

Original Articles

The development of corporal third-party punishment

Julia Marshall*, Anton Gollwitzer, Karen Wynn, Paul Bloom

Yale University, United States



ARTICLE INFO

Keywords:
Development
Punishment
Violence
Aggression

ABSTRACT

Previous research has demonstrated that toddlers are willing to punish those who harm others. This work, however, has predominantly focused on punishment in the form of resource reduction—taking away a resource or withholding access to a resource from an antisocial other. Here, in two studies, we examined whether 4- to 7-year-old children ($N = 141$) engage in direct, corporal punishment against antisocial others in third-party contexts. Children were given the opportunity to press buttons so that antisocial and prosocial puppets would be hit with a hammer. In Study 1, younger children (~4-year-olds) hit the antisocial and prosocial puppets indiscriminately, whereas older children (~7-year-olds) tended to preferentially hit the antisocial puppet. In Study 2, we tested a larger sample of 4- to 7-year-olds, and found that none of the children engaged in corporal punishment. Collapsing across both Studies 1 and 2 also indicated a null effect—children did not engage in third-party corporal punishment. We observed these findings even though children evaluated the antisocial puppet as mean and understood that pressing the hit button hurt the puppets. These findings suggest that children lack a strong desire to corporally punish third-party social wrongdoers. Our results illustrate the importance of considering different types of punishment in assessing the development of third-party punishment, and raise questions about the development of corporal third-party punishment.

Third-party punishment is an important aspect of human's psychological repertoire (Bshary & Bshary, 2010; Clutton-Brock & Parker, 1995). This type of punishment involves individuals punishing antisocial others, even though the punishers are not directly affected by the antisocial others' actions (e.g., Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Fischbacher, 2004; Fehr & Gächter, 2000; Henrich et al., 2006). Third-party punishment has been extensively documented in economic game paradigms with both children and adults (e.g., Fehr & Gächter, 2000; Fischbacher, Gächter, & Fehr, 2001; McAuliffe, Jordan, & Warneken, 2015), and it shows up in real-life cases—for example, someone may contact the authorities to report a colleague who is evading taxes, even though they are not personally harmed by this act.

Interestingly, while common in most human societies, this form of punishment is rare or non-existent in non-humans – non-human primates appear uninterested in punishing in third-party contexts (e.g., Riedl, Jensen, Call, & Tomasello, 2012). The presence of third-party punishment in humans therefore raises a question about how such punishment develops. Some have argued that third-party punishment has evolved in humans through group selection; perhaps groups with punishers do better than groups without punishers (e.g., Boyd & Richerson, 1992; Gintis, 2000; Henrich & Boyd, 2001). Alternatively, it could arise through individual-level selection; those who punish are

seen as more desirable to interact with than those who do not, and hence there is selective advantage to being seen as a punitive individual (e.g., Barclay, 2006; Jordan, Hoffman, Bloom, & Rand, 2016).

Others argue, however, that an evolved propensity to engage in third-party punishment does not exist in the first place. For instance, Guala (2012) claims that spontaneous third-party punishment is rare in small-scale societies. From this perspective, a desire for third-party punishment might arise (in some societies) through cultural learning. Or, alternatively, it might derive from an appetite for second-party punishment, caused by empathizing with the person who is victimized (Bloom, 2013). As Adam Smith (2010, pp. 98–99) put it, “When we see one man oppressed or injured by another, the sympathy which we feel with the distress of the sufferer seems to serve only to animate our fellow-feeling with his resentment against the offender. We are rejoiced to see him attack his adversary in his turn, and are eager and ready to assist him.”

One way to evaluate these arguments is to investigate the emergence of third-party punishment over the course of development. Do very young children engage in third-party punishment? Or does this behavior emerge much later in development?

There are hints of early emergence. Preverbal infants prefer individuals who punish antisocial targets relative to those who are nice to

* Corresponding author.

E-mail address: julia.marshall@yale.edu (J. Marshall).<https://doi.org/10.1016/j.cognition.2019.04.029>

Received 24 November 2018; Received in revised form 12 April 2019; Accepted 26 April 2019

Available online 23 May 2019

0010-0277/ © 2019 Elsevier B.V. All rights reserved.

antisocial targets (Hamlin, 2014; Hamlin, Wynn, Bloom, & Mahajan, 2011). And toddlers are more prone to direct negative behaviors, such as withholding or taking away resources, toward antisocial rather than prosocial characters (Hamlin et al., 2011; Riedl, Jensen, Call, & Tomasello, 2015; Van de Vondervoort, Aknin, Kushnir, Slevinsky, & Hamlin, 2018). Furthermore, Van de Vondervoort and Hamlin (2017) find that when 3-year-olds are asked “Who should get in trouble?”, they tend to select antisocial others.

Studies have also found that children appear motivated to punish others in third-party contexts even when it is costly. Yudkin, Van Bavel, and Rhodes (2019) documented that children even as young as 3 will shut down a slide to prevent an antisocial child from playing on it—even when doing so prevents the participant from playing on the slide as well. And, Marshall, Yudkin, and Crockett (submitted for publication) find that 4- to 7-year-olds will sacrifice the opportunity to play a fun iPad game to punish an antisocial peer in third-party contexts. Finally, research utilizing economic game paradigms (e.g., Fehr & Gächter, 2000; Fischbacher et al., 2001) finds that 6-year-olds (but not 5-year-olds) sacrifice their own resources (e.g., candy) to prevent a previously selfish player from receiving resources (Jordan, McAuliffe, & Warneken, 2014; McAuliffe et al., 2015).

Punishment takes many forms in the real-world, however. Most relevant for the studies presented here, adults sometimes engage in corporal punishment. As Fiske and Rai (2014, pp. 38) argue, “In most cultures, in many relationships, the most natural, intuitive satisfying punishment for the grave transgression of any relationship is corporal violence, such as flogging or execution for theft, heresy, or homicide.”

It is unclear whether children possess a desire to enact violent punishment in third-party contexts. As illustrated by the studies reviewed above, research on children’s third-party punishment desires has predominately focused on punishment in terms of resource reduction. For example, children took resources away from a mean puppet (Hamlin et al., 2011), prevented a puppet or person from receiving a resource after acting meanly or unfairly (McAuliffe et al., 2015; Riedl et al., 2015; Van de Vondervoort et al., 2018), or removed the opportunity to engage in a fun activity (Marshall et al., submitted for publication; Yudkin et al., 2019). Van de Vondervoort and Hamlin (2017) documented that children think that antisocial others should get in trouble, but the type of punishment was left unspecified. None of this work looked at corporal punishment—physical aggression against an antisocial other.

There are two studies that bear more directly on this issue. Mendes, Steinbeis, Bueno-Guerra, Call, and Singer (2017) found that 6-year-olds but not 4- or 5-year-olds will pay a cost (in the form of tokens) to watch a puppet who had previously withheld resources from the participant be physically injured. But this study only assessed second-party punishment and also did not test whether children would themselves harm an antisocial other. And, Kenward and Osth (2015) found that 5-year-olds preferentially gave antisocial (rather than neutral) characters a bad-tasting treat when given the opportunity to do so—but only when they were told that the recipient of the treats would not know that the child had allocated them. It is unclear, though, whether giving a bad-tasting treat rises to the level of direct corporal punishment.

Here, to explore whether children engage in third-party violent, corporal punishment, we conducted two studies with children ranging in age from 4 to 7 years old.¹ To do so, we developed a paradigm in which children could interact with a prosocial and an antisocial puppet. Children could press buttons on a board that allowed them to either tickle or hit the prosocial and antisocial puppets. In Study 2, we also utilized this paradigm, but additionally confirmed that all children

actually considered the antisocial puppet as mean and the prosocial puppet as nice, and that children believed that hitting the puppet actually inflicted pain.

1. Study 1

In Study 1, we developed a paradigm in which children could interact with an antisocial and a prosocial character either positively (by tickling it) or negatively (by hitting it). Using this paradigm, we investigated two questions. First, do children possess a desire to violently punish antisocial others in third-party contexts? And, second, does age moderate this effect?

1.1. Method

1.1.1. Participants

We tested children between the ages of 4 and 7 years old. We selected this age range in light of findings suggesting that children older than 6 years of age but not younger engage in second-party corporal punishment (Mendes et al., 2017). We aimed to test 50 children between the ages of 4 and 7 years of age.

We ultimately tested five additional children because of experimenter error, resulting in a final sample of 55 children ($M_{age} = 6.07$, $SD_{age} = 1.12$, range: 4.00–7.98; 25 females). Forty-two of the children were tested in local schools, 11 were tested in a lab at a northeastern university, and the remaining 2 were tested at a local museum. Although we did not collect participants’ socio-economic status information about our sample, it is likely that the children in our sample were middle-to-upper class, given the locations in which we collected our data. The administration of the study did not vary across these different testing locations. Lastly, 8 additional children were tested but excluded because they failed comprehension checks, which required children to remember which character was nice and which was mean.

Sensitivity power analyses revealed that, in collecting a final sample of 55 children, we had 80% power to detect an effect of small-to-medium size or greater (effect size $d_z = 0.38$), and 90% power to detect an effect of medium size or greater (effect size $d_z = 0.45$).

1.1.2. Materials

We built a button board (see Fig. 1). This button board included four buttons – two for each puppet; one hit button per puppet and one tickle button per puppet. Attached to each button was a cord connected to a circuit board (www.MakeyMakey.com), which was connected to a laptop computer. The MakeyMakey circuit board enabled the buttons to generate a response on the laptop screen. Both puppets were depicted on the computer screen. Depending on the button pressed, the computer played a video of the respective puppet being tickled by feathers and giggling, or being hit by a hammer and saying “Ouch!” in the voice of a real child that had been laid over the videos (view videos on OSF: https://osf.io/6u7h8/?view_only=d2aa20f4b4d245d2a8cfe1406b4323b9).

The study involved a pre-recorded puppet show. This puppet show featured three characters: a protagonist, a hitter, and a hugger. We included a protagonist puppet who was the target of the hitter’s or hugger’s actions because we wanted to remove the possibility that children would model the puppet’s actions by either hitting the hugger puppet or by tickling the hitting puppet. The puppet show also involved two scenes: one scene for the hitter puppet and another for the hugger puppet. For both the helper and hitter scenes, the protagonist puppet was the only puppet present in the video and waved back and forth and jumped up and down for several seconds. Then either the hitter or hugger entered the scene. The protagonist and the hitter or hugger puppet looked at one another several times; then the puppet either hit or hugged the protagonist puppet. Which scene came first – either the hitter or hugger puppet – was counter-balanced across participants.

¹ We conducted two additional studies prior to the two studies presented here. The results of these studies align with the data presented in the main manuscript although their methods are different. See [Supplemental Analyses](#) for details on these studies.

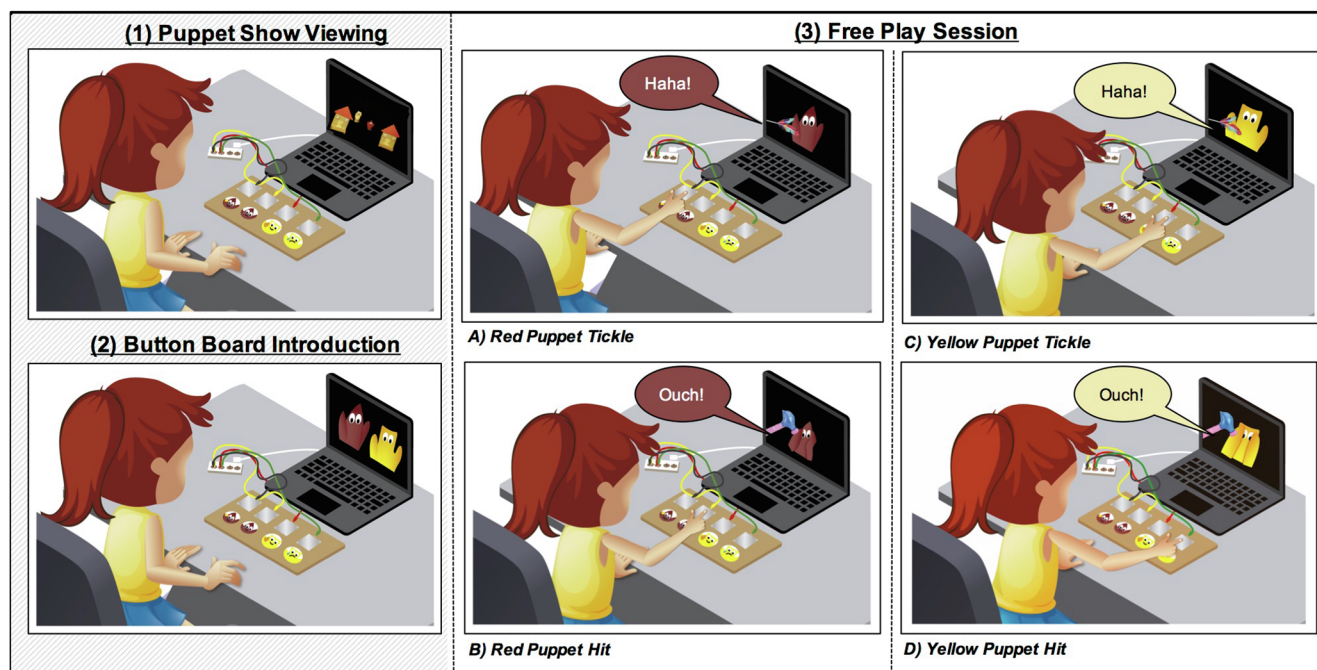


Fig. 1. Materials for presented studies. Children were shown a puppet show depicting an antisocial and prosocial puppet on a laptop (see panel [1]). Children were then introduced to a button-board, which allowed them to interact with each puppet, by an experimenter (see panel [2]). Children could thereafter tickle or hit each of the puppets as often as they chose via the button board (see panel [3]). The voice of the puppets were real children's voices expressing that they were hurt ["Ouch!"] or that they were happy [laughter]). Children activated the board by pressing the metal buttons.

1.1.3. Procedure

The experimenter first told children that they were going to learn about a special button machine (see Fig. 1). The experimenter described what each of the buttons did in a demonstration and then prompted the child to use the button board one time to learn how the buttons worked: "Here are four buttons. When you press them, you have to tap lightly and lift your finger. Then a video will play on the screen, and you have to watch the whole thing until you want to press the next button. Can you try?" The experimenter ensured that the child pressed one of the buttons, but no video accompanied the button press at this point. The experimenter then said, "These pictures [pointing to the individual pictures] tell you what the buttons do: If you tap this button, you hurt the [yellow/red] puppet and make it sad. If you tap this button, you tickle the [yellow/red] puppet and make it happy. If you tap this button, you hurt the [red/yellow] puppet and make it sad. If you tap this button, you tickle the [red/yellow] puppet and make it happy. Now do you want to give it a try?" The experimenter allowed the child to press each of the four buttons, playing the associated videos.

The experimenter then told the participant that they were going to learn about how the puppets act in real life. Here, the child and the experimenter together watched the puppet show. Once the puppet show was finished, the experimenter told the participant: "Now, this is the puppet who [hit/hugged] the blue puppet (pointing to either the red or yellow puppet), and this is the puppet who [hugged/hit] the blue puppet (pointing to either the red or yellow puppet). You get to use the buttons for real and, this time, the puppets will really get hurt and really get tickled! You can press the buttons no times, 1 time, 2 times, or however many times you want. Let me know when you're done."

After the participant indicated he or she was finished, the study ended with a set of *comprehension questions*: (1) Who hit the blue puppet? (2) Who hugged the blue puppet? If a child incorrectly responded to either of these questions, they were removed from analyses.²

1.2. Results

We examined children's overall button press behavior. To do so, we conducted a Poisson loglinear generalized estimating equation (GEE) with Puppet Character (prosocial, antisocial) and Button Action (tickle, hit) as within-subjects factors. We opted for this model because we are utilizing count data within a repeated-measures design. The analysis did not reveal a main effect of either Puppet Character or Button Action (p s > 0.419). The Puppet Character $\times$ Button Action interaction was also not significant, $\chi^2(1, N = 55) = 0.95, p = .330; \eta_p^2 = 0.02, 95\% \text{ CI } [0.00, 0.13]$.³

We then examined whether age moderated our effects. To do so, we conducted the same model as previously but added age as a continuous predictor. We found a Puppet Character $\times$ Button Action $\times$ Age interaction, $\chi^2(1, N = 55) = 4.80, p = .028; \eta_p^2 = 0.07, 95\% \text{ CI } [0.00, 0.22]$. Given this interaction, we examined the Puppet Character $\times$ Button Action interaction at younger ages (-1.5 SD ; Age: 4.38) and at older ages ($+1.5 \text{ SD}$; Age: 7.75). The Puppet Character $\times$ Button Action interaction for older children was not significant (although it approached significance), $\chi^2(3, N = 55) = 7.67,$

(footnote continued)

and 7-year-olds) two additional sets of questions for exploratory purposes. Specifically, we asked: (1) two *forced-choice questions* (i.e., "If you could only press one button for this puppet [pointing to one of the puppets], which button would you press?"), and (2) two *button-belief questions* (i.e., "Do you think this button and this button [pointing to the two hitting buttons] really hurt the puppets? Yes or no?"). Because we did not ask all children these questions, these analyses are included in Supplemental Analyses. We ask these questions to all children in Study 2, however.

³ We chose to report η_p^2 effect sizes (calculated via repeated measures GLMs) rather than Odds (calculated using GEEpack in R [see Hothorn & Everitt, 2014]). We did so because we find η_p^2 values to be more informative in terms of interpreting the effect sizes of interaction effects. Additionally, it is important to note that η_p^2 effect sizes cannot be negative because they are squared values; therefore, the inclusion of zero in the confidence interval signifies a non-significant effect.

² At the end of the procedure, we asked only 4- and 5-year-old children (not 6-

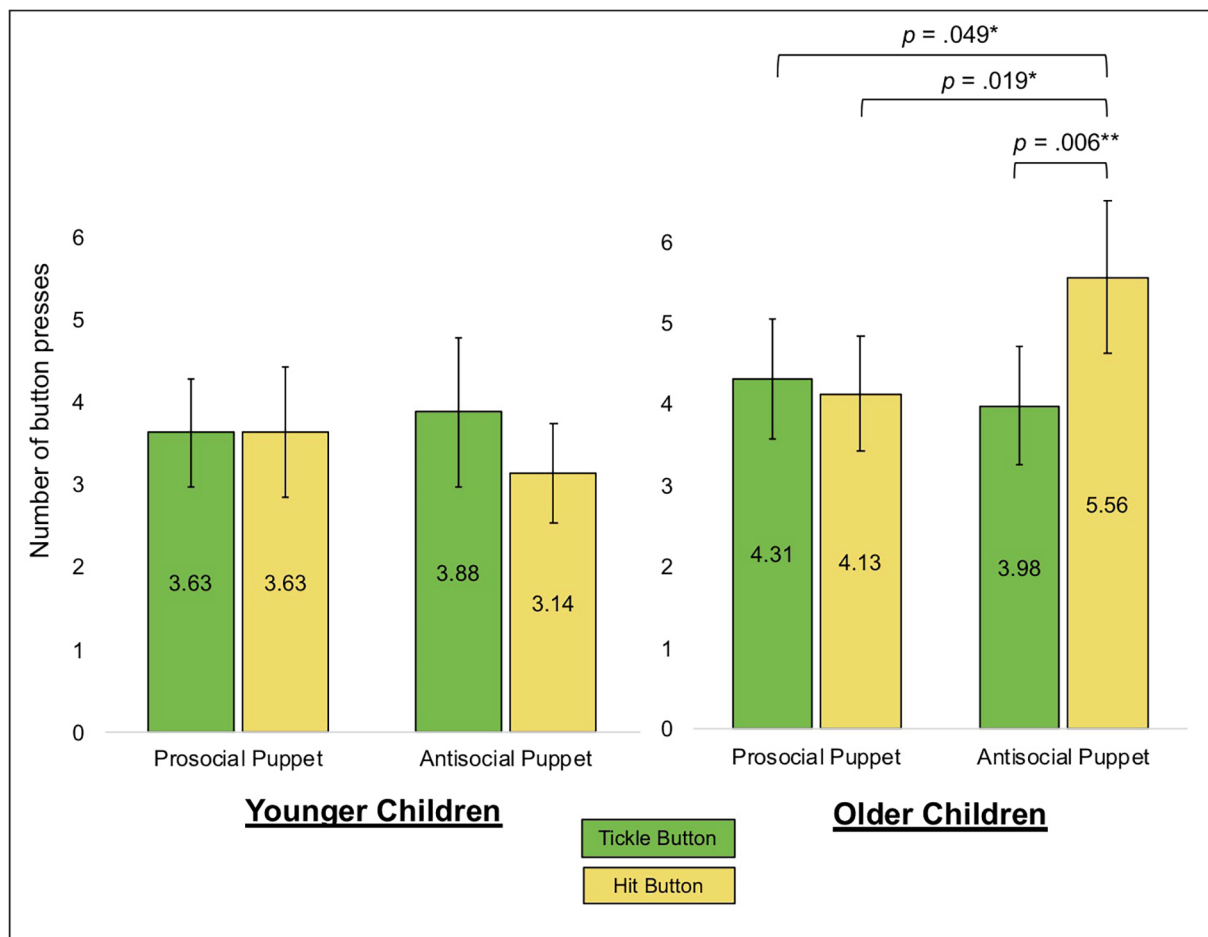


Fig. 2. Younger and older children's number of button presses in Study 1 depending on puppet; $^*p < .05$, $^{**}p < .01$; errors bars represent ± 1 SEM.

$p = .053$. We did not find a Puppet Character $\times$ Button Action interaction for younger children, $\chi^2(3, N = 55) = 2.37, p = .498$. See Fig. 2.

We also examined the pairwise comparisons within both older and younger children. Older children tended to select the hit button more than the tickle one for the antisocial puppet, $\chi^2(1, N = 55) = 7.63, p = .006; d = 0.26, 95\% \text{ CI } [-0.12, 0.63]$.⁴ Furthermore, older children also tended to select the hit button for the antisocial puppet more than they did for the prosocial one, $\chi^2(1, N = 55) = 5.51, p = .019; d = 0.24, 95\% \text{ CI } [-0.14, 0.61]$. And, finally, older children tended to select the hit button more for the antisocial puppet than they selected the tickle button for the prosocial puppet, $\chi^2(1, N = 55) = 3.87, p = .049; d = 0.20, 95\% \text{ CI } [-0.18, 0.57]$. No other effects were significant for older children, and none of the pairwise comparisons were significant for younger ones; see Fig. 2.

1.3. Discussion

In sum, in Study 1 we observed a significant tendency for older children (~7-year-olds) to engage in corporal punishment toward an antisocial puppet, while younger children (~4-year-olds) did not.

⁴ Based on recommendations by Dunlap, Contina, Vaslow, and Burke (1996) and Lakens (2013), we opted to utilize Cohen's d effect sizes and 95% confidence intervals (rather than d_z or alternative repeated measures effect sizes). We adopted this same approach in Section 2.3.2. By doing so, the effect sizes and confidence intervals provided here are more conservative estimates. As a result, the estimates cross zero despite being significant. Confidence intervals were calculated using an online calculator provided by Uanhoro (2017).

2. Study 2

We conducted Study 2 for two reasons. First, we wanted to more thoroughly ensure that children understood the contingencies of the experiment. We were especially interested in confirming that the youngest participants understood the study, considering they acted indiscriminately toward the puppet in Study 1. To accomplish this goal, we added a set of measures in which we asked children about (a) which button they would select for each puppet if they had to select one, (b) whether they considered the actions of the prosocial and antisocial puppet nice and mean, and (c) whether they believed the buttons on the board actually affected the puppets. Second, we also aimed to replicate the effects documented in Study 1.

2.1. Participants

We tested 86 children from the ages of 4 to 7 ($M_{age} = 6.04, SD_{age} = 1.16$, range: 4.00–7.94; 38 females). A sensitivity analyses revealed that, in collecting this sample of children, we had 80% power to detect an effect of small-to-medium size or greater (effect size $d_z = 0.30$), and 90% power to detect an effect of medium size or greater (effect size $d_z = 0.35$). Sixty-two of the children were tested at a natural history museum, and 24 were tested in a lab at a northeastern university. We did not find any effects of testing location, and the study was run identically across both testing locations. Like Study 1, we did not collect participants' socio-economic status information about our sample, although it is likely that the children in our sample were middle-to-upper class, given where we collected data.

We ran an additional 15 children who were excluded (8 4- and 5-

year-olds and 7 6- and 7-year-olds). Fourteen of them were excluded for failing comprehension questions (described in Section 2.2), and one was excluded because of experimental error.⁵ None of the presented results change if we include these participants.

2.2. Materials and procedure

The materials and procedures of Study 2 were largely identical to Study 1. The only change made was that we added several additional questions after the free-play session. Specifically, the experimenter first asked two *forced choice questions* in counter-balanced order: (A) “If you could only press one button for this puppet [pointing to the antisocial puppet], which button would you press?” (B) “If you could only press one button for this puppet [pointing to the prosocial puppet], which button would you press?”. The experimenter then asked the two *comprehension questions* from Study 1 in counter-balanced order: (A) “Who hugged the blue puppet?” (B) “Who hit the blue puppet?”.

Next, the experimenter asked two *moral evaluation questions* – one for the antisocial puppet and one for the prosocial puppet – to confirm that children did indeed consider the antisocial puppet mean and the prosocial puppet nice. Specifically, the experimenter asked: “Do you think this puppet was nice or mean? How nice (mean)? Just a teeny bit nice (mean), a little bit nice (mean), or very nice (mean)?” And, finally, the experimenter asked two *button-belief questions* to confirm the participant believed pressing the buttons truly affected the puppets: (A) “Do you think this button and this button [pointing to the hit buttons] really hurt the puppets?” (B) “Do you think this button and this button [pointing to the tickle buttons] really tickle the puppets?”. If the child responded ‘yes’, the experimenter followed up by asking, “Do you think that a teeny bit or a lot?”

2.3. Results

2.3.1. Button-press behavior

We first assessed children’s overall button-press behavior. To do so, we conducted the same Poisson loglinear model as in Study 1 with Puppet Character (prosocial, antisocial) and Button Action (tickle, hit) as within-subjects factors. We did not find a main effect of either Puppet Character or of Button Action, $p_s > 0.99$. We also did not find a Puppet Character $\times$ Button Action interaction, $\chi^2(1, N = 86) = 0.02, p = .884$; $\eta_p^2 < 0.01$, 95% CI [0.00, 0.03]. See Fig. 3.

Importantly, unlike in Study 1, age did not moderate this interaction, $\chi^2(1, N = 86) = 2.25, p = .133$; $\eta_p^2 = 0.02$, 95% CI [0.00, 0.12]. Still though, we examined the Puppet Character $\times$ Button Action interaction in addition to all pairwise comparisons at younger ages (-1.5 SD; Age: 4.26) and older ones ($+1.5$ SD; Age: 7.78) in exploratory analyses. For both older and younger children, the Button Action $\times$ Puppet Character interaction was not significant, $p_s > 0.566$; furthermore, none of the pairwise comparisons were significant (all $p_s > 0.169$).

2.3.2. Button-press behavior (Study 1 and 2 combined)

Given that we found different results between Study 1 and 2, we combined the data from these two studies. The studies’ procedures were identical (the additional questions in Study 2 were presented *after* the main dependent variables). We did not find a Puppet Character $\times$ Button Action interaction, $\chi^2(1, N = 141) = 0.273, p = .601$; $\eta_p^2 < 0.01$, 95% CI [0.00, 0.04]. Furthermore, we did not

⁵ Because we tested the majority of the children at a local museum, we also had to test children older than 7 years of age because they were at the museum and wanted to participate – specifically, we tested a number of 8-year-olds ($n = 20$) and also 9-year-olds ($n = 9$). We opted to exclude those children from analyses here primarily to keep the Methods and Participants consistent with Study 1 and because the results do not change if we include them.

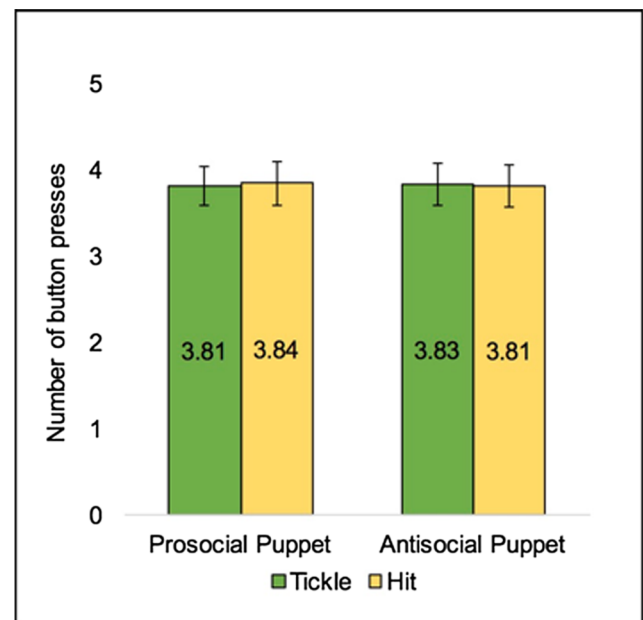


Fig. 3. Children’s aggregate button presses from Study 2 across all ages; errors bars represent ± 1 SEM.

find a Puppet Character $\times$ Button Action $\times$ Age interaction, $\chi^2(1, N = 141) = 0.01, p = .911$; $\eta_p^2 < 0.01$, 95% CI [0.00, 0.02]. Additionally, for both older and younger children, the Puppet Character $\times$ Button Action interaction was not significant, $p_s > 0.492$; furthermore, none of the pairwise comparisons were significant (all $p_s > 0.130$).

Because small samples tend to lead to increased Type II error, we also tested for null effects using Bayesian statistics using JASP (<https://jasp-stats.org/>). These analyses compare models according to a null hypothesis and an alternative hypothesis for best fit (Wagenmakers, 2007). Using this software, we calculated the Bayes factor in favor of the null hypothesis using a Bayesian Repeated Measures analysis of variance (ANOVA) with Puppet Character (prosocial, antisocial) and Button Action (hit, tickle) as within-subjects factors. The estimated Bayes factor (B01) suggested that the data were approximately 608 to 1 in favor of the null hypothesis; this was categorized as a decisive null interaction effect (Jarosz & Wiley, 2014). Similarly, the Puppet Character $\times$ Button Action $\times$ Age interaction was also categorized as a decisive null interaction (1.57e5 in favor of the null hypothesis).

In addition to examining these overall interaction effects, we also assessed the evidence in favor of the null for the two pairwise comparisons that bear most directly on whether children engage in corporal punishment. We examined the pairwise comparison between hitting the antisocial puppet (corporal punishment) and tickling the antisocial puppet (e.g., ‘prosocial behavior’) in addition to the pairwise comparison between hitting the antisocial puppet (corporal punishment) and hitting the prosocial puppet (e.g., ‘antisocial behavior’). To do so, we conducted two Bayesian t-tests using JASP. For the comparison between hitting the antisocial puppet and tickling the antisocial puppet, the estimated Bayes factor (B01) suggested that the data were 9:1 in favor of the null hypothesis; this qualifies as substantial evidence in favor of the null hypothesis (Jarosz & Wiley, 2014). Similarly, for the comparison between hitting the antisocial puppet and hitting the prosocial puppet, the estimated Bayes factor (B01) suggested that the data were 8:1 in favor of the null hypothesis; this also qualifies as substantial evidence in favor of the null hypothesis.

Furthermore, we also examined the 95% confidence intervals of the effect sizes of these two pairwise comparisons (Thompson, 2002). As illustrated in Fig. 4, we do not find evidence for a medium or large effect of children engaging in third-party corporal punishment. This is

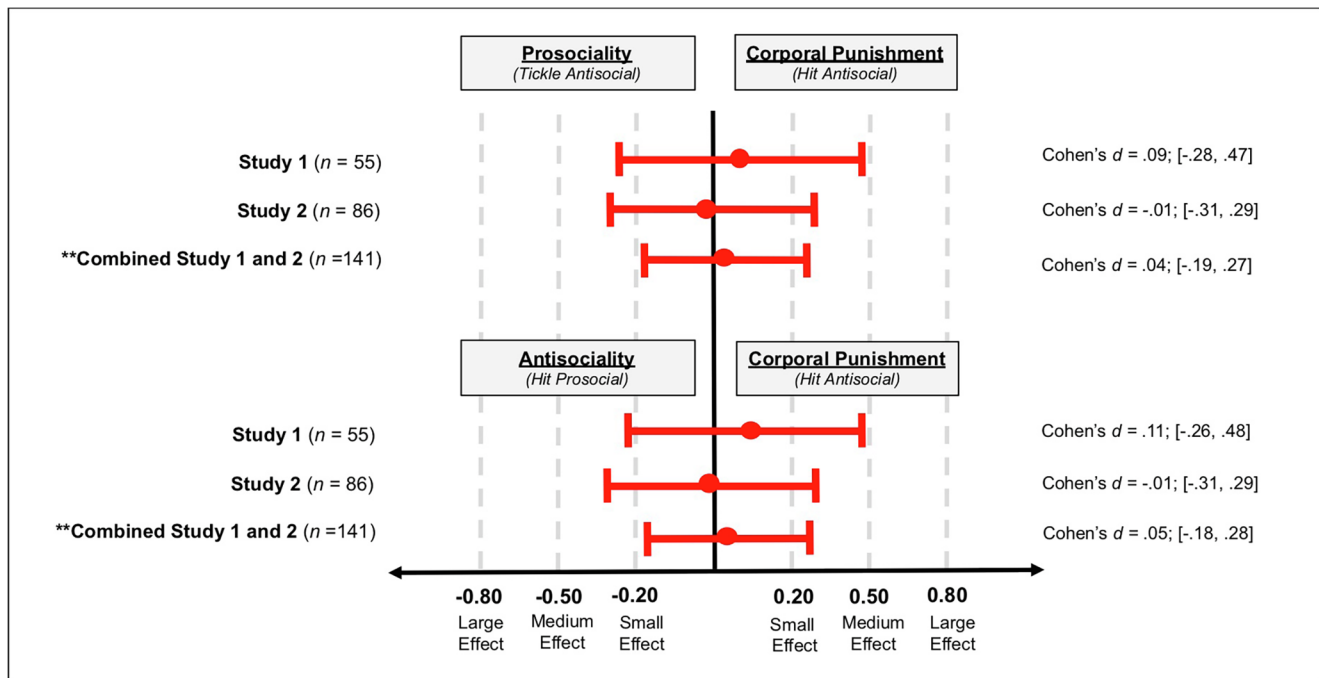


Fig. 4. 95% confidence intervals for the effect sizes documented in Study 1, Study 2, and combined Studies 1 and 2 (collapsed across age). The top component represents the pairwise comparison between tickling the antisocial puppet and hitting the antisocial puppet. The bottom component represents the pairwise comparison between hitting the prosocial puppet and hitting the antisocial puppet.

because the potential effect size of children hitting the antisocial puppet compared to tickling it ranged from -0.19 to 0.27 . And, the potential effect size of children hitting the antisocial puppet compared to hitting the prosocial one ranged from -0.18 to 0.28 .

We also conducted sensitivity power analyses of the combined samples of Study 1 and 2 ($n = 141$). These analyses indicated that we had 95% power to detect an effect of small-to-medium size or greater (effect size $d_z = 0.31$), and 80% power to detect an effect of small size or greater (effect size $d_z = 0.24$). Yet, we found no such effect. These sensitivity analyses, together with the Bayesian analyses and confidence intervals of the effect sizes, indicate that either children do not engage in third-party corporal punishment at all or that their tendency to do so is small.

2.3.3. Forced-choice responses

We then assessed children's forced-choice button responses. To do so, we coded choices to press the hit button as 0, and choices to press the tickle button as 1. We first conducted a logistic GEE analysis with Puppet Character (prosocial, antisocial) as a within-subjects factor and child's age as a continuous predictor. We did not find Puppet Character $\times$ Age interaction, $\chi^2(1, N = 86) = 1.07, p = .302$. We then collapsed across age; in doing so, we found a main effect of Puppet Character, $\chi^2(1, N = 86) = 6.91, p = .009; \eta^2 = 0.04$, 95% CI [0.01, 0.12]: Participants of all ages tended to select the tickle button (rather than the hit button) more for the prosocial puppet, $M = 0.73$, $SD = 0.45$, than for the antisocial one, $M = 0.55$, $SD = 0.50$. Binomial follow-up tests revealed that participants tended to select the tickle button more frequently than the hit button compared to chance for the prosocial puppet, $p < .001$, but they did not select either button more frequently than compared to chance for the antisocial puppet, $p = .451$. See Fig. 5.

2.3.4. Moral evaluation questions

We coded children's responses to the moral evaluation questions such that a 3 reflected children's choice that the character was very nice, 2 a little nice, and 1 a teeny bit nice. The same values (negative) represented children's choices when they indicated a character was

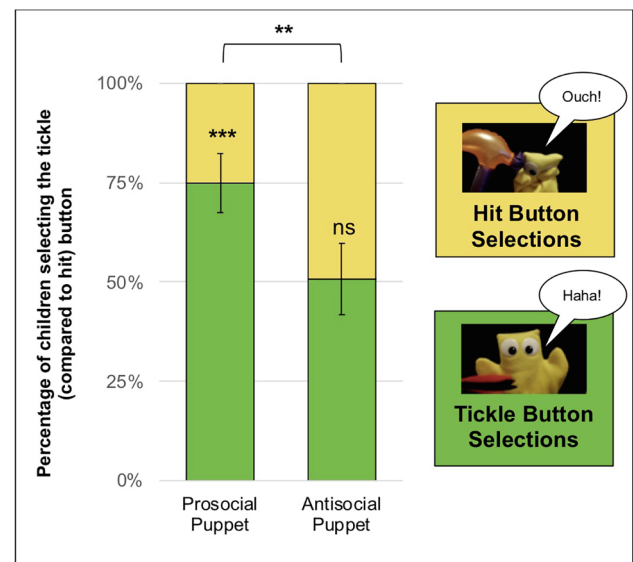


Fig. 5. Children's forced-choice responses for each puppet; errors bars represent ± 1 SEM; ** $p < .01$; *** $p < .001$.

mean. A repeated measures ANOVA with Puppet Character (prosocial, antisocial) as a within-subjects factor revealed a significant effect of Puppet Character, $F(1, 85) = 743.07, p < .001, \eta_p^2 = 0.897$, 90% CI [0.86, 0.92]. Children reported that the hitting character was very mean ($M = -2.35, SD = 1.25$) and that the hugging character was very nice ($M = 2.80, SD = 0.76$). Age did not interact with these effects, $p = .897$.

2.3.5. Button-belief questions

We also analyzed participants' responses to the button-belief questions. We created a scale from 1 to 3. Three represented that the child believed that the buttons really did hit or tickle the puppets a lot; 2 represented that the child believed that the buttons really did hit or

tickle the puppets a teeny bit; 1 represented that the child did not believe that the buttons hit or tickled the puppets. Indeed, for both the tickle and hurt button, participants tended to say that the buttons actually tickled, $M = 2.38$, $SD = 0.85$, and actually hurt the puppets, $M = 2.60$, $SD = 0.76$. Furthermore, we treated participants' responses as binary responses ("no" = 0; "yes" = 1) and conducted two one-sample t-tests to assess whether participants' button beliefs were significantly different from chance. Participants tended to believe that the tickle button was actually tickling the puppets ($M = 0.76$, $SD = 0.43$) and that the hit button was actually hurting the puppets ($M = 0.84$, $SD = 0.37$) compared to chance, both $ps < 0.001$. These data verify that children did think that pressing the buttons actually tickled and actually hurt the puppets.

These effects did not differ as a function of children's age, $ps > 0.870$. Including only children who believed the tickle button tickled and the hit button hurt in our analyses did not change the button-press behavior results reported earlier—the Puppet Character $\times$ Button Action was still not significant, $p = .774$.

2.4. Discussion

In sum, we do not find evidence that children exhibit an interest in engaging in corporal third-party punishment in third-party contexts. Importantly, these null findings are not a function of children failing to understand the contingencies of the experiment, as all children evaluated the prosocial and antisocial puppet as nice and mean, respectively, and also believed that the button board truly affected the puppets.

3. General discussion

Across two studies, we examined whether children engage in corporal third-party punishment. We find little evidence in favor of this hypothesis, at least in our paradigm. Specifically, we gave 4- to 7-year-olds the opportunity to press different buttons which either tickled or hit puppets who had acted in either an antisocial or prosocial manner. In Study 1, we found that younger children (~4-year-olds) did not exhibit a desire to corporally punish antisocial puppets, but older children (~7-year-olds) did hit the antisocial puppet more than they tickled it, and hit the antisocial puppet more than the prosocial one. In Study 2, in a larger sample, we again tested 4- to 7-year-olds and found that both younger and older children did not selectively hurt the puppet who acted antisocially. Collapsed across Studies 1 and 2, younger and older children did not exhibit a tendency to engage in corporal third-party punishment. Importantly, we observed this null effect despite that children of all ages judged the antisocial other as mean and reported that pressing the hit button actually hurt the puppets.

Children did, however, act discriminately toward the puppets in a forced-choice question context, suggesting that our methods are adequate to elicit preferential behavior. Specifically, when asked which button they would prefer to select for each puppet, we found that participants selected the tickle button more for the prosocial puppet compared to the antisocial one and the hit button more for the antisocial puppet compared to the prosocial puppet. Furthermore, we found that children across all ages pressed the tickle button for the prosocial puppet significantly more than chance, though they did not press the hit button for the antisocial puppet significantly more than chance. We interpret these findings to mean that, when forced to choose, children are nicer to prosocial others compared to antisocial ones, are meaner to antisocial others compared to prosocial ones, and are overall nice to prosocial others. This tells us that our participants were not acting randomly; they understood the actions of the puppets and responded to them appropriately.

Do these forced-choice responses provide evidence for an appetite for corporal punishment? Not necessarily. The difference in hitting frequency might not be due to children's attitudes toward the antisocial

puppet at all; children might be compelled, reasonably enough, not to hit the prosocial puppet who acted nicely. Alternatively, they may hit the antisocial puppet because they want to avoid acting nicely toward the antisocial puppet (i.e., they do not want to reward mean actions)—and hitting is the only alternative in our study to acting nicely. And, even if one were to interpret our forced-choice findings as signifying corporal punishment, the effect size documented here (Study 2 forced-choice: Cohen's d of 0.38; a small-to-medium effect) would still be markedly smaller than the average effect size documented in developmental research on resource-reduction punishment (Cohen's d of 1.09; a large effect; see [Supplemental Analyses](#) for information).

Our findings overall indicate that, although children as young as 3 and older engage in third-party punishment in resource-reduction paradigms (Marshall et al., submitted for publication; Riedl et al., 2015; Van de Vondervoort et al., 2018; Yudkin et al., 2019), children do not seem particularly interested in directly harming antisocial others—at least in the current paradigm. Hence the emergence of children's third-party punishment behavior may depend on the type of punishment, with children exhibiting a much stronger and early-emerging desire to engage in resource-reduction punishment compared to direct, corporal third-party punishment.

Why would children be so uninterested in enacting violent, corporal punishment? One possibility is that children are uninterested in acting aggressively in general. But actually, in all our experiments, children of all ages were willing to press the hit buttons. And, past research, as well as everyday observation of how toddlers interact, indicates children are often quite willing to act aggressively (Bandura, Ross, & Ross, 1961).

We find two explanations more promising. First, perhaps it is much later in development that people view hurting or injuring another person as inherently *punitive*. That is, children do not make a connection between hitting and punishing. Aligning with this possibility, Marshall and Bloom (submitted for publication) find that, when asked to spontaneously generate descriptions of how others enact third-party punishment, children rarely come up with violent actions. Furthermore, children generally do not consider retributive violence—hitting another person back in response to being hit—to be a morally appropriate course of action (Astor, 1994). And, while children in our studies frequently experience resource-reducing punishment (i.e., losing privileges) in their everyday lives, they are less likely to experience aggressive punishment, such as being hit. One interesting way of testing such a hypothesis would involve conducting our studies in a culture where hitting is a more prominent form of punishment.

Second, Fiske and Rai (2014) argue that corporal violence emerges largely out of certain types of relational dynamics often defined by authority and hierarchy. Such relationships were not present in our studies; children lacked any relationship to the puppets or any context for why they were given the opportunity to interact with the puppets. If children were given particular roles, such as being appointed to the "Puppet Police", children may indeed preferentially hurt the antisocial puppet. Relatedly, research with adults finds that people are especially motivated to engage in violence toward others in intergroup contexts (Cikara, 2015; Leach, Spears, Branscombe, & Doosje, 2003), so perhaps children too would preferentially hurt an antisocial other if the antisocial other were an out-group member. Indeed, Jordan et al. (2014) finds that children are more motivated to punish out-group members than in-group ones. All of this is grist for future research.

Both of these explanations raise a more general question about the evolutionary roots of punishment. Possibly, third-party corporal punitive desires are socially learned and manifest later in life. In favor of the social learning possibility, Salali, Juda, and Henrich (2015) find that children imitate an experimenter who engages in third-party costly punishment, and are increasingly more imitative as they age. These findings suggest that cultural learning may play an important role in the transmission of punitive norms in young children.

This is not to say that there is no evolutionary element of punitive desires. It could be that the desire to engage in corporal third-party

punishment is evolved, but just does not emerge until much later in life (e.g., in adolescence). Similarly, it could also be that the desire to engage in corporal third-party punishment is early-emerging, but by age 4, children recognize that violence is a socially inappropriate form of punishment. This may be particularly the case in the United States where children are not frequently punished for their misdeeds through corporal means. Again, future developmental research in cultures where corporal punishment is more prominent may help to clarify some of these issues.

Consider some potential objections to the studies presented here. First, one may be concerned with statistical power, especially considering we are reporting null findings. However, we had 95% power to detect an effect of small-to-medium size or greater (effect size $d_z = 0.31$) when combining Studies 1 and 2. This means that, if a corporal punishment effect of small-to-medium size (or greater) exists, we would have had the power to detect it. We only had a reasonable chance though of detecting a small effect size, though ($\sim 65\text{--}80\%$; $d_z = 0.20\text{--}0.22$). Based on this, we conclude that children either do not engage in corporal third-party punishment or that such a tendency is quite small.

Second, children may have responded differently had we utilized actual humans and not puppets. Indeed, Kenward and Osth (2015) find that 5-year-olds are willing to selectively allocate bad-tasting treats to antisocial adults. Although our results may have differed had we used humans compared to puppets, we opted to use puppets for two reasons. First, using real humans would have raised potential ethical issues, given that we were specifically interested in whether children would directly physically harm antisocial others (Fiske & Rai, 2014). Second, we aimed to compare the present results to past work studying resource-reducing punishment (Hamlin et al., 2011; Riedl et al., 2015) and second-party corporal punishment (Mendes et al., 2017), which utilized puppets in their paradigms. Still, it would be interesting to see what difference using actual people would make.

In a similar vein, it is possible that different contexts may elicit corporal third-party punishment. For instance, children may be more inclined to engage in corporal third-party punishment in contexts where the moral violation is more severe, such as a situation where the antisocial puppet consistently hits others. Furthermore, we only ever offer one corporal punishment option – hitting with a hammer. Perhaps children would engage in corporal punishment of another type, such as endorsing a slap on the wrist for hitting another puppet. Future studies can investigate these possibilities.

Beyond these methodological limitations, another qualification of the present studies arises from our exclusive focus on third-party punishment. It remains to be seen whether children engage in second-party corporal punishment specifically in our paradigm. Aligning with the arguments put forth here, there is evidence to suggest that children younger than 6 years of age would *not* engage in corporal second-party punishment. This is because Mendes et al. (2017) found that only children 6 and older will engage in violent punishment when they are personally affected. Still though, future research should investigate whether children of different ages engage in corporal punishment in second-party contexts utilizing our paradigm.

One final concern relates to generalizability. The sample included largely White children presumably from middle to upper class families. Indeed, many studies have documented a negative relationship between socioeconomic status and corporal punishment such that families of lower socio-economic status use corporal punishment more frequently (see Gershoff, 2002 for review).

In sum, our studies provide evidence that the desire to engage in corporal third-party punishment is not particularly potent in children. Our studies illustrate the heterogeneity of third-party punishment behaviors and the importance of assessing different types of punishment actions in the context of developmental research. Future work should take the form of punishment into account (e.g., resource-reducing, corporal) when examining the development of third-party punishment.

Acknowledgements

The authors would like to thank the Mind and Development lab for helpful feedback on the design and methodology of the present studies, and Emmett Larson, Maya Rodriguez, Rob Henderson, Karli Cecil, and Arianna Hairston for assistance with data collection. We would also like to thank the children, parents, and schools who participated in this research, and Clara Colombatto for her invaluable technical assistance. This work was funded by the Klaus J. Jacobs Research Prize awarded by the Jacobs Foundation.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2019.04.029>.

References

- Astor, R. A. (1994). Children's moral reasoning about family and peer violence: The role of provocation and retribution. *Child Development*, 65, 1054–1067.
- Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, 63, 575–582.
- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27, 325–344.
- Bloom, P. (2013). *Just babies: The origins of good and evil*. New York, NY: Crown.
- 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.
- Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–195.
- Bshary, A., & Bshary, R. (2010). Self-serving punishment of a common enemy creates a public good in reef fishes. *Current Biology*, 20, 2032–2035.
- Cikara, M. (2015). Intergroup schadenfreude: Motivating participation in collective violence. *Current Opinion in Behavioral Sciences*, 3, 12–17.
- Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, 373, 209–216. <https://doi.org/10.1038/373209a0>.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170–177. <https://doi.org/10.1037/1082-989X.1.2.170>.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25, 63–87.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public good experiments. *The American Economic Review*, 90, 980–994.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters*, 71, 397–404.
- Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.
- Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review. *Psychological Bulletin*, 128, 539–579.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Guala, F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35, 1–15.
- Hamlin, J. K. (2014). Context-dependent social evaluation in 4.5-month-old human infants: the role of domain-general versus domain-specific processes in the development of social evaluation. *Frontiers in Psychology*, 5, 614.
- 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*, 108, 19931–19936.
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208, 79–89.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... Ziker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.
- Hothorn, T., & Everitt, B. S. (2014). *A handbook of statistical analyses using R*.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7, 2–9.
- Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530, 473–476.
- Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences*, 111, 12710–12715.
- Kenward, B., & Osth, T. (2015). Five-year-olds punish antisocial adults. *Aggressive Behavior*, 41, 413–420.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of Personality and Social Psychology*, 84, 932–943.

- Marshall, J. & Bloom, P. (2019). Noble but naïve: children's early reasoning about obligations to punish (submitted for publication).
- Marshall, Yudkin, & Crockett (2019). How consequential and retributive concerns shape costly third-party punishment in young children (submitted for publication).
- McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, 134, 1–10.
- Mendes, N., Steinbeis, N., Bueno-Guerra, N., Call, J., & Singer, T. (2017). Preschool children and chimpanzees incur costs to watch punishment of antisocial others. *Nature Human Behaviour*, 2, 45–51.
- Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences*, 109, 14824–14829.
- Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology*, 25, 1731–1735.
- Salali, G. D., Juda, M., & Henrich, J. (2015). Transmission and development of costly punishment in children. *Evolution and Human Behavior*, 36, 86–94.
- Smith, A. (2010). *The theory of moral sentiments*. Penguin.
- Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31, 25–32.
- Uanhoro, J. O. (2017). Effect size calculators. Available online at: <https://effect-size-calculator.herokuapp.com/>.
- Van de Vondervoort, J. W., Aknin, L. B., Kushnir, T., Slevinsky, J., & Hamlin, J. K. (2018). Selectivity in toddlers' behavioral and emotional reactions to prosocial and antisocial others. *Developmental Psychology*, 54, 1–14.
- Van de Vondervoort, J. W., & Hamlin, J. K. (2017). Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *Journal of Experimental Child Psychology*, 164, 136–151.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.
- Yudkin, D., Van Bavel, J., & Rhodes, M. (2019). Young children police in-group members at a personal cost. *Journal of Experimental Psychology: General*.