. *Child development*, 82(3), 793–811. 10.1111/j.1467-8624.2011.01577.x [PubMed: 21413937]
- Dunham Y, Durkin A, & Tyler TR (2018). The development of a preference for procedural justice for self and others. *Scientific Reports*, 8(1), 17740. 10.1038/s41598-018-36072-1 [PubMed: 30531864]
- Ebner NC, Riediger M, & Lindenberger U (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1), 351–362. 10.3758/BRM.42.1.351 [PubMed: 20160315]
- Elliott SN (1988). Acceptability of behavioral treatments: Review of variables that influence treatment selection. *Professional Psychology: Research and Practice*, 19, 68–80. 10.1037/0735-7028.19.1.68
- Evans IM, Galyer KT, & Smith KJH (2001). Children's perceptions of unfair reward and punishment. *The Journal of Genetic Psychology*, 162(2), 212–227. 10.1080/00221320109597962 [PubMed: 11432606]
- Fehr E, Bernhard H, & Rockenbach B (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079–1083. 10.1038/nature07155 [PubMed: 18756249]
- Fehr E, & Fischbacher U (2003). The nature of human altruism. *Nature*, 425(6960), 785–791. 10.1038/nature02043 [PubMed: 14574401]
- Flanagan CA, Cumsille P, Gill S, & Galloway LS (2007). School and community climates and civic commitments: Patterns for ethnic minority and majority students. *Journal of Educational Psychology*, 99(2), 421–431. 10.1037/0022-0663.99.2.421
- Fraser HP, Kemp S, & Keenan T (2007). How do children behave when they distribute rewards from task participation? *European Journal of Developmental Psychology*, 4(2), 198–219. 10.1080/17405620600602445
- Gao L, Wang Z, Pansini R, Li Y-T, & Wang R-W (2015). Collective punishment is more effective than collective reward for promoting cooperation. *Scientific Reports*, 5(1), 17752. 10.1038/srep17752 [PubMed: 26634907]
- Geraci A. (2021). Toddlers' expectations of corporal third-party punishments against the non-defender puppet. *Journal of Experimental Child Psychology*, 210, 105199. 10.1016/j.jecp.2021.105199 [PubMed: 34146988]
- Geraci A, & Surian L (2021). Toddlers' expectations of third-party punishments and rewards following an act of aggression. *Aggressive Behavior*, 47(5), 521–529. 10.1002/ab.21979 [PubMed: 34101839]
- Geraci A, & Surian L (2023). Preverbal infants' reactions to third-party punishments and rewards delivered toward fair and unfair agents. *Journal of Experimental Child Psychology*, 226, 105574. 10.1016/j.jecp.2022.105574 [PubMed: 36332434]
- Gowing A (2019). Peer-peer relationships: A key factor in enhancing school connectedness and belonging. *Educational and Child Psychology*, 36(2), 64–77. 10.53841/bpsecp.2019.36.2.64

- Guala F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1), 1–15. 10.1017/s0140525x11000069 [PubMed: 22289303]
- Hamlin JK, Wynn K, Bloom P, & Mahajan N (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the USA*, 108(50), 19931–19936. 10.1073/pnas.1110306108 [PubMed: 22123953]
- Harber C. (2021). Punishment. In Harber C (Ed.), *Post-covid schooling: Future alternatives to the global normal* (pp. 97–119). Springer International Publishing. 10.1007/978-3-030-87824-5_4
- Hechter M. (1990). The attainment of solidarity in intentional communities. *Rationality and Society*, 2(2), 142–155. 10.1177/1043463190002002004
- Hechter M, & Kanazawa S (1993). Group solidarity and social order in Japan. *Journal of Theoretical Politics*, 5(4), 455–493. 10.1177/0951692893005004002
- Heckathorn DD (1990). Collective sanctions and compliance norms: A formal theory of group-mediated social control. *American Sociological Review*, 55(3), 366–384. 10.2307/2095762
- Henrich J, Ensminger J, McElreath R, Barr A, Barrett CT, Bolyanatz A, Cárdenas JC, Gurven M, Gwako ELM, Henrich N, Lesorogol C, Marlowe FW, Tracer DP, & Ziker JP (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484. 10.1126/science.1182238 [PubMed: 20299588]
- Hofferth SL, & Sandberg JF (2001). How American children spend their time. *Journal of Marriage and Family*, 63(2), 295–308. 10.1111/j.1741-3737.2001.00295.x
- Hook JG, & Cook TD (1979). Equity theory and the cognitive ability of children. *Psychological Bulletin*, 86, 429–445. 10.1037/0033-2909.86.3.429
- House BR, Kanngiesser P, Barrett HC, Yilmaz S, Smith AM, Sebastián-Enesco C, Erut A, & Silk JB (2020). Social norms and cultural diversity in the development of third-party punishment. *Proceedings of the Royal Society B: Biological Sciences*, 287(1925), 20192794. 10.1098/rspb.2019.2794
- Huppert E, Cowell JM, Cheng Y, Contreras-Ibáñez C, Gomez-Sicard N, Gonzalez-Gadea ML, Huepe D, Ibanez A, Lee K, Mahasneh R, Malcolm-Smith S, Salas N, Selcuk B, Tungodden B, Wong A, Zhou X, & Decety J (2019). The development of children's preferences for equality and equity across 13 individualistic and collectivist cultures. *Developmental Science*, 22(2), e12729. 10.1111/desc.12729 [PubMed: 30207638]
- Imber SC (1973). Relationship of trust to academic performance. *Journal of Personality and Social Psychology*, 28(1), 145–150. 10.1037/h0035644
- Kanakogi Y, Inoue Y, Matsuda G, Butler DL, Hiraki K, & Myowa-Yamakoshi M (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behaviour*, 1(2), 0037. 10.1038/s41562-016-0037
- Kanakogi Y, Miyazaki M, Takahashi H, Yamamoto H, Kobayashi T, & Hiraki K (2022). Third-party punishment by preverbal infants. *Nature Human Behaviour*, 6(9), 1234–1242. 10.1038/s41562-022-01354-2
- Kenward B, & Östh T (2012). Enactment of third-party punishment by 4-year-olds. *Frontiers in Psychology*, 3, 373. 10.3389/fpsyg.2012.00373 [PubMed: 23162486]
- Kenward B, & Östh T (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior*, 41(5), 413–420. 10.1002/ab.21568 [PubMed: 26918430]
- Killen M, Breton S, Ferguson H, & Handler K (1994). Preschoolers' evaluations of teacher methods of intervention in social transgressions. *Merrill-Palmer Quarterly*, 40(3), 399–415. <http://www.jstor.org/stable/23087352>
- Kinzler KD, Dupoux E, & Spelke ES (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30), 12577–12580. 10.1073/pnas.0705345104
- Kogut T. (2012). Knowing what I should, doing what I want: From selfishness to inequity aversion in young children's sharing behavior. *Journal of Economic Psychology*, 33(1), 226–236. 10.1016/j.joep.2011.10.003
- Lee Y-E, & Warneken F (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. *Cognition*, 205, 104374. 10.1016/j.cognition.2020.104374 [PubMed: 32819708]

- Lewis R, & Lovegrove MN (1987). The teacher as disciplinarian: How do students feel? *Australian Journal of Education*, 31(2), 173–186. 10.1177/000494418703100205
- LoBue V, & Thrasher C (2014). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology*, 5, 1532. 10.3389/fpsyg.2014.01532 [PubMed: 25610415]
- Loke IC, Heyman GD, Forgie J, McCarthy A, & Lee K (2011). Children’s moral evaluations of reporting the transgressions of peers: Age differences in evaluations of tattling. *Developmental Psychology*, 47, 1757–1762. 10.1037/a0025357 [PubMed: 21910529]
- Markus HR, & Kitayama S (1991). Cultural variation in the Self-Concept. In Strauss J & Goethals GR (Eds.) *The Self: Interdisciplinary Approaches* (pp. 18–48). Springer, New York, NY. 10.1007/978-1-4684-8264-5_2
- Marlow C, Kelsey CM, & Vaish A (2023). Cheat to win: Children’s judgements of advantageous vs. disadvantageous rule breaking. *Cognitive Development*, 66, 101328. 10.1016/j.cogdev.2023.101328
- Marshall J, Gollwitzer A, & Bloom P (2022). Why do children and adults think other people punish? *Developmental Psychology*, 58(9), 1783–1792. 10.1037/dev0001378 [PubMed: 35511518]
- Marshall J, & McAuliffe K (2022). Children as assessors and agents of third-party punishment. *Nature Reviews Psychology*, 1(6), 334–344. 10.1038/s44159-022-00046-y
- Marshall J, Mermin-Bunnell K, & Bloom P (2020). Developing judgments about peers’ obligation to intervene. *Cognition*, 201, 104215. 10.1016/j.cognition.2020.104215 [PubMed: 32464406]
- Marshall J, Yudkin DA, & Crockett MJ (2020). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*, 5(3), 361–368. 10.1038/s41562-020-00975-9
- Martin JW, Martin S, & McAuliffe K (2021). Third-party punishment promotes fairness in children. *Developmental Psychology*, 57(6), 927–939. 10.1037/dev0001183 [PubMed: 34424010]
- Masters JC, & Furman W (1981). Popularity, individual friendship selection, and specific peer interaction among children. *Developmental Psychology*, 17(3), 344–350. 10.1037/0012-1649.17.3.344
- McAuliffe K, Jordan JJ, & Warneken F (2015). Costly third-party punishment in young children. *Cognition*, 134, 1–10. 10.1016/j.cognition.2014.08.013 [PubMed: 25460374]
- McElroy CE, Kelsey CM, Oostenbroek J, & Vaish A (2023). Beyond accidents: Young children’s forgiveness of third-party intentional transgressors. *Journal of Experimental Child Psychology*, 228, 105607. 10.1016/j.jecp.2022.105607 [PubMed: 36584664]
- McGillicuddy-de Lisi AV, Watkins C, & Vinchur AJ (1994). The effect of relationship on children’s distributive justice reasoning. *Child Development*, 65(6), 1694–1700. 10.2307/1131288 [PubMed: 7859550]
- McLoughlin N, Tipper SP, & Over H (2018). Young children perceive less humanness in outgroup faces. *Developmental Science*, 21(2), e12539. 10.1111/desc.12539
- Meristo M, & Surian L (2013). Do infants detect indirect reciprocity? *Cognition*, 129(1), 102–113. 10.1016/j.cognition.2013.06.006 [PubMed: 23887149]
- Miller JG, Bersoff DM, & Harwood RL (1990). Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology*, 58(1), 33–47. 10.1037/0022-3514.58.1.33 [PubMed: 2308074]
- Misch A, Over H, & Carpenter M (2014). Stick with your group: Young children’s attitudes about group loyalty. *Journal of Experimental Child Psychology*, 126, 19–36. 10.1016/j.jecp.2014.02.008 [PubMed: 24842584]
- Nakao K, & Chai S-K (2011). Criminal conflict as collective punishment. *Economics of Peace & Security Journal*, 6(1), 5–11.
- Oostenbroek J, & Vaish A (2019). The benefits of forgiving: Young children respond positively to those who forgive. *Journal of Experimental Psychology: General*, 148(11), 1914–1924. 10.1037/xge0000576 [PubMed: 31021151]
- Packer MJ, & Moreno-Dulcey FA (2022). Theory of puppets?: A critique of the use of puppets as stimulus materials in psychological research with young children. *Cognitive Development*, 61, 101146. 10.1016/j.cogdev.2021.101146

- Pereira A, & van Prooijen J-W (2018). Why we sometimes punish the innocent: The role of group entitativity in collective punishment. *PLoS ONE*, 13(5), e0196852. 10.1371/journal.pone.0196852 [PubMed: 29723255]
- Riedl K, Jensen K, Call J, & Tomasello M (2015). Restorative Justice in Children. *Current Biology*, 25(13), 1731–1735. 10.1016/j.cub.2015.05.014 [PubMed: 26096976]
- Robbins E, & Rochat P (2011). Emerging signs of strong reciprocity in human ontogeny. *Frontiers in Psychology*, 2, 353. 10.3389/fpsyg.2011.00353 [PubMed: 22194730]
- Rotenberg KJ, Fox CL, Green S, Ruderman L, Slater K, Stevens KR, & Carlo G (2005). Construction and validation of a children's interpersonal trust belief scale. *British Journal of Developmental Psychology*, 23(2), 271–293. 10.1348/026151005x26192
- Rotenberg KJ, Addis N, Betts LR, Corrigan A, Fox CL, Hobson Z, Rennison S, Trueman M, & Boulton MJ (2010). The relation between trust beliefs and loneliness during early childhood, middle childhood, and adulthood. *Personality and Social Psychology Bulletin*, 36(8), 1086–1100. 10.1177/0146167210374957 [PubMed: 20585058]
- Rotenberg KJ, Qualter P, Holt NL, Harris RA, Henzi SP, & Barrett L (2014). When trust fails: The relation between children's trust beliefs in peers and their peer interactions in a natural setting. *Journal of Abnormal Child Psychology*, 42(6), 967–980. 10.1007/s10802-013-9835-8 [PubMed: 24469229]
- Salali GD, Juda M, & Henrich J (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior*, 36(2), 86–94. 10.1016/j.evolhumbehav.2014.09.004
- Schmidt MF, Rakoczy H, & Tomasello M (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, 124(3), 325–333. 10.1016/j.cognition.2012.06.004 [PubMed: 22766522]
- Schmidt MFH, & Tomasello M (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21(4), 232–236. 10.1177/0963721412448659
- Selman RL, & Dray AJ (2006). Risk and prevention. In Renniger KA, Siegel IE, Damon W, & Lerner RM (Eds.), *Handbook of child psychology: Child psychology in practice*, (Vol. 4, 6th ed., pp. 378–419). John Wiley & Sons Inc.
- Shaw A, & Olson KR (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, 141(2), 382–395. 10.1037/a0025907 [PubMed: 22004168]
- Smetana JG, & Ball CL (2017). Young children's moral judgments, justifications, and emotion attributions in peer relationship contexts. *Child Development*, 89(6), 2245–2263. 10.1111/cdev.12846 [PubMed: 28586086]
- Smetana JG, Jambon M, & Ball C (2014). The social domain approach to children's moral and social judgments. In Killen M & Smetana JG (Eds.), *Handbook of moral development*, (2nd ed., pp. 23–45). Psychology Press.
- Smetana JG, Ball CL, Jambon M, & Yoo HN (2018). Are young children's preferences and evaluations of moral and conventional transgressors associated with domain distinctions in judgments? *Journal of Experimental Child Psychology*, 173, 284–303. 10.1016/j.jecp.2018.04.008 [PubMed: 29772455]
- Smith CE, Blake PR, & Harris PL (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS ONE*, 8(3), e59510. 10.1371/journal.pone.0059510 [PubMed: 23527210]
- Smith CE, & Warneken F (2016). Children's reasoning about distributive and retributive justice across development. *Developmental Psychology*, 52(4), 613–628. 10.1037/a0040069 [PubMed: 26845506]
- Stengelin R, Haun DBM, & Kanngiesser P (2023). Simulating peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development? *Child Development*, 94, 1117–1135. 10.1111/cdev.13913 [PubMed: 36779431]
- Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, & Kirsch P (2011). Beyond revenge: Neural and genetic bases of altruistic punishment. *NeuroImage*, 54(1), 671–680. 10.1016/j.neuroimage.2010.07.051 [PubMed: 20673803]

- Thomas KJ, De Cunha JM, De Souza DA, & Santo JB (2019). Fairness, trust, and school climate as foundational to growth mindset: A study among Brazilian children and adolescents. *Educational Psychology*, 39(4), 510–529. 10.1080/01443410.2018.1549726
- Thornberg R. (2008). School children's reasoning about school rules. *Research Papers in Education*, 23(1), 37–52. 10.1080/02671520701651029
- Tisak MS (1993). Preschool children's judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly*, 39(3), 375–390. <http://www.jstor.org/stable/23087427>
- Vaish A, Carpenter M, & Tomasello M (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, 81(6), 1661–1669. 10.1111/j.1467-8624.2010.01500.x [PubMed: 21077854]
- Vaish A, Herrmann E, Markmann C, & Tomasello M (2016). Preschoolers value those who sanction non-cooperators. *Cognition*, 153, 43–51. 10.1016/j.cognition.2016.04.011 [PubMed: 27135711]
- 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. 10.1348/026151010X532888 [PubMed: 21288257]
- Waterman AS (1984). *The psychology of individualism*. Praeger, New York, NY.
- Yucel M, & Vaish A (2018). Young children tattle to enforce moral norms. *Social development*, 27(4), 924–936. 10.1111/sode.12290
- Zelazo PD, Helwig CC, & Lau AY (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, 67(5), 2478–2492. 10.1111/j.1467-8624.1996.tb01869.x
- Ziv T, Whiteman JD, & Sommerville JA (2021). Toddlers' interventions toward fair and unfair individuals. *Cognition*, 214, 104781. 10.1016/j.cognition.2021.104781 [PubMed: 34051419]

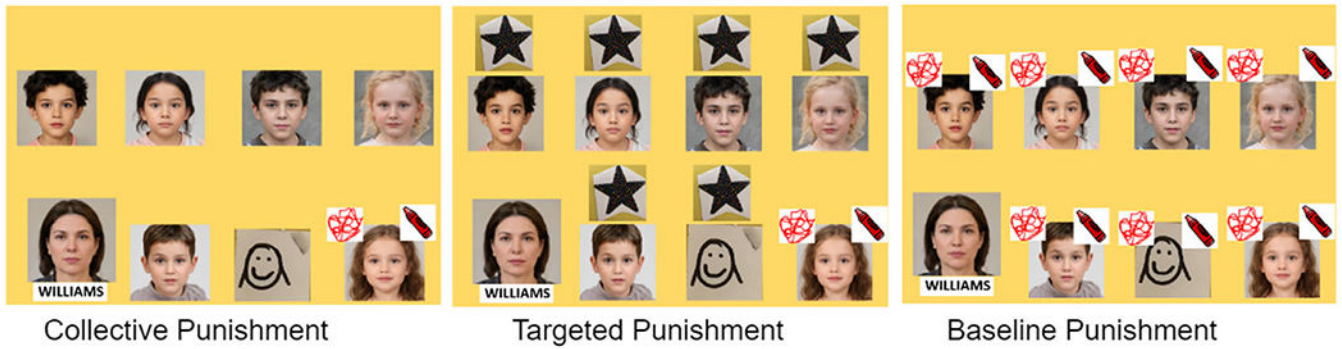


FIGURE 1.

Example of the experimental setup for each of the punishment types. *Note:* The participant is represented by a self-portrait, the transgressor(s) are identified by the presence of the crayon and scribbles, and the stars are the “prize” stickers given out in the scenario. The images used in the present study were taken from standardized datasets (Ebner et al., 2010; LoBue & Thrasher, 2014). However, for copyright reasons, the images shown here are AI-generated photos simulated using <https://generated.photos/faces/child>.

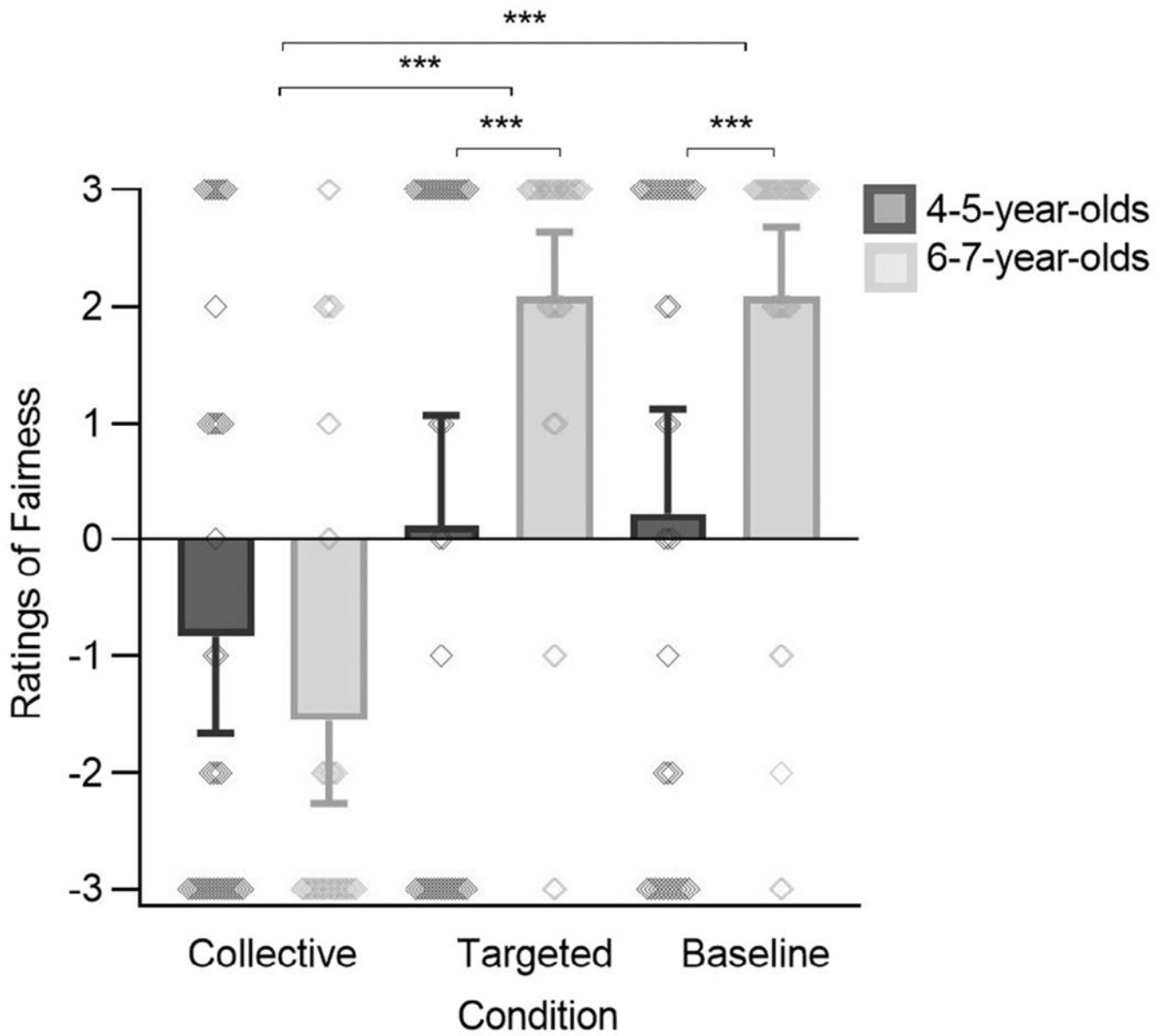


FIGURE 2. Mean fairness ratings in each condition, by age. *Note:* Error bars represent the 95% confidence interval and individual data points are indicated with diamond shapes. Symbols: *** $p < .001$.

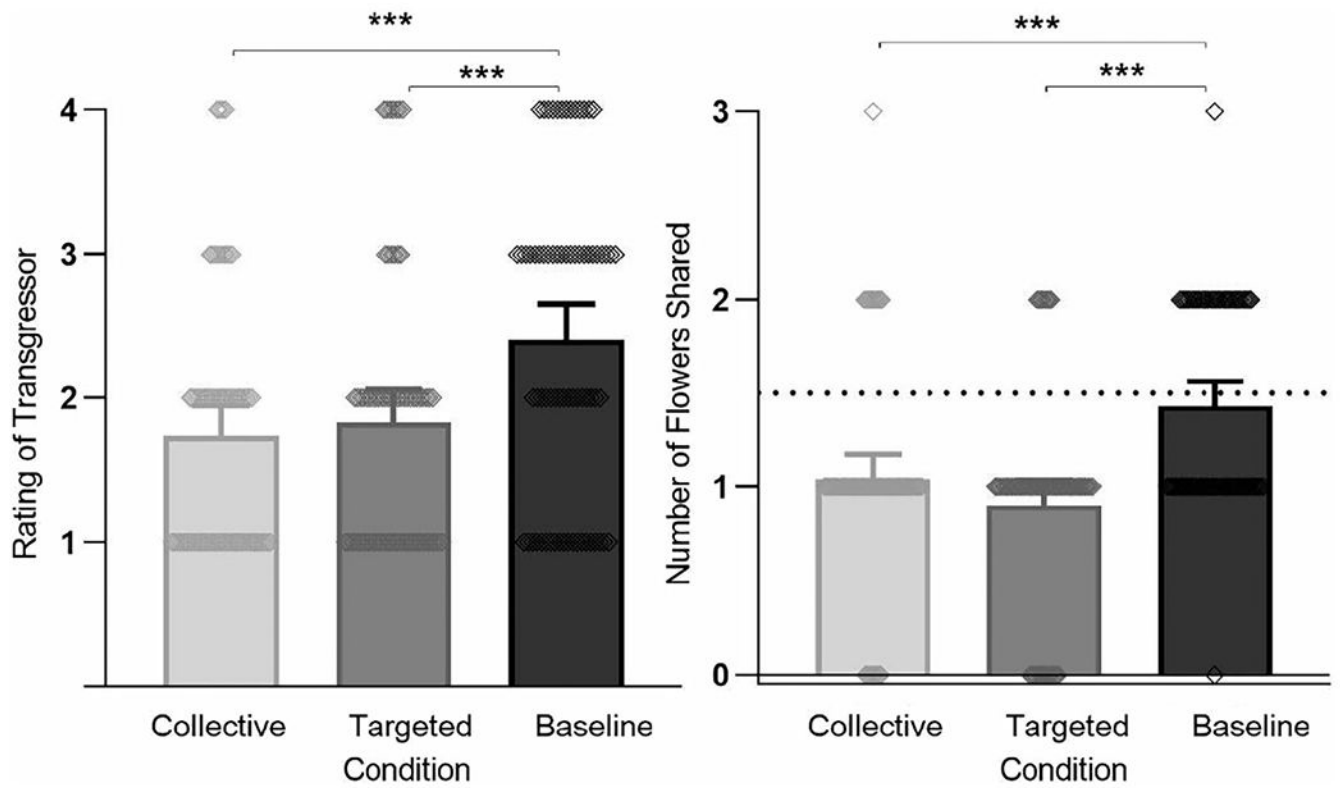


FIGURE 3.

Mean ratings of liking the transgressor and average number of flowers shared with the transgressor in each condition. *Note:* Error bars represent the 95% confidence interval and individual data points are indicated with diamond shapes. Symbols: *** $p < .001$.

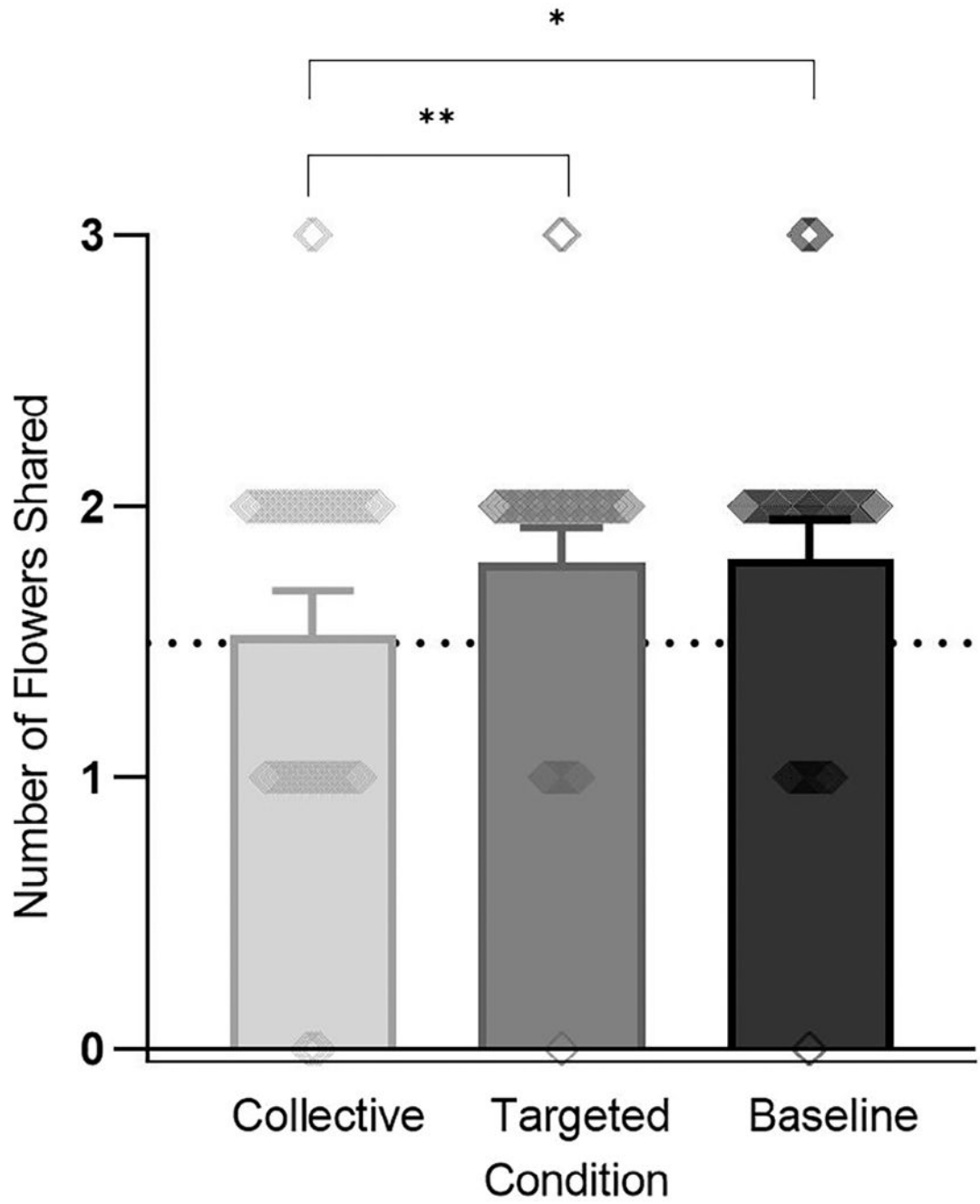


FIGURE 4.

Mean number of flowers shared with the teacher in each condition. *Note:* Error bars represent the 95% confidence interval and individual data points are indicated with diamond shapes. * $p < .05$, ** $p < .01$.