

Young Children Police Group Members at Personal Cost

Daniel A. Yudkin
Yale University

Jay J. Van Bavel and Marjorie Rhodes
New York University

Humans' evolutionary success has depended in part on their willingness to punish, at personal cost, bad actors who have not harmed them directly—a behavior known as costly third-party punishment. The present studies examined the psychological processes underlying this behavior from a developmental perspective, using a novel, naturalistic method. In these studies (ages 3–6, total $N = 225$), participants of all ages enacted costly punishment, and rates of punishment increased with age. In addition, younger children (ages 3–4), when in a position of authority, were more likely to punish members of their own group, whereas older children (ages 5–6) showed no group- or authority-based differences. These findings demonstrate the developmental emergence of costly punishment, and show how a sense of authority can foster the kind of group-regulatory behavior that costly punishment may have evolved to serve.

Keywords: third-party punishment, reputation, cooperation, fairness, development

Humans' remarkable social and technological achievements were made possible in part by their unique ability to maintain cooperative arrangements among groups of unrelated individuals (Boyd, Gintis, Bowles, & Richerson, 2003). Such cooperation may have been facilitated by people's willingness to pay personal costs to punish those who violate group norms—a behavior known as costly third-party

punishment (Hauert, Traulsen, Brandt, Nowak, & Sigmund, 2007). Because such behavior provides a deterrent for would-be cheaters and free riders, it can promote cooperation over time (Balliet, Mulder, & Van Lange, 2011; Fehr & Gächter, 2002; Mathew & Boyd, 2011).

Although costly punishment has been widely observed in a variety of human societies (Boyd et al., 2003; Henrich et al., 2006), the psychological motivations underlying it remain uncertain (Balafoutas, Nikiforakis, & Rockenbach, 2014; Krasnow, Delton, Cosmides, & Tooby, 2016). Some evidence suggests that people punish primarily for reputational reasons (dos Santos, Rankin, & Wedekind, 2011; Krasnow, Cosmides, Pedersen, & Tooby, 2012; Pedersen, Kurzban, & McCullough, 2013; Pedersen, McAuliffe, & McCullough, 2018). For instance, people are more likely to engage in costly punishment when observed (Kurzban, DeScioli, & O'Brien, 2007). People also judge others who have enacted punishment as more trustworthy (Jordan, Hoffman, Bloom, & Rand, 2016) and are subsequently more willing to cooperate with them (Nelissen, 2008), supporting the notion that punishment may be used as a signal of virtue.

By contrast, other work suggests that costly punishment may stem from the desire to promote good behavior in cooperative communities, a concept termed "in-group policing" (Boyd et al., 2003; Crockett, Özdemir, & Fehr, 2014; Fearon & Laitin, 1996; Piazza & Bering, 2008). Group policing implies that punishment should be disproportionately directed toward in-group versus out-group members, since it is the former with whom future cooperation is most likely. Accordingly, some studies have found that people punish transgressors within their own group more harshly (e.g., Schmidt, Rakoczy, & Tomasello, 2012; Shinada, Yamagishi, & Ohmura, 2004). However, other research finds that out-group members are punished more harshly (Bernhard, Fehr, & Fischbacher, 2006; Delton & Krasnow, 2017; Jordan, McAuliffe, & Warneken, 2014; Yudkin, Rothmund, Twardawski, Thalla, & Van Bavel, 2016; see also Balafoutas et al., 2014), suggesting that people may be using punishment as a means of inflicting damage on potential competitors. Thus, the basic psychological motivations driving people to perform costly third-party punishment remain unclear.

This article was published Online First May 16, 2019.

Daniel A. Yudkin, Department of Psychology, Yale University; Jay J. Van Bavel, Center for Neural Science and Department of Psychology, New York University; Marjorie Rhodes, Department of Psychology, New York University.

This research has been disseminated in several conferences and conventions. Daniel A. Yudkin presented the research at the Justice and Morality Preconference at the 2018 meeting of the Society for Personal and Social Psychology and at the 2015 meeting of the Association for Psychological Science. Marjorie Rhodes presented the research at the 2017 meeting of the Cognitive Development Society. The manuscript has been posted on PsyArXiv (psyarxiv.com/rc4ta). Daniel A. Yudkin, Jay J. Van Bavel, and Marjorie Rhodes designed the research, Daniel A. Yudkin collected the data, Daniel A. Yudkin analyzed the data with input by Jay J. Van Bavel and Marjorie Rhodes, and Daniel A. Yudkin wrote the paper, with critical edits by Jay J. Van Bavel and Marjorie Rhodes.

We thank Elyana Feldman, Lisa Kaggen, Christine Tai, Rachel Vitale, and Kat Yee for help with data collection and the New York University Social Perception and Evaluation Lab for feedback on this research. We further thank Eva Jost and Jack Knowles for their dramatic performances. This work was supported by a Research Challenge Grant for Women in Science and by a James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition–Scholar Award to Marjorie Rhodes and National Science Foundation (Grant 1349089; to Jay J. Van Bavel). We thank the Children's Museum of Manhattan for their support of this research, and Jordan Wylie for valuable feedback.

Correspondence concerning this article should be addressed to Daniel A. Yudkin, Department of Psychology, Yale University, 10 Hillhouse Avenue, New Haven, CT 06511, or to Marjorie Rhodes, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: day235@nyu.edu or marjorie.rhodes@nyu.edu

Overview of Current Research

The current research uses a developmental approach to better understand the motivations underlying costly punishment. Working with young children confers several advantages that are ideally suited to answering basic questions regarding the psychology of this behavior. First, it is easier to create situations children find psychologically meaningful, which helps contribute to a more robust test of behavior than the more removed scenarios often used in adult research. Second, because important cognitive capacities develop at various points in childhood, it becomes possible to examine the specific impact these capacities have on punishment behavior. Third, because young children have had less chance to engage in cultural learning than at a later point in life, studying their behavior provides an opportunity to get a clearer picture of basic human tendencies inherent in social cognition.

The current research consists of two studies designed to assess whether children do indeed enact costly punishment at young ages, and to begin to identify the circumstances under which they are more or less likely to do so. The rationale for this was that, by better understanding the conditions under which people enact costly punishment, we could help shed light on the motivations underlying this behavior while at the same time better understand the development of moral cognition more generally.

Our research focused on several primary factors. First, building on the research described above, we examined the influence of group membership on rates of costly punishment. In addition, we sought to test the influence of the sense of authority within the group. Past research suggests that hierarchical arrangements can help solve coordinated-action dilemmas (Glowacki & von Rueden, 2015; Van Vugt, 2006). For instance, groups with leaders willing to punish free riders are more successful at preserving shared resources (Kosfeld & Rustagi, 2015). Thus, people who feel a sense of leadership or authority may be more inclined to utilize punishment to improve group members' behavior.

We tested this hypothesis in young children (ages 3–6), because we expected that, if a sense of authority really does elicit group-regulatory efforts, such behaviors may emerge early in development. By age 3, children have an abstract understanding of what it means to be a member of a social group (Rhodes, 2013), and of the importance of maintaining and enforcing group norms (Schmidt et al., 2012). For example, 3-year-olds expect members of the same group not to harm one another (Rhodes, 2012) and view transgressions that occur among members of the same group as particularly problematic (Rhodes & Chalik, 2013). Young children also take actions to correct transgressions that harm others (e.g., returning stolen objects to their rightful owners, Riedl, Jensen, Call, & Tomasello, 2015), are willing to incur a cost to avoid interacting with norm violators (Tasimi & Wynn, 2016), reject antisocial others (Hamlin, Wynn, Bloom, & Mahajan, 2011), intervene against moral transgressions (Vaish, Missana, & Tomasello, 2011), and appear to expect antisocial actions to be punished (Kenward & Öst, 2012).

Whereas all of this previous work suggests that young children have the requisite capacities and motivations to engage in costly punishment—particularly if it serves a functional role in maintaining group norms—previous studies have not found evidence of costly punishment in children younger than age 6 (Jordan et al., 2014; McAuliffe, Blake, Steinbeis, & Warneken, 2017; McAuliffe,

Jordan, & Warneken, 2015; Salali, Juda, & Henrich, 2015). One possible explanation is that earlier studies on costly punishment in children used tasks modeled on those from the adult literature on this topic, which required numerical comparisons that may have been challenging for young children (Chernyak & Sobel, 2016). Thus, a key goal of the present studies was to test for evidence of costly punishment in young children using a naturalistic experimental design, and to test whether it arises under conditions conducive to group regulation.

Study 1

Method

In order to render our experimental method accessible to young children, we developed a novel method for assessing costly punishment behavior that avoided the need for quantitative comparisons. In Study 1, we conducted a preliminary validation of this method to see whether children in our target age range would enact costly punishment. Participants were randomly assigned to either a “transgression” or a “no transgression” condition and were then given the opportunity to punish the target. We expected that, if the punishment was indeed costly, children would be reluctant to enact punishment unless they observed a transgression that was, indeed, deserving of punishment. All data, experimental scripts, and protocols are available at the Open Science Framework (osf.io/y9n5a/) and all studies were approved by the New York University Ethics Board (Protocol IRB-FY2016-760). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants. Given that people tend to punish harmful more than harmless behavior (e.g., Rossi, Waite, Bose, & Berk, 1974), we anticipated a large effect size ($V = .5$, power = 80%) and so sought a sample of 35 participants. The sample was 17 female, 18 male, nine White, nine Black, two Asian, two Hispanic, nine mixed race/other, four unreported ($M_{\text{age}} = 5.4$, range 4–6 years). Children were recruited from and tested at the Children's Museum of Manhattan, New York.

Procedure. Children were brought into a private classroom in the museum, with a large slide (Figure 1). They were given the opportunity to test out the slide by going down it once (all did) and were asked whether and how much they liked going down it. Participants were then brought to a small table and asked to draw a picture of something they did at the museum that day. Afterward, they were introduced to two additional children, shown on video:

Now, I'd like to tell you a little story about what happened a few minutes ago with the last children who were in here. A few minutes ago, these two girls were in here, and they drew pictures just like you did. One girl's name was Jill, and one's name was Stacey. This is Stacey. When Jill was done with her drawing, she had to go to the bathroom, so she gave Stacey her picture to hold for her until she came back. Now I'm going to show you a video of what happened next. Remember, Jill asked Stacey to hold her picture. See, here is Stacey holding Jill's picture. Let's see what she does next.

By random assignment, children were then shown either a “harmless” video, in which the child carefully held the picture, or a “harmful” video, in which she crumpled the picture and threw it on the ground.

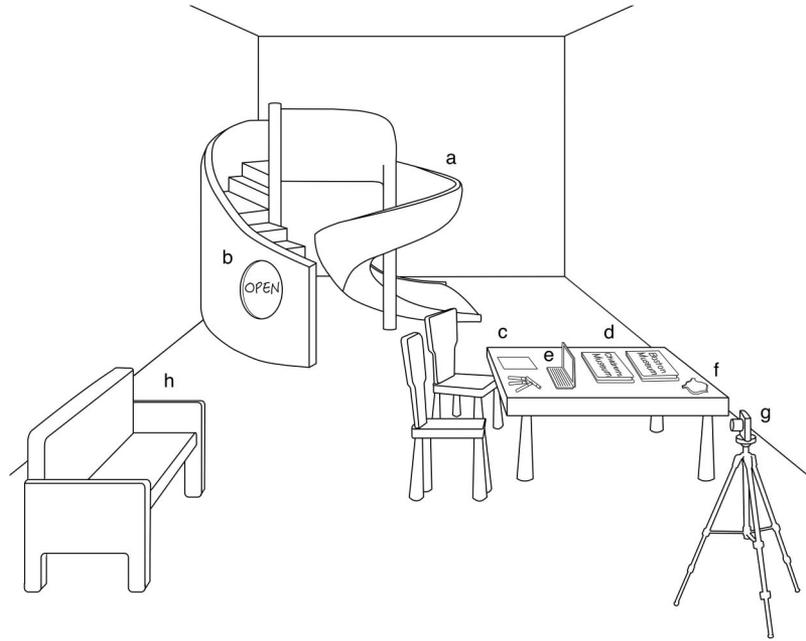


Figure 1. Schematic of experiment space. (A) Slide to be revoked as punishment. (B) Open/closed sign. (C) Paper for drawing project. (D) Books in which the art project would be stored. (E) Computer on which transgressor video was played. (F) Sheriff's badge randomly assigned to participants. (G) Video camera on which interaction was recorded. (H) Couch where transgression occurred.

Afterward, children were told that the person who either held or ruined the picture wanted to come back and go down the slide again, and that the child's task was to decide what should happen:

Now I want to ask you something. See here I have a sign. On one side it says OPEN. On the other side it says CLOSED. We're going to put this on the slide. If we want to let Stacey go down the slide, then we should put up the OPEN sign. If we do not want to let her go down, we should put the CLOSED sign. But remember, if we put the CLOSED sign, then you cannot go down the slide either.

Children were then allowed to place the sign on the slide to mark it as either open or closed. Children who chose to close the slide enacted costly punishment, because in doing so, they sacrificed their own future opportunity play with the slide in the service of punishing the child for her transgression. Afterward, children completed a series of questions to assess their understanding of and feelings about the situation that they had just experienced, including whether they initially enjoyed the slide, whether they understood that closing the slide meant they could not go down it again, whether the actor's behavior was right or wrong, and whether she should get in trouble or not. In addition, we included a measure of noncostly punishment. At the conclusion of the study, children were offered the chance to choose a sticker as thanks for their participation. The experimenter then told them, "I also have [a sticker] that I could give to Stacey. Do you think I should give a sticker to Stacey?" Noncostly punishment was reflected in participants saying that the experimenter should not give a sticker to the transgressor (for full scripts, see osf.io/y9n5a/). In order to ensure the findings generalized across contents of the video, four transgression stimuli were used, including different variations of

the transgression (crumpling vs. tearing the drawing) and actors (female and male transgressors).

Results

We expected that participants would not enact costly punishment in situations in which there was no reason to do so (harmless condition), but would be willing to punish transgressors (harmful condition). Indeed, 57% of participants punished in the harmful condition, whereas none did so in the harmless condition,¹ $\chi^2(1, N = 35) = 10.79, p = .001, V = .55$. Confirming that children understood the nature of the punishment they were administering, 100% of the children who closed the slide understood that they could not go down the slide again, despite the fact that all had stated they enjoyed going down the slide initially. Thus, we were confident that these children knowingly engaged in the costly punishment of transgressors.

Confirming that children judged the transgression as wrong, more children in the harmful condition stated the actor's behavior was wrong (87%), compared with the harmless condition (9%), $\chi^2(1, N = 34) = 19.1, p < .001, V = .75$, and more children stated that the transgressor should get into trouble (77%), compared with the harmless condition (0%), $\chi^2(1, N = 34) = 18.5, p < .001, V =$

¹ In order to make best use of the statistical power available to us, and in accordance with statistical theory (Fleiss, Levin, & Paik, 2004), we apportioned more participants to the harmful ($n = 23$) than the harmless condition ($n = 12$). This allowed us to better ascertain the base rate of punishment in the sample while simultaneously establishing statistical difference between these groups.

.74. These judgments of permissibility and reprimand were both positively associated with a willingness to enact costly punishment, $\chi^2(1, N = 34) = 13.0, p < .001, V = .61$, and $\chi^2(1, N = 34) = 12.9, p < .001, V = .61$, respectively.

Sanctioning of bad behavior was also evident in the patterns of noncostly punishment (choosing to withhold a sticker from the child), which more children enacted in the harmful condition (77%) than the harmless condition (8%), $\chi^2(1, N = 34) = 14.8, p < .001, V = .66$. There were no effects of video version, participant gender, transgressor gender, or interactions with these variables, all $ps > .25$. There was no linear effect of age ($p = .17$). Overall, these data provide the first evidence that children below the age of 6 will punish transgressors at personal cost and suggest that they do so because they recognize a moral transgression.

Study 2

Method

Having established that very young children engage in costly punishment, we then examined the conditions that elicit costly punishment in early childhood. To do so, all children in this experiment witnessed the transgression. To test if children are more likely to engage in costly punishment behavior in order to regulate the behavior of group members, we manipulated whether children shared group membership with the transgressor or not, as well as whether they were given a sense of authority in the group or not. We also wanted to see whether there were differences in punishment according to age. Previous research has revealed developmental differences in other-regarding preferences that emerge between the ages of 3–4 and 5–6 (e.g., Engemann, Herrmann, & Tomasello, 2016; Fehr, Bernhard, & Rockenbach, 2008). Additionally, little research has directly tested willingness to enact costly punishment among children ages 3–4, thus we were particularly interested the behavior of this younger age group of children on their own. Finally, past work has found a range of social-cognitive skills that emerge at around age 5, but not before (e.g., Best, Miller, & Jones, 2009; Engemann & Rapp, 2018; Wellman, 1992). Thus, we tested for punishment differences between younger (ages 3–4) and older children (ages 5–6).

Participants. A power analysis performed in G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) suggested a sample of approximately 200 would be capable of detecting the small-to-medium effects and interactions of age and our experimental manipulations that we anticipated (Cohen's $f^2 = .06$, number of predictors = 4, power = 80%). Our initial sample consisted of 203 participants, of which 13 were excluded from all analysis on the basis of a priori exclusion criteria: five for inattention or a learning disability, three for experimenter error, and five for parental interference (further information on excluded participants can be found at osf.io/y9n5a/). The sample retained for analyses included 190 participants (90 male, 100 female, 53 White, 15 Black, 25 Asian, 17 Hispanic, 60 mixed race/other, 21 unreported, $M_{\text{age}} = 5.2$, range 3–6 years). There were 106 children ages 5–6, 78 children ages 3–4, and six with no age indicated.

Procedure. Procedures were similar to those of Study 1, with several key exceptions. First, it was possible that demand effects contributed to the rates of punishment observed in Study 1: children might have felt pressured into punishing by the presence of an

experimenter who had also witnessed the transgression and thus expected them to punish. In order to reduce these pressures, we employed two experimenters. Experimenter 1 conducted the first part of the experiment until the transgression occurred, then said she was going to do some work and would get her friend to help with the rest of the activity. She left the room and Experimenter 2 entered, ostensibly blind to the transgression, and said, "So my friend told me you're deciding which sign to put up. Are you ready to put up the sign?" In this way, we sought to minimize the pressure children would feel to act in a certain way due to experimenter demand.

Second, we manipulated several features of the context to test the motivations underlying costly punishment. In particular, participants were brought to a table to perform a drawing activity similar to Study 1, but with more context about the goal of the drawing activity to introduce the group membership component of the design. Group membership was manipulated through the introduction of different books into which the participants' and the actors' drawings would belong to. Participants were told:

We're going to use our book to show other people how creative all the kids who come to our museum are. Here is the book! [Shows book]. See what it says? "Children's Museum!" That means it's the book for this museum. Children at lots of different museums in other cities are making books just like this. Here is a book for a different museum. [Shows book.] See what it says? "Boston Museum." That means it's the book for a different museum—in a different city.

In the in-group condition, children were told that the characters in the video "come to our museum all the time. So their pictures aren't for the *other* book—they'll go into *our* book." In the out-group condition, children were told that the characters in the video "belong to the other museum. So their pictures aren't for *our* book—they go in this *other* book." In order to help ensure that the manipulation affected participants' actual sense of group identity, instead of mere inferences regarding the preferences or similarity of the transgressor (Mussweiler & Ockenfels, 2013), we included in the experimental protocol explicit references to group membership, probing children about which book the characters' picture would belong to ("the book for *our* museum" vs. "the book for *their* museum").

We also manipulated children's sense of authority. Previous research suggests that social cues may have a significant impact on children's attitudes and behavior (e.g., Karniol et al., 2011). Accordingly, we used a symbol of leadership—a sheriff's badge—to manipulate participants' sense of authority. In "badge" conditions, the experimenter started the experiment wearing the sheriff's badge and then, just before the presentation of the transgression, told children that they were "in charge" and gave them the Sheriff's badge. In "no-badge" conditions, there was no badge present and children were not told that they were in charge.

After the child placed the sign, Experimenter 2 asked a series of open-ended and scale questions to probe their understanding and feelings about the situation, including why they chose to leave the slide open or closed, whether they understood that this meant they could no longer go down it, and whether they felt in charge (*No/A little/Sort of/Very*), like the "boss" (*No/Yes*), responsible for making sure other people were following the rules (*No/Yes*), whether it was important to teach the transgressor a lesson (*No/A little/Sort of/Very*), how bad the behavior was (*Not at all/Just a little/Pretty!*

Very), and whether the transgressor should get in trouble (*No/A little/A lot*). They were then thanked for their participation and parents were fully debriefed.

Results

Almost all (97%) of participants rated the slide as enjoyable; and 83% children responded correctly on the first try to a manipulation check of transgressor group membership; the rest were prompted until they gave the correct response (and there were no age differences on this manipulation check). There were no differences in these responses according to age, gender, video version, or their interactions (all $ps > .15$).

Children in the badge condition (54%) were more likely to say that they felt like the boss than those in the no-badge condition (34%), $\chi^2(1, N = 175) = 6.96, p = .008, V = .2$, which was subsequently associated with participants' saying they felt it was their job to make sure everyone was following the rules, $\chi^2(1, N = 97) = 13.4, p < .001, V = .37$. Participants were also more likely to say that they felt in charge when they were in the badge condition (77%) than the no-badge condition (54%), $\chi^2(1, N = 188) = 11.20, p < .001, V = .244$. Overall, almost all children viewed the transgressor's behavior as at least somewhat wrong (95%) and almost all believed it was important for the transgressor to change his or her behavior (89%). None of these effects were moderated by age, gender, or condition, $ps > .25$, except that older children were more likely than younger to say it was important for the transgressor to change his or her behavior, $\chi^2(1, N = 182) = 8.23, p = .004, V = .213$.

Next, we examined our primary dependent variable: rates of costly punishment according to group membership and authority. Overall, 48% of children engaged in costly punishment. The likelihood of punishment was higher among the older (74.4%, 95% CI [5.4, 69.7]) than the younger cohort (28.2%, 95% CI [18.6, 39.5]), $\chi^2(1, N = 184) = 18.68, p < .001, V = .32$; nevertheless, both of these proportions differed significantly from 0 (all $ps < .001$; Figure 2). As evidence that children understood the personal cost of their decision to close the slide, 91% of those who closed the slide responded correctly to a follow-up comprehension check by confirming that they could not go down the slide again.

We also conducted a complete text analysis of participant verbal responses following punishment. These responses were open-ended questions allowing participants to indicate, in their own words, their motivations for their behavior.² Among those who closed the slide, 77% referenced the transgressor's behavior as a reason for doing so (e.g., "Because she crumpled up someone's picture"). In addition, among those who left the slide open, the most frequent specific explanation for this behavior made a direct reference to the cost either to themselves or the transgressor (e.g., "Because then I wouldn't get to go"). These explanations serve as further evidence that participants understood both the reasons for and the consequences of their actions. In addition, even following punishment, children appeared to continue viewing the resource they had given up as valuable: at the conclusion of the experiment, participants were given an opportunity to go down the slide one more time by the experimenter temporarily flipping the sign to "open." Of those who had closed the slide, 85% chose to go down again. Overall, these data provide clear evidence from a large

sample that children below the age of 6 knowingly engaged in costly punishment behavior.

We next tested the effects of age, group manipulation, and authority manipulation on punishment, using a 2 (transgressor: in-group vs. out-group) \times 2 (authority: badge vs. no badge) \times 2 (age: younger vs. older) binary logistic regression model to see if there was any evidence of in-group policing in the authority condition. The analysis revealed a significant three-way interaction between group, role, and age ($B = -.376, SE = .171, \text{Wald } \chi^2 = 4.84, p = .028, \text{odds ratio } [OR] = .69$),³ as well as a two-way interaction between group and role among younger children, $B = -.629, SE = .275, \text{Wald's chi-square} = 5.22, p = .022, OR = .53$. Pairwise contrasts within condition among the younger cohort suggested this interaction was driven by the different treatment of in-group and out-group members according to authority role. Specifically, children assigned to the authority role punished in-group members (43%) at a significantly higher rate than out-group members (17%), $p = .038$, whereas children not assigned to this role did not show an effect of group membership (indeed, the pattern of means was in the opposite direction, though not significantly so, $p = .109$; Figure 3). Meanwhile, older children demonstrated no main effects, simple effects, or interactions of transgressor group membership or authority role on punishment (all $ps > .25$; though as noted above, older children punished at higher rates overall). In sum, younger, but not older, children demonstrated different patterns of punishment according to group membership and authority.

In order to see whether these behavioral tendencies were reflected in children's attitudes, we performed parallel analyses on children's sense of responsibility for regulating others' behavior. Results showed that the extent to which children felt in charge mirrored the pattern of effects in punishment, with a three-way interaction between age, badge condition, and group condition, $B = .38, SE = .14, \text{Wald's chi-square} = 6.78, p = .009$. Specifically, while older children's sense of responsibility did not differ as a function of their group or leadership condition, $p > .25$, younger children's did, $B = -.55, SE = .21, \text{Wald's chi-square} = 6.37, p = .012$. Simple effects tests showed that, when punishing in-group members, participants in an authority role felt significantly more responsible than those in the control, $B = .75, SE = .30, \text{Wald's chi-square} = 6.39, p = .011$; no significant effects of authority emerged in the out-group transgressor of authority emerged in the out-group transgressor condition, $p > .25$ (see the supplementary materials available at the Open Science Framework: osf.io/y9n5a/). Finally, there was a positive correlation between costly punishment and children's reports that it was important for the transgressor to change his or her behavior, $r = .32, p < .001$, and belief it was important for them to teach the transgressor a lesson, $r = .23, p = .001$. These findings suggest that participants view costly punishment as to some extent serving

² Two independent raters blind to condition coded the responses. The raters demonstrated a high rate of agreement (Cohen's $\kappa = .91$). Ratings that differed from each other (approximately 3% of responses) were arbitrated by a third rater who was blind to condition. Verbal responses and codes are available at osf.io/y9n5a/

³ At reviewers' request we also include an analysis of the group and badge manipulations with age treated as a continuous (rather than categorical) variable: $B = -.31, SE = .20, \text{Wald } \chi^2 = 2.46, p = .117, OR = .73$.

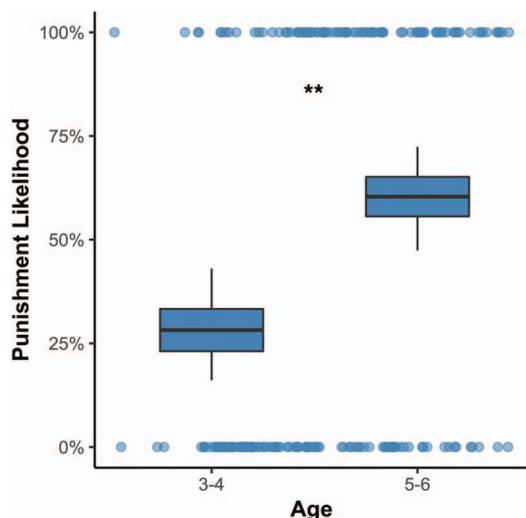


Figure 2. Likelihood of costly punishment (y-axis) predicted by age (x-axis; $n = 190$, Study 2). Center lines within boxes represent group means; edges, *SEM*; whiskers, Clopper-Pearson 95% confidence intervals. Dots: individual participants by exact age. X-axis: exact age with labels centered on each age cohort. ** $p < .001$. See the online article for the color version of this figure.

a pedagogical function—teaching group members how to behave appropriately. Whereas other motivations may also be at play, costly punishment appears to be associated with the desire to ensure that transgressors learn a lesson and change their behavior.

Noncostly punishment (withholding a sticker from the transgressor) was administered by 63.6% of participants, a significantly higher proportion than those who delivered costly punishment (47.6%), $z = 3.12$, $p = .002$. Rates of noncostly punishment significantly increased with age, $B = .49$, $SE = .19$, Wald's chi-square = 2.82, $p = .010$, $OR = 1.63$. There was no Age \times Punishment Type interaction, $p = .37$, and no significant three-way interaction of age, role, and group, $p = .61$; thus, in this respect, noncostly punishment did not mirror the group regulatory pattern of costly punishment. Noncostly punishment was predicted by participants' sense of authority (feeling like "the boss"), $\chi^2(1, N = 183) = 7.64$, $p = .006$, $V = .21$, and by their belief that the transgressor should be reprimanded for his or her behavior, $\chi^2(1, N = 178) = 15.03$, $p < .001$, $V = .29$, corroborating the effects observed in costly punishment.

Discussion

The current research sheds light on the developmental origins and psychological processes underlying costly punishment. Past research has been mixed as to whether people punish in-group or out-group members more harshly. Our research suggests that a critical determinant of these effects may be people's sense of authority: whereas young children in an authority role tended to punish in-group members more harshly than out-group members, children without authority demonstrated no such differences. Children's feelings of responsibility mirrored this pattern of results—thereby corroborating the primary findings and supporting the general conclusions. Taken together, these results suggest that,

under certain conditions, costly punishment may be used to help regulate the behavior of members of the group.

Our results affirm previous theorizing on costly punishment and on the role of hierarchical structures in cooperative communities, both of which are believed to serve group-regulatory purposes. It is known that costly punishment can help promote cooperation in groups by discouraging bad behavior (e.g., Boyd et al., 2003; Fehr & Gächter, 2002). Similarly, existing theories suggest that hierarchical social structures can encourage cooperation by helping to coordinate collective action (Glowacki & von Rueden, 2015; Van Vugt, 2006). We now have evidence that a sense of authority may play a critical role in eliciting the sorts of group-regulatory efforts that costly punishment may have evolved to serve.

This research further supports the general conclusion that young children enact costly third-party punishment. Previous research has found evidence of costly punishment among children age 6 but not younger (Jordan et al., 2014; McAuliffe et al., 2015; Salali et al., 2015). Our findings are consistent with this in that we found an increase in rates of costly punishment across this age range, but even the youngest children engaged in some costly punishment behavior as well. Even among the older cohort, rates of punishment were significantly higher than those observed in previous research. For example, Jordan et al. (2014) observed rates of costly punishment among 6-year-olds of approximately 30%. By contrast, the older cohort in our sample punished at more than twice that rate (74%). One possible explanation for this is that the more naturalistic aspect of our procedures allowed for greater comprehension and psychological realism across both the older and younger cohorts.

But although older children clearly demonstrated high rates of punishment overall, they showed no differences in their rates of

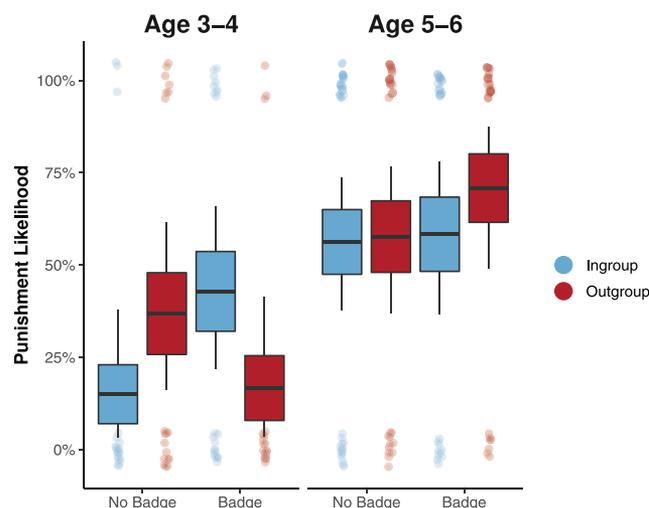


Figure 3. Costly punishment (y-axis) as a function of age, group membership, and role ($n = 190$). Boxes: center lines represent group means; edges, *SEM*; whiskers, Clopper-Pearson 95% confidence intervals. Dots: jittered individual participants by group and badge condition. Younger (age 3–4), but not older (age 5–6) children punish in-group members more harshly than out-group when placed in a position of authority; older children show no such effect. * $p < .05$. See the online article for the color version of this figure.

punishment according to either authority or group membership. While we did expect a general interaction between authority and group condition, we did not expect to observe this only in the younger cohort, so any explanation for this effect is post hoc. However, we may speculate on the reasons for this phenomenon. One possible reason concerns the development of generalizable rules regarding justice and fairness. Past research has shown that children around age 5 begin to develop a stronger concept of moral principles that can be abstracted from the exigencies of their current situation (e.g., Best et al., 2009; Bunge & Zelazo, 2006; Kohlberg, 1981). If this is the case, then older children might be less susceptible to local, contextual cues such as group membership and authority when administering punishment, and thus have administered punishment equally across conditions. Indeed, other research has shown that greater capacity for deliberation increases egalitarian punishment tendencies in adults (Yudkin et al., 2016).

Another, more mundane explanation for the lack of effect among the older cohort is that they did not find the group or authority manipulations as compelling. However, there were no age-based differences in manipulation or comprehension checks of either manipulation, which is not consistent with this possibility. Yet, it remains possible that there were differences in the psychological reality of these manipulations that were not captured by our questions that ultimately impacted (or failed to impact) participants' behavior. Further research should aim to better understand the different ways in which group membership, authority, and age influence rates of costly punishment across childhood.

Our research also speaks to ongoing debates regarding the motivations underlying punishment. Several articles have cast doubt on whether people ever truly enact costly third-party punishment (Pedersen et al., 2013; Pedersen et al., 2018). On this view, punishment is primarily motivated not by the intrinsic desire to punish bad behavior but rather by the desire to attain benefits for the self through, for instance, enhancing one's reputation (Jordan et al., 2016; Kurzban et al., 2007; Nelissen, 2008). Although we cannot speak directly to the extent that reputational motivations played a role in children's desire to punish, previous research has documented that children's behavior is motivated by reputational concerns around age 5, and not earlier (Engelmann, Herrmann, & Tomasello, 2012, 2018; Engelmann & Rapp, 2018; Leimgruber, Shaw, Santos, & Olson, 2012; Warneken & Tomasello, 2013). Children in our study not only engaged in costly punishment before this age, but also explicitly cited the transgressor's behavior as their motive for punishing. This implies that reputational concerns may not be the only factors motivating costly punishment in humans.

If reputational motivations are not the only driving factors here, then what is? Past research presents three important possibilities. First, there are pedagogical considerations (also known as deterrence motives; see Carlsmith, Darley, & Robinson, 2002). In other words, children may have punished because they wanted the transgressor to change their behavior or learn a lesson. Indeed, this is supported by our data: moderate correlations (of around r of .2 or .3) emerged between measures of these motivations and rates of punishment, and children explicitly mentioned the transgressor changing her behavior as one of their motives. Next, there are retributive motivations. Children may have punished simply out of an intuitive desire to ensure the transgressor received her just deserts. Unfortunately, because the experimental design explicitly

informed participants that the punisher would be told that the slide was closed "[b]ecause of her behavior," we cannot disentangle the former from the latter motivation in this experimental design. However, ongoing research suggests that this, too, plays an important role in costly punishment in children (Marshall, Yudkin, & Crockett, 2019).

A final possible motivation is that children punished in order to avoid interacting with the transgressor or to avoid harm to themselves. Indeed, prior research suggests that people may enact third-party punishment in order to deter transgressors from harming them directly (Delton & Krasnow, 2017; Krasnow et al., 2016). However, there are several facts that speak against this explanation. First, once participants had finished their own drawing, the experimenter instructed children that they would be placing their drawing in the book to "keep it safe." Thus, children believed the transgressor would not have access to their drawing. Second, the experimental script stated that the transgressor was going to be coming back "later" in the day, after children had left the play area. Thus, there was no reason to believe they would be interacting with the transgressor directly. Third, an analysis of participants' verbal explanations for their punishment decisions revealed that no children referenced future interaction with the transgressor as a reason for enacting punishment. Fourth, such an explanation could not account for the correlations between rates of punishment and children's reports that it was important for the transgressor to change his or her behavior and to learn a lesson.

On the contrary, we did observe correlations between children's rate of punishment and their indication that they did not want to play with the transgressor ($r = .43, p < .001$). This would be consistent with a self-preservation account of punishment. However, it is also consistent with the possibility that kids simply did not want to play with morally bad actors (Hamlin et al., 2011). Overall, the evidence is insufficient in this experiment to conclusively demonstrate the necessity or sufficiency of any of the motivations described above. More research will be needed to fully disentangle these diverging accounts.

There are several limitations of the present design. One possible concern relates to demand characteristics: due to the constraints of our testing environment (children could not be left alone in a room in the museum), we were obligated to have an experimenter present during the experiment. In Study 2, we attempted to mitigate this concern by including a second experimenter. Nevertheless, it is possible that rates of punishment would be lower in an experimental context in which no adult is present. Even so, we note that demand effects cannot explain condition-based differences in punishment, since they were equally present across conditions. Thus, while it will be important for future research to examine rates of costly punishment in contexts completely free of demand, there is no reason to expect this would impact the between-condition effects.

Another limitation concerns the manipulation of group membership. It is possible that, instead of manipulating group membership per se, our procedure involving the different museum memberships and books impacted participants' sense of similarity with the transgressor. While these concepts are closely related, the distinction is important, as past research has shown that similarity and group membership breed different responses in the context of costly punishment (Mussweiler & Ockenfels, 2013). Our experiment included no direct way of differentiating between these

constructs, and so we cannot rule out the possibility that similarity, not group membership, was responsible for the effects. Since museum membership explicitly entails a reference to belonging to a community, however, and no references were made to specific qualities of the transgressors (aside from their membership to the museum), the latter appears less likely. We also note that we did not assess individual differences with regard to participants' identification with the museum, so it is likely that the group membership manipulation had effects of varying magnitude according to participants' specific relationship to the museum (Tajfel & Turner, 1986). We recognize this as a limitation of the current design, and suggest that future research in this vein conduct a more thorough assessment of the precise impact of the group membership manipulation on each individual in order to obtain a more precise understanding of its overall effects. Nevertheless, research on the effect of minimal groups in young children suggests that the manipulation of museum membership could be effective among our participants, even if they did not feel a sense of connection to the museum ahead of time (Dunham, Baron, & Carey, 2011)

Another interesting observation taken from this data is the different patterns of punishment according to whether the punishment was costly or noncostly. One reason for this is that costly punishment may send a particularly strong signal that the punisher disapproved of the transgression, since the punisher was willing to incur a personal cost to sanction it. As people may be sensitive to punishment's communicative intent (Funk, McGeer, & Gollwitzer, 2014), costly punishment may serve as a more effective means of group regulation (because it sends such a strong signal of disapproval)—thus making it more sensitive to the manipulations of group membership and authority. Another, more mundane reason why different patterns were observed for costly versus noncostly punishment regards order effects: because children always enacted costly punishment before enacting noncostly punishment, they may have already had a chance to act in accordance with any group-regulatory motivations that were elicited by the group and authority manipulations, thereby reducing their sensitivity to these measures.

Coda

By regulating the behavior of individuals not directly related to the self, costly punishment is believed to underlie more abstract notions of justice and fairness. For instance, the fact that members of a society will endure considerable inconvenience and effort to ensure just desserts for transgressors with whom they have had no direct involvement is the basis of the modern legal system. And a willingness to intervene against injustice, even at considerable personal sacrifice, animates many of the organizations fighting for social justice. Here we find that a variety of factors influence punishment decisions in children, including group membership and authority. The fact that young children police group members at personal cost suggests that the underpinnings of such vital human institutions emerge early in life.

Context of the Research

This research is part of a larger effort by the authors to understand the psychological origins of intergroup relations and moral judgment. The research began as an outgrowth of prior research by

the first author showing that reflexive judgment leads to intergroup bias in third-party punishment (Yudkin et al., 2016), and by the third author that moral judgments incorporate information about group memberships from early in development (Rhodes, 2012, 2013). In future work, we plan to further investigate the underlying motivations of costly punishment, particularly regarding retributive versus communicative motives. Do children punish transgressors even when they do not believe the transgressor will learn a lesson? In other research, the authors are examining the ideological underpinnings of costly punishment: specifically, the extent to which conservatism versus liberalism in both adults and children impacts punishment of in-group and out-group members.

References

- Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 15924–15927. <http://dx.doi.org/10.1073/pnas.1413170111>
- Balliet, D., Mulder, L. B., & Van Lange, P. A. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*, 594–615. <http://dx.doi.org/10.1037/a0023489>
- Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *The American Economic Review*, *96*, 217–221. <http://dx.doi.org/10.1257/000282806777212594>
- Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*, 180–200. <http://dx.doi.org/10.1016/j.dr.2009.05.002>
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 3531–3535. <http://dx.doi.org/10.1073/pnas.0630443100>
- Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*, 118–121. <http://dx.doi.org/10.1111/j.0963-7214.2006.00419.x>
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299. <http://dx.doi.org/10.1037/0022-3514.83.2.284>
- Chernyak, N., & Sobel, D. M. (2016). Equal but not always fair: Value-laden sharing in preschool-aged children. *Social Development*, *25*, 340–351. <http://dx.doi.org/10.1111/sode.12136>
- Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General*, *143*, 2279–2286. <http://dx.doi.org/10.1037/xge0000018>
- Delton, A. W., & Krasnow, M. M. (2017). The psychology of deterrence explains why group membership matters for third-party punishment. *Evolution and Human Behavior*, *38*, 734–743. <http://dx.doi.org/10.1016/j.evolhumbehav.2017.07.003>
- dos Santos, M., Rankin, D. J., & Wedekind, C. (2011). The evolution of punishment through reputation. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*, 371–377.
- Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of “minimal” group affiliations in children. *Child Development*, *82*, 793–811. <http://dx.doi.org/10.1111/j.1467-8624.2011.01577.x>
- Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE*, *7*, e48433. <http://dx.doi.org/10.1371/journal.pone.0048433>
- Engelmann, J. M., Herrmann, E., & Tomasello, M. (2016). Preschoolers affect others' reputations through prosocial gossip. *British Journal of Developmental Psychology*, *34*, 447–460. <http://dx.doi.org/10.1111/bjdp.12143>

- Engelmann, J. M., Herrmann, E., & Tomasello, M. (2018). Concern for group reputation increases prosociality in young children. *Psychological Science*, 29, 181–190.
- Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, 20, 92–95. <http://dx.doi.org/10.1016/j.copsyc.2017.08.024>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fearon, J. D., & Laitin, D. D. (1996). Explaining interethnic cooperation. *The American Political Science Review*, 90, 715–735. <http://dx.doi.org/10.2307/2945838>
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454, 1079–1083. <http://dx.doi.org/10.1038/nature07155>
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140. <http://dx.doi.org/10.1038/415137a>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2004). Determining sample sizes needed to detect a difference between two proportions. In J. L. Fleiss, B. Levin, & M. C. Paik (Eds.), *Statistical methods for rates and proportions* (3rd ed., pp. 64–85). Hoboken, NJ: John Wiley & Sons.
- Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message. *Personality and Social Psychology Bulletin*, 40, 986–997. <http://dx.doi.org/10.1177/0146167214533130>
- Glowacki, L., von Rueden, C. (2015). Leadership solves collective action problems in small-scale societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 1683. <http://dx.doi.org/10.1098/rstb.2015.0010>
- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 19931–19936. <http://dx.doi.org/10.1073/pnas.1110306108>
- Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., & Sigmund, K. (2007). Via freedom to coercion: The emergence of costly punishment. *Science*, 316, 1905–1907. <http://dx.doi.org/10.1126/science.1141588>
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770. <http://dx.doi.org/10.1126/science.1127333>
- Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530, 473–476. <http://dx.doi.org/10.1038/nature16981>
- Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 12710–12715. <http://dx.doi.org/10.1073/pnas.1402280111>
- Karniol, R., Galili, L., Shtilerman, D., Naim, R., Stern, K., Manjoch, H., & Silverman, R. (2011). Why superman can wait: Cognitive self-transformation in the delay of gratification paradigm. *Journal of Clinical Child and Adolescent Psychology*, 40, 307–317. <http://dx.doi.org/10.1080/15374416.2011.546040>
- Kenward, B., & Östh, T. (2012). Enactment of third-party punishment by 4-year-olds. *Frontiers in Psychology*, 3, 373. <http://dx.doi.org/10.3389/fpsyg.2012.00373>
- Kohlberg, L. (1981). *Essays on moral development: Vol. 2. The psychology of moral development*. San Francisco, CA: Harper & Row.
- Kosfeld, M., & Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in Ethiopia. *The American Economic Review*, 105, 747–783. <http://dx.doi.org/10.1257/aer.20120700>
- Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for? *PLoS ONE*, 7, e45662. <http://dx.doi.org/10.1371/journal.pone.0045662>
- Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science*, 27, 405–418. <http://dx.doi.org/10.1177/0956797615624469>
- Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75–84. <http://dx.doi.org/10.1016/j.evolhumbehav.2006.06.001>
- Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PLoS ONE*, 7, e48292. <http://dx.doi.org/10.1371/journal.pone.0048292>
- Marshall, J., Yudkin, D. A., & Crockett, M. (2019). *Children enact costly punishment both for retributive and communicative purposes*. Manuscript in preparation.
- Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 11375–11380. <http://dx.doi.org/10.1073/pnas.1105604108>
- McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, 1, 0042.
- McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, 134, 1–10. <http://dx.doi.org/10.1016/j.cognition.2014.08.013>
- Mussweiler, T., & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 19318–19323. <http://dx.doi.org/10.1073/pnas.1215443110>
- Nelissen, R. M. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, 29, 242–248. <http://dx.doi.org/10.1016/j.evolhumbehav.2008.01.001>
- Pedersen, E. J., Kurzban, R., McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122723.
- Pedersen, E. J., McAuliffe, W. H. B., & McCullough, M. E. (2018). The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically. *Journal of Experimental Psychology: General*, 147, 514–544. <http://dx.doi.org/10.1037/xge0000410>
- Piazza, J., & Bering, J. M. (2008). The effects of perceived anonymity on altruistic punishment. *Evolutionary Psychology*, 6, 487–501.
- Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology*, 25, 1731–1735.
- Rhodes, M. (2012). Naïve theories of social groups. *Child Development*, 83, 1900–1916. <http://dx.doi.org/10.1111/j.1467-8624.2012.01835.x>
- Rhodes, M. (2013). How two intuitive theories shape the development of social categorization. *Child Development Perspectives*, 7, 12–16. <http://dx.doi.org/10.1111/cdep.12007>
- Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, 24, 999–1006.
- Rossi, P. H., Waite, E., Bose, C. E., & Berk, R. E. (1974). The seriousness of crimes: Normative structure and individual differences. *American Sociological Review*, 39, 224–237. <http://dx.doi.org/10.2307/2094234>
- Salali, G. D., Juda, M., & Henrich, J. (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior*, 36, 86–94. <http://dx.doi.org/10.1016/j.evolhumbehav.2014.09.004>
- Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, 124, 325–333. <http://dx.doi.org/10.1016/j.cognition.2012.06.004>
- Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies. *Evolution and Human Behavior*, 25, 379–393. <http://dx.doi.org/10.1016/j.evolhumbehav.2004.08.001>
- Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Hall.

- Tasimi, A., & Wynn, K. (2016). Costly rejection of wrongdoers by infants and children. *Cognition, 151*, 76–79. <http://dx.doi.org/10.1016/j.cognition.2016.03.004>
- Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*(Pt. 1), 124–130. <http://dx.doi.org/10.1348/026151010X532888>
- Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review, 10*, 354–371. http://dx.doi.org/10.1207/s15327957pspr1004_5
- Warneken, F., & Tomasello, M. (2013). Parental presence and encouragement do not influence helping in young children. *Infancy, 18*, 345–368. <http://dx.doi.org/10.1111/j.1532-7078.2012.00120.x>
- Wellman, H. M. (1992). *The child's theory of mind*. Cambridge, MA: The MIT Press.
- Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of Experimental Psychology: General, 145*, 1448–1459. <http://dx.doi.org/10.1037/xge0000190>

Received August 1, 2018

Revision received March 11, 2019

Accepted March 26, 2019 ■