nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00975-9

Check for updates

Children punish third parties to satisfy both consequentialist and retributive motives

Julia Marshall[1]✉, Daniel A. Yudkin[1,2] and Molly J. Crockett[1]✉

Adults punish moral transgressions to satisfy both retributive motives (such as wanting antisocial others to receive their 'just deserts') and consequentialist motives (such as teaching transgressors that their behaviour is inappropriate). Here, we investigated whether retributive and consequentialist motives for punishment are present in children approximately between the ages of five and seven. In two preregistered studies ($N = 251$), children were given the opportunity to punish a transgressor at a cost to themselves. Punishment either exclusively satisfied retributive motives by only inflicting harm on the transgressor, or additionally satisfied consequentialist motives by teaching the transgressor a lesson. We found that children punished when doing so satisfied only retributive motives, and punished considerably more when doing so also satisfied consequentialist motives. Together, these findings provide evidence for the presence of both retributive and consequentialist motives in young children.

Punishment is considered a hallmark of our moral psychological repertoire[1]. Across cultures, people believe that those perceived as violating rules, laws, or norms should be punished with negative sanctions, through monetary fines, physical or emotional pain, or even death, depending on the severity of the transgression[2,3]. Individuals are motivated to impose such negative sanctions on transgressors even at a personal cost, both in direct retaliation for being harmed via 'second-party punishment'[4] and in response to merely witnessing a transgression via 'third-party punishment'[5,6]. Researchers have argued that such costly punishment behaviour may play a vital role in sustaining cooperation by deterring antisocial behaviour[7–9].

Philosophical theories of justice highlight two possible proximate motives for costly punishment behaviour[3]. On one hand, people may be motivated by retributive concerns[10], such as wanting antisocial others to suffer as a form of 'just desert'. On the other hand, people may be motivated by consequentialist concerns[11], such as wanting to deter future harms by teaching the transgressor a lesson. In practice, punishment typically satisfies both motives. If a colleague publicly ridicules a co-worker for stealing someone else's lunch from the breakroom, it will satisfy a retributive motive (because the thief will suffer by being publicly ridiculed) as well as a consequentialist motive (because the thief, as well as other observers, will learn that stealing food is wrong). Thus, it is typically very challenging to directly infer motives for punishment from observing punishment behaviour alone.

Importantly, experimental studies can test for the presence of both consequentialist and retributive motives by manipulating whether punishment is 'communicative' or not—that is, whether the punishment is delivered alongside an explicit communication that a norm has been violated[12]. Communicative punishment satisfies both retributive and consequentialist motives (Fig. 1): it satisfies retributive motives by inflicting damage on the transgressor, and it satisfies consequentialist motives by communicating that the damage is inflicted because they violated a norm, thereby potentially teaching the transgressor a lesson. Though most real-world punishment is communicative, occasionally it is not—consider the disgruntled employee who surreptitiously slashes their evil boss's tires one day after work, or the harried server who makes a rude customer's food unbearably spicy before bringing it out from the kitchen. Such 'non-communicative' punishment cannot satisfy consequentialist motives because the transgressor does not know why they suffer misfortune, and such knowledge is a prerequisite for learning a lesson. However, non-communicative punishment still satisfies retributive motives because the punisher knows that their actions cause the transgressor to suffer (Fig. 1). Thus, while retributive punishment merely involves inflicting harm, consequentialist punishment additionally involves communicating to the transgressor that they have violated a norm (otherwise, the transgressor would never have the opportunity to learn).

Research has shown that adults are willing to engage in both communicative and non-communicative punishment[13–19]. The observation of non-communicative punishment in adults provides evidence that at least some punishment is likely to be motivated by pure retribution. Interestingly though, adults are more willing to engage in communicative than non-communicative punishment[15,16]. Because both types of punishment inflict damage on the transgressor, the increased demand for communicative punishment is thought to result from its additional ability to satisfy consequentialist motives. Consistent with this, the strength of self-reported consequentialist motives—for example, a desire for the transgressor learn a lesson—is positively correlated with the increased engagement in communicative relative to non-communicative punishment[15].

Importantly, little to no work has investigated punishment motives from a developmental perspective. We know from previous research that toddlers are willing to punish antisocial individuals in third-party contexts[20] and that children around the age of four will tattle on their antisocial peers[21–23]. Additional research has demonstrated that children, even across cultures[24], will go as far to sacrifice their own personal resources, such as stickers, candies, or time playing on a slide, to punish a transgressor who had acted unfairly or unkindly[25–28]. We also know that children, when they punish, tend to care about restoring justice to victims[29,30]. For example, children will remove a resource from a thief and return it to the victim rather than keep it for themselves. But it remains unclear why young children are willing to punish at all in the first place: do

[1]Department of Psychology, Yale University, New Haven, CT, USA. [2]Social and Behavioral Science Initiative, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: julia.marshall@yale.edu; molly.crockett@yale.edu

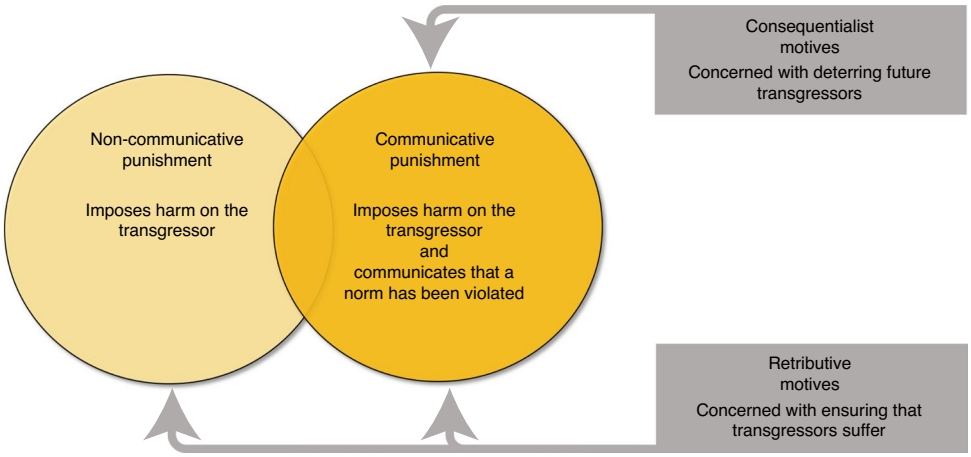

Fig. 1 | Communicative versus non-communicative punishment. A visual explanation of how consequentialist and retributive motives map onto communicative and non-communicative punishment.

children, like adults[15,18], punish for both consequentialist and retributive reasons?

The present work directly addresses this question by testing young children's propensities for both communicative and non-communicative punishment. One possibility is that young children are 'naive retributivists', punishing indiscriminately regardless of whether punishment can communicate that a norm has been violated. Such a finding would suggest that consequentialist motives develop in later childhood and adolescence, perhaps through observing and experiencing lesson-learning via punishment. Alternatively, it is possible that children are 'naive consequentialists', punishing exclusively when it communicates that a norm has been violated. This observation would imply that retributive motives develop later, perhaps through repeatedly experiencing schadenfreude when observing a transgressor suffer and associating those positive emotions with punishment behaviour. Finally, it is possible that young children are 'naive pluralists', with both consequentialist and retributive motives present from a young age, resulting in a pattern of punishment behaviour similar to that observed in adults[15]. Such a finding would suggest that both consequentialist and retributive motives develop early in life.

To distinguish between these possibilities, in an initial study, we assigned children ranging from four to seven years old ($N=113$) to one of three experimental conditions in a costly third-party punishment paradigm (Fig. 2). In two conditions, the participants watched a video depicting an antisocial child who ripped up another child's artwork; in the third condition, the participants watched a video depicting a neutral child who simply looked at another child's artwork. Manipulation checks verified that the participants considered the antisocial other meaner than the neutral actor and reported that the antisocial other elicited more negative and less positive emotions compared with the neutral actor (see the Supplementary Information for the details). After watching the video, the participants were given the opportunity to punish the antisocial (or neutral) child by giving up a valued resource: time playing on an iPad. Manipulation checks here too verified that the participants enjoyed playing on the iPad and wanted to play more, confirming that the punishment in this paradigm was costly for the participants (Supplementary Information).

Children were then given the option to place the iPad in a locked box. Doing so would prevent the other child from playing on the iPad, but it also meant that the participants would lose access to the iPad themselves (thereby ensuring that selecting the locked box was costly to the participants). If the participants decided to put the iPad in the locked box, they were further asked how long the other child should be restricted from playing. See Open Science Framework (OSF) for the full script: https://osf.io/ht7j6/. All conditions involved harming the antisocial (or neutral) other by preventing them from playing on the iPad; this was required to ensure that both conditions involved punitive behaviour[3]. Importantly though, the consequences of punishment were described differently across conditions to isolate consequentialist motives from retributive ones[15,18] (Fig. 2). In the non-communicative condition, the participants were told that, if they decided to lock up the iPad, the antisocial child would not be told why they could not play on the iPad, and thus they could not learn a lesson. In the communicative condition, the participants were instructed that, if they decided to punish, the antisocial child would be told why they could not play with the iPad and consequently would learn a lesson.

To rule out the possibility that punishment in the non-communicative condition was merely the result of a preference for locking up iPads—in which people punish others merely to inflict damage on them rather than in response to a transgression—we included a baseline control condition in which the character merely held, rather than tore up, the drawing. In this condition, the participants were told that, if they punished, the neutral child would not know why they could not play on the iPad. In all conditions, the participants were told that, if punished, the antisocial (or neutral) other would feel sad because they would not be allowed to play on the iPad.

We verified that the participants understood that the punished child would know why they were punished in the communicative condition, and that they would not know why they were punished in the non-communicative and baseline control conditions. We did so because we wanted to ensure that children believed that their decision about whether to punish carried real social consequences. Specifically, we confirmed that children believed that the transgressor would know why they could not play with the iPad if they were punished, whether the transgressor would be sad if they could not play with the iPad, and whether the transgressor would learn a lesson about ripping up drawings if they could not play with the iPad (see Methods for the exact language for the comprehension checks and Supplementary Information for the full details). We also confirmed that the participants believed punishment would make the punished child equally sad across conditions (Supplementary Information). Those who did not exhibit an understanding of the contingencies of the boxes were excluded from the main analyses ($N=22$). Furthermore, the participants made their punishment

Fig. 2 | Visualizations of the communicative, non-communicative, and baseline control conditions. The child actor was not always female. Across participants, we used either a female (as depicted here) or a male child actor (not depicted here; counterbalanced across participants). See OSF for the verbatim scripts per condition. Permission to use the participant images was obtained by the first author.

decisions in private, and the experimenter's identity changed after the participants made their decisions to minimize task demands and reduce the possibility that participants punish because of reputational concerns[31]. We preregistered our methods at aspredicted.org (communicative and non-communicative conditions, https://aspredicted.org/mz2pc.pdf; baseline control condition, https://aspredicted.org/ct965.pdf).

Results

In our first study, we found an effect of condition on the participants' punishment decisions ($\chi^2(2, N=113)=26.71$, $P<0.001$, $R^2=0.269$). The effect of condition remained when including participants who were excluded for failing the comprehension checks ($P=0.007$). The participants in the communicative condition were more likely to punish (mean (M) = 0.78, s.e.m. = 0.07) than the participants in the non-communicative condition ($M=0.39$, s.e.m. = 0.08, $\chi^2(1, N=75)=10.91$, $P=0.001$, odds ratio (OR) = 5.56, 95% confidence interval (CI) (2.01, 15.38)), as illustrated in Fig. 3. Furthermore, the participants in the communicative condition were more likely to punish than the participants in the baseline control condition ($M=0.13$, s.e.m. = 0.06, $\chi^2(1, N=75)=25.86$, $P<0.001$, OR = 23.93, 95% CI (7.04, 81.34)). Finally, the participants in the non-communicative condition were more likely to punish than the participants in the baseline control condition ($\chi^2(1, N=76)=6.26$, $P=0.012$, OR = 4.30, 95% CI (1.37, 13.51)). Condition did not interact with children's continuous age ($\chi^2(2, N=113)=1.70$, $P=0.427$); see the Supplementary Information for the full information. We found an identical pattern of results when analysing children's judgements regarding how long the antisocial other should be restricted from playing on the iPad (Supplementary Information). These findings provide initial evidence that young children are naive pluralists: they punish both because they want to inflict emotional damage on the transgressor and because they care about the transgressor learning a lesson and thereby potentially reforming their behaviour.

Beyond assessing the presence of each motive in young children, the additive nature of the experimental design also allowed us to examine the comparative strength of consequentialist motives compared with purely retributive motives. That is, one can infer the strength of solely consequentialist motives by subtracting the percentage of participants who punished in the non-communicative condition from the percentage of participants who punished in the communicative one. Doing so reveals that the participants punished 39% more (OR = 5.56, 95% CI (2.01, 15.38)) in the communicative condition than in the non-communicative condition. Furthermore, one can calculate the strength of solely retributive motives by subtracting the percentage of participants who punished in the baseline control from the percentage of participants who punished in the non-communicative condition. Doing so reveals that the participants punished 26% more (OR = 4.30, 95% CI (1.37, 13.51)) in the non-communicative condition than in the baseline control condition. Because the effect size CIs of the consequentialist effect and the retributive effect overlap, the data suggest that the strength of

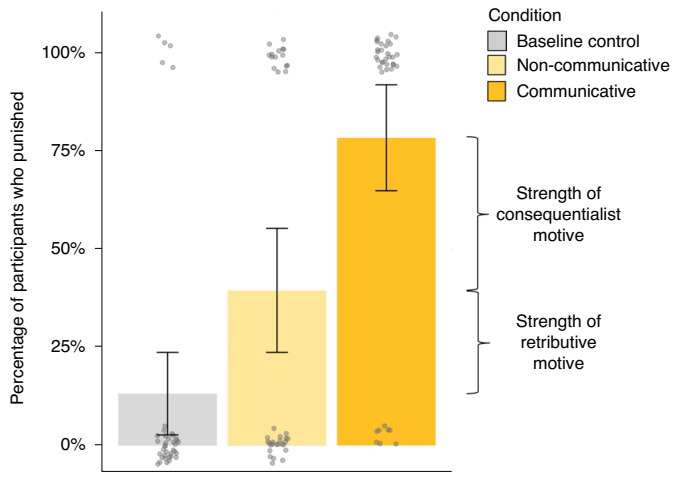

Fig. 3 | Percentages of participants who punished depending on condition in Study 1 ($N = 113$). The error bars represent bootstrapped 95% CIs. The brackets indicate estimates of the strengths of retributive and consequentialist motives in children, as reflected in the difference in punishment rates between the non-communicative and baseline control conditions (for the former) and the non-communicative and communicative conditions (for the latter).

children's retributive motives were not significantly different from the strength of children's consequentialist motives.

The results of Study 1 provide initial evidence that children predominantly between the ages of five and seven are naive pluralists—they, like adults, punish for both consequentialist and retributive reasons. Beyond documenting the presence of both motives, we also find that children's retributive and consequentialist desires do not significantly differ in strength. Two questions remained, which we tested in a second experiment ($N = 138$; the methods were preregistered at https://aspredicted.org/9su64.pdf). First, did children punish in the non-communicative condition because they falsely believed that the other child would reform their behaviour following punishment? In Study 2, we tested this by asking the participants to predict whether the antisocial other would re-offend in the future. Second, did the participants punish considerably more in the communicative condition because we told them that the punished child would learn a lesson? In reality, it is never guaranteed that punishment targets will learn; thus, instructing the participants that the punished other 'would learn a lesson' may have artificially inflated punishment in the communicative condition. In Study 2, we therefore instructed the participants that the punished child 'could' (instead of 'would') learn a lesson.

We made two minor changes in addition to those two main changes. First, we added several exploratory measures, including children's general attitudes about punishment, how the participants felt when they made their decision to punish or not, and how deserving of punishment they considered the antisocial (or neutral) other. See the Supplementary Information for the full details. Second, Study 2 focused on five- to seven-year-olds largely because four-year-olds disproportionately struggled with the methodology in Study 1 (~50% of comprehension check failures in Study 1 were four-year-olds; see the Supplementary Information for the details).

Study 2 fully replicated our findings from Study 1 (Fig. 4). We found an effect of condition on children's punishment decisions ($\chi^2(2, N = 138) = 20.20$, $P < 0.001$, $R^2 = 0.193$). The effect of condition remained when including participants who were excluded for failing the comprehension checks ($P < 0.001$). The participants in the communicative condition were more likely to punish ($M = 0.57$, s.e.m. $= 0.07$) than the participants in the non-communicative condition ($M = 0.33$, s.e.m. $= 0.07$, $\chi^2(1, N = 92) = 5.21$, $P = 0.022$, OR $= 2.69$, 95% CI (1.15, 6.28)). Furthermore, the participants in the communicative condition were more likely to punish than the participants in the baseline control condition ($M = 0.07$, s.e.m. $= 0.04$, $\chi^2(1, N = 92) = 19.22$, $P < 0.001$, OR $= 18.63$, 95% CI (5.04, 68.89)). Finally, the participants in the non-communicative condition were more likely to punish than the participants in the baseline control condition ($\chi^2(1, N = 92) = 8.23$, $P = 0.004$, OR $= 6.94$, 95% CI (1.85, 26.04)), again indicating that non-communicative punishment cannot be explained by a preference for locking up boxes. Thus, across two studies, we find consistent evidence for the presence of both consequentialist and retributive motives, thereby suggesting that children, like adults, are naive pluralists.

As in Study 1, we also examined the comparative strength of different motives. In doing so, we found that the participants punished 24% more (OR $= 2.69$, 95% CI (1.15, 6.28)) in the communicative condition than in the non-communicative condition. Furthermore, the participants punished 26% more (OR $= 6.94$, 95% CI (1.85, 26.04)) in the non-communicative condition than in the baseline control condition. As in Study 1, because the CIs for both the consequentialist and the retributive effect overlap, it suggests that children's retributive desires were not significantly stronger than their consequentialist ones or vice versa.

Unlike in Study 1, condition did interact with children's continuous age ($\chi^2(2, N = 138) = 7.85$, $P = 0.020$). Further investigation revealed that this interaction was largely driven by the fact that only three children punished in the baseline control condition, and all of these children were five years old. If we analyse only the communicative and non-communicative conditions, we do not find an age × condition effect ($\chi^2(1, N = 92) = 2.08$, $P = 0.150$). Because of this and because we did not find an age interaction in Study 1, we did not consider this age interaction any further, although see the Supplementary Information for more details about age effects in Studies 1 and 2.

We next considered whether non-communicative punishment could be explained by children in this condition erroneously inferring that they could teach transgressors a lesson by punishing them (despite us explicitly informing them to the contrary). If this was the case, then the participants who punished in both the communicative and non-communicative conditions should believe that the target of punishment would be less likely to re-offend. If not, only the participants in the communicative condition should believe that punishment would reduce re-offending. A logistic regression revealed an interaction between punishment decision (yes, no) and condition (communicative, non-communicative) when predicting beliefs about re-offending ($\chi^2(1, N = 92) = 3.94$, $P = 0.047$, $R^2 = 0.092$, Fig. 5). In the communicative condition, participants who punished were less likely to think that the transgressor would re-offend than participants who did not punish ($\chi^2(1, N = 46) = 7.33$, $P = 0.007$, OR $= 20.46$, 95% CI (2.30, 181.73)). However, in the non-communicative condition, there was no effect of punishment decision on beliefs about re-offending ($\chi^2(1, N = 46) = 0.36$, $P = 0.551$, OR $= 1.51$, 95% CI (0.39, 5.90)). This interaction was not moderated by participant age ($\chi^2(1, N = 92) = 0.76$, $P = 0.383$). These findings demonstrate that participants who punished in the non-communicative condition did not do so because they erroneously believed that the transgressor would reform their behaviour.

Finally, in exploratory analyses, we examined children's self-reported motives about punishment in general. Specifically, we asked children four questions—two of which measured consequentialist motives (for example, "Do you think people who do bad things should be punished because they should change their behavior?") and two of which measured retributive motives (for example, "Do you think people who do bad things should be punished because they deserve to feel sad?"). Paralleling findings with

adults[15], we found that children differentially endorsed each of these items ($F(3, 135) = 82.97$, $P < 0.001$, partial eta squared (η_p^2) = 0.648).

Specifically, children endorsed the two consequentialist items the most—they indicated that antisocial others should be punished in general because they should change their behaviour ($M = 4.37$, s.d. = 1.74) and because antisocial others need to learn a lesson ($M = 4.16$, s.d. = 1.86). The degree of endorsement did not differ between these two items ($t(137) = 1.24$, $P = 0.219$, Cohen's $d = 0.12$, 95% CI (−0.07, 0.30)). Children also endorsed retributive motives, but less so. Specifically, children endorsed the idea that antisocial others should be punished because it is the right thing to do ($M = 2.66$, s.d. = 1.92), and the degree of endorsement for this item differed from those for the item about changing behaviour ($t(137) = 9.00$, $P < 0.001$, $d = 0.93$, 95% CI (0.70, 1.17)) and for the item about lesson-learning ($t(137) = 8.91$, $P < 0.001$, $d = 0.79$, 95% CI (0.60, 0.99)). Children also endorsed the idea that antisocial others should be punished because antisocial others deserve to feel sad ($M = 2.01$, s.d. = 1.76), although children endorsed this item less than the item about changing behaviour ($t(137) = 13.90$, $P < 0.001$, $d = 1.35$, 95% CI (1.10, 1.60)), the item about lesson-learning ($t(137) = 11.73$, $P < 0.001$, $d = 1.19$, 95% CI (0.94, 1.43)) and the item about whether punishment is the right thing to do ($t(137) = 3.49$, $P = 0.001$, $d = 0.35$, 95% CI (0.15, 0.56)). We do find that, for all motives, children's responses significantly differed from 'no' (all P values < 0.001). These findings mirror children's actual behaviour: children in general endorsed both retributive and consequentialist motives, but, similar to their punitive behaviour, they endorsed consequentialist motives more than retributive ones (see the Supplementary Information for additional information).

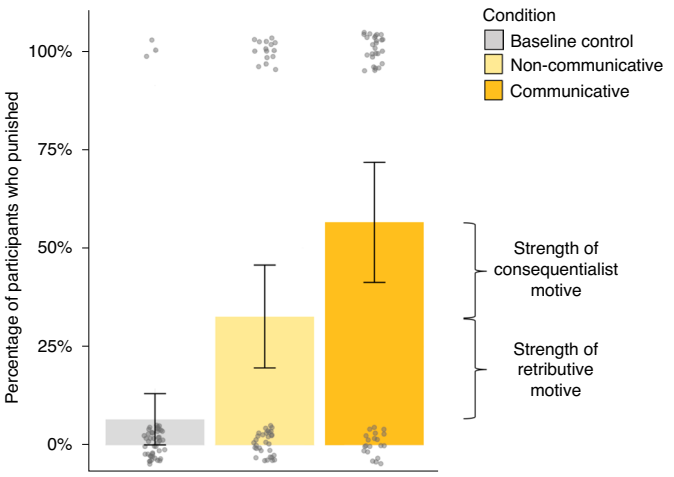

Fig. 4 | Percentages of participants who punished depending on condition in Study 2 ($N = 138$). As in Fig. 3, but for Study 2. The error bars represent bootstrapped 95% CIs. The brackets indicate estimates of the strengths of retributive and consequentialist motives in children, as reflected in the difference in punishment rates between the non-communicative and baseline control conditions (for the former), and the non-communicative and communicative conditions (for the latter).

Discussion

Overall, these two preregistered studies provide clear evidence for the presence of both consequentialist and retributive motives in young children, supporting the naive pluralism hypothesis. Our observations cohere with past research showing that children between the ages of five and seven[25–28] are willing to engage in costly third-party punishment, and reveal the motives behind children's punitive behaviour. Children reliably engaged in purely retributive punishment: they punished solely to make an antisocial other sad without any possibility of deterring future antisocial behaviour. Children did not punish in the non-communicative condition out of a preference for locking iPads in boxes, shown by the fact that children punished less in the baseline control condition. Furthermore, non-communicative punishment could not be explained by erroneous beliefs that punishing would teach the transgressor a lesson. This demonstrates that young children are not pure consequentialists. Rather, our data suggest that young children engaged in costly third-party punishment for purely retributive reasons.

Yet our data also demonstrate that young children are not solely retributivists—they punished more when doing so conferred a social benefit than when it did not, and concurrently believed that the target of punishment would be less likely to re-offend in the future. This aligns with research showing that adults punish more when transgressors will know that they are being punished[15]. One could argue that communicative punishment could be driven partly by retributive motives above and beyond the non-communicative condition, if the participants believe that communicating to the antisocial child that they did something wrong would inflict added emotional damage. Nevertheless, we find it unlikely that the increase in punishment between the non-communicative and communicative conditions can be entirely explained by the retributive component of the communicative condition because our participants did not predict that the antisocial child, if punished, would be sadder in the communicative than the non-communicative condition (Supplementary Information). Instead, we think it is plausible that children punish more in the communicative condition because lesson-learning is often considered positively and valued in adulthood[32] and also in childhood, as evidenced by children's self-reported punitive motives in Study 2. Taken together, then, these findings support the naive pluralist hypothesis.

Our data not only speak to the presence of consequentialist and retributive motives in young children but also speak to the comparative strength of these motives. Specifically, children's consequentialist motives (that is, punishment rates in the communicative condition relative to the non-communicative condition) did not significantly differ from children's retributive motives in either study (that is, punishment rates in the non-communicative condition relative to the baseline control condition). The present findings therefore align with work suggesting that children punish out of prosocial concerns about restoring justice to victims insofar as children in our studies punish because they are concerned with promoting cooperation[29,30].

In general, children's punitive motives align with those of adults—both children and adults are concerned with consequentialist and retributive punishment. As a result, these findings suggest that ample social experience with punishment may be minimally required for the emergence of both motives in young children. Instead, children around the age of five seem inclined to weigh both consequentialist and retributive concerns when deciding whether to punish. This raises questions about how certain motives can subsequently be promoted, mitigated, or maintained through social learning and cultural values. For instance, some philosophers have argued that punishment should be less rooted in retributive interests and more grounded in consequentialist outcomes[33–39]. Because children seem to value both concerns, it seems plausible that they are capable of reasoning about punishment in consequentialist terms, and therefore, a foundation exists to promote such reasoning throughout early childhood and into adulthood.

We should note, however, that we only tested children from predominantly middle- to upper-class backgrounds and did not test enough children within the different age categories. Because of this, our generalizability is limited, and we have limited power to detect age effects. In light of this limitation, we think future research should investigate the degree to which children in different cultures (for example, non-WEIRD cultures) and at different ages may be

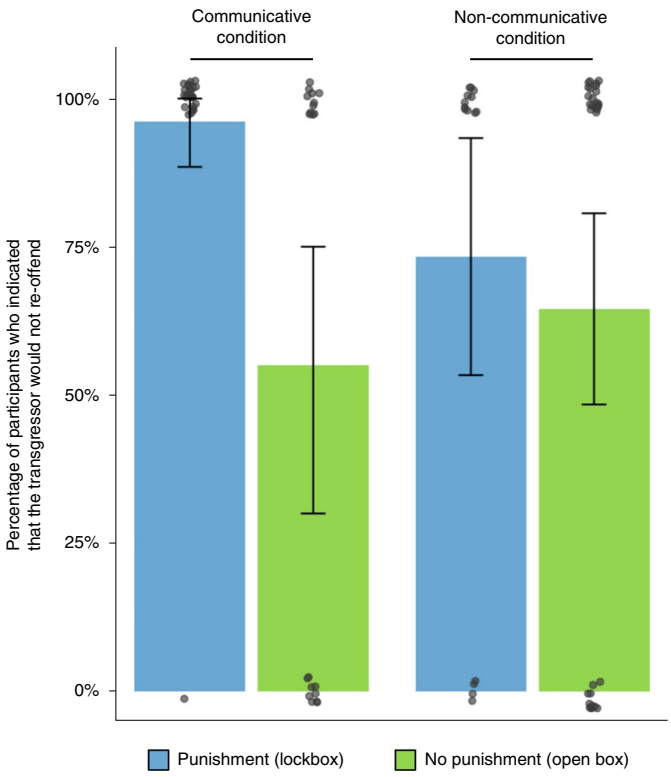

Fig. 5 | Percentages of participants who predicted that the transgressor will not re-offend. Percentages of participants who predicted that the transgressor will not re-offend as a function of condition (either communicative or non-communicative) and of their punishment decision (either punish or not). The error bars represent bootstrapped 95% CIs.

inclined to punish for consequentialist versus retributive purposes, and whether childhood motives for punishment in a given culture are related to patterns of norm enforcement observed among adults in that culture.

There are several additional limitations worth noting. First, our methodology, for practical purposes, involved witnessing a transgression on a video rather than in real life. This feature of the design may have affected whether participants truly believed that they were interacting with a real person and that their actions elicited real outcomes. Our comprehension checks suggest that children believed their decisions would result in real consequences, but we cannot know for sure. Furthermore, our methodology was highly verbal, and because of this, younger children may have struggled more with following the different components of the study.

An additional consideration pertains to whether adults prefer punishment over alternative means of rectifying injustice[40,41]. It is possible that punishment may be attenuated in a context whereby there are alternative ways to intervene, such as compensating the victim or mere verbal intervention. Additional work is needed to test whether children prefer punishment—imposing a cost on the transgressor (such as removing the iPad or removing the stolen good from the thief)—over compensatory actions that only help the victim and do not impose a direct cost on the transgressor (such as giving the victim extra candy). We think future research can answer this question fully, and there is some emerging evidence that children prefer punishment to compensation, as evidenced by McAuliffe and Dunham (unpublished manuscript) and others[42].

Other factors may moderate the results found here. For instance, the punishment in our studies is moderately harmful. We speculate that, as the punishment becomes disproportionate to the crime, children may be less inclined to punish retributively, although future research is required to assess this. Additionally, punishment could be made more or less costly by varying how much the participant has to give up to punish the transgressor. It seems plausible that, as punishment becomes cheaper, children may be more inclined to punish in general, thereby minimizing the difference between retributive and consequentialist punishment. Finally, our study involves verbal scaffolding to ensure that the participants understand the consequences of some punishments compared with others; it remains unclear whether children would respond similarly in situations that involve more spontaneity.

In conclusion, two preregistered studies provide evidence that, from a young age, children willingly engage in costly third-party punishment even when doing so exclusively satisfies retributive concerns. At the same time, young children are more willing to punish when doing so teaches the transgressor a lesson. These findings provide evidence that young children are naive pluralists in their punishment behaviour, calibrated to both retributive and consequentialist concerns.

Methods

Study 1. *Participants and design.* These studies were approved by the Yale Human Subjects Committee (Protocol No. 1302011578), and we obtained informed parental consent and verbal assent from all participants.

The sample included 113 participants: 37 participants in the communicative condition (M_{age} = 6.59, s.d.$_{age}$ = 1.05; 22 females), 38 in the non-communicative condition (M_{age} = 6.29, s.d.$_{age}$ = 1.13; 17 females) and 38 in the baseline control condition (M_{age} = 6.15, s.d.$_{age}$ = 1.03; 16 females). The participants were recruited for the communicative and non-communicative conditions first, followed by the baseline control condition. This is because the necessity of the baseline control condition was contingent on the results from the other two experimental conditions. This condition was necessary only if children did engage in punishment in the non-communicative condition to verify that punishment was not a result of a preference for locking up iPads. We thus ran the communicative and non-communicative conditions first. For more information about this and for information about the preregistration and power analyses, see the Supplementary Information.

The sample comprised 16 four-year-olds, 25 five-year-olds, 31 six-year-olds and 41 seven-year-olds (M = 6.34, s.d. = 1.08). The sample was predominantly white (N = 67). Eight additional children were mixed race, 7 were Hispanic, 3 were Black, 3 were Asian, and 2 were South Asian. The parents of the remaining children opted to not report their child's ethnicity. Fifty-eight of the participants were male (55 females). All participants were tested in the Mind and Development Lab at Yale University. The participants were covertly videorecorded for the purposes of data coding. Several additional children (N = 24) were tested but excluded in accordance with our preregistered exclusion criteria (described in 'Materials and procedure'). See the Supplementary Information for the full information.

Materials and procedure. All verbatim scripts are available on OSF (https://osf.io/ht7j6/).

Introduction of the iPad and costly self-report validation. We first introduced children to an iPad, let them play an iPad game, and measured the degree to which they liked playing on the iPad and how much they wanted to continue playing. Specifically, the children were shown either a game called Scoops, where the aim of the game is to collect ice cream scoops and avoid onions and tomatoes, or Happy Fall, where the aim of the game is to help a candy collect coins while going down a tunnel. We chose these games because children would probably not have any previous familiarity with either game (no children expressed having played the game previously), and neither game involved violence. See the Supplementary Information for the full information.

Antisocial (or neutral) child introduction. The experimenter then told the child a story about another child, named either Jessica or John. We counterbalanced across participants such that approximately half of the female and half of the male participants witnessed John as the transgressor (for clarity's sake, the script described here and in the remainder of the paper will involve Jessica). We examine stimulus effects in the Supplementary Information. In Study 1, we find that the participants punished the female transgressor more than the male transgressor regardless of their own gender, but these stimulus effects did not replicate in Study 2.

In the story, it was described that Betty and Jessica were drawing pictures, and Betty went to go to the bathroom. In the communicative and non-communicative conditions, the experimenter then showed the participant child a short video clip in which Jessica ripped up Betty's artwork. In the baseline control condition, the

experimenter showed the participant child a short video clip in which Jessica just looked at Betty's artwork. For both Study 1 and Study 2, the exact stimulus videos are not included on OSF because of privacy concerns. Not all parents consented to having their children's video posted on the internet. The videos are available on request from the first author. Importantly, though, all children in our stimuli were the same age (six years old) and were provided the same instructions when filming the stimulus videos.

Meanness validation item and emotion questions. The experimenter then asked the meanness validation item and three emotion items (happiness, sadness and anger) in a randomized order. See the Supplementary Information for the wording and findings.

Locked box and open box introduction. The experimenter then introduced the participant to two black boxes (Fig. 2; whether the box was placed on the left or right was counterbalanced). One box had a functioning lock, and the other box did not. The experimenter explained that they could choose to place the iPad in the locked box (which would prevent Jessica from playing on it) or the open box (which would allow Jessica to play on it). The experimenter also explained that, if the locked box was selected, the participant could no longer play on the iPad, and that, if the open box was selected, the participant could play on the iPad. We did not specify the amount of time the participant would give up on the iPad if they chose to do so. But given the nature of the experimental study sessions in the lab, the participant knew that they would be leaving the lab shortly after having finished the study, so children in reality were giving up approximately 10–15 minutes of time on the iPad (depending on how long the family wanted to stay in the lab). The non-punitive children were generally told that the study was over and that they should select a prize for participating (standard protocol in the lab); a few children stuck around to play on other toys that were present in the waiting room.

Next, the experimenter asked two comprehension checks. For each box, the experimenter asked, "Can you tell me: if you decide to put the iPad in this box, will you get to play with the iPad anymore? Yes or no?" If the participant answered incorrectly, the experimenter corrected them and re-asked the question. If the participant continued to respond incorrectly, they were excluded from analyses for comprehension check failure.

Isolating retributive motives. First, the experimenter introduced the participant to a sign associated with the open box (the sign read "Yes iPad" with a large green check mark). Next, the experimenter introduced the participant to a sign associated with the locked box. The locked box sign read either "No iPad because you ripped up the drawing" (in the communicative condition) or just "No iPad" (in the non-communicative and baseline control conditions). These signs depicted an iPad with a large red X over the iPad. The visualizations on the cards (that is, the green check mark and the red X) aided participants who may have difficulty reading the signs.

Because some of the younger participants probably could not read the language on the signs, verbal instructions were given to the participants, and these descriptions varied across conditions. In the communicative condition, the participants were told that, if they placed the iPad in the locked box, Jessica would know why she couldn't play with the iPad and would thereby learn a lesson. In the non-communicative condition, the participants were told that, if they placed the iPad in the lockbox, Jessica would not know why she couldn't play and thereby would not learn a lesson. In the baseline control condition, the participants were told that, if they placed the iPad in the lockbox, Jessica would not know why she couldn't play with the iPad; we did not mention anything about learning a lesson because there is no lesson to be learned when acting neutrally. Critically, the participants were told in all conditions that, if they placed the iPad in the lockbox, Jessica would feel sad.

The experimenter then asked the participant several comprehension checks in randomized order: (1) for knowledge comprehension, "If you do decide to put the iPad in this box, will Jessica know why she can't play with the iPad? Or will she not know why?"; and (2) for sadness comprehension, "If you do decide to put the iPad in this box, will Jessica feel sad that she can't play?" (After the sadness comprehension check, we also asked the participants how sad they thought the antisocial child would be. We asked, "How sad do you think Jessica will be? A teeny bit sad, a little bit sad, or very sad?" This resulted in a three-point Likert-style scale of projected sadness.) For the communicative and non-communicative conditions, the participants were also asked (3) for lesson-learning comprehension, "Can you tell me? If you do decide to put the iPad in this box, will Jessica learn a lesson about not ripping up people's drawings?" This third question was not included in the baseline control condition because it was not relevant.

For each question, if the participant responded incorrectly, the experimenter corrected them and asked the question a second time. If the participant continued to answer incorrectly, they were excluded.

Punishment decision. We had the participants make their punishment decisions in private, and we instructed the participants before making their decisions that a different experimenter would come back into the room after the participant had made their decision. We viewed these two methodological components as very important given research indicating that reputational concerns play a role in shaping participants' punitive decisions[31]. More specifically, the participants were instructed that the first experimenter would leave the room, and once they had decided which box to put the iPad in, they should ring a bell, and a different experimenter would return. Once the child rang the bell, this second experimenter returned to the room, introduced themselves to the children and recorded the two dependent punishment measures. For the binary punishment measure, the experimenter indicated which box the child had selected (the locked or unlocked box). We also asked a continuous punishment measure (Supplementary Information). The experimenter also asked the participant why they made their punitive selection: "Why did you choose that box?" Finally, the experimenter allowed the child to play with the iPad regardless of their punishment choice.

Study 2. *Participants and design.* Study 2 involved the same three conditions as in Study 1: communicative condition (antisocial child), non-communicative condition (antisocial child) and baseline control condition (neutral child). The assignment of condition was randomized across all participants. The sample resulted in a total of 138 children: 46 participants in the communicative condition (M_{age} = 6.61, s.d.$_{age}$ = 0.84; 23 females), 46 in the non-communicative condition (M_{age} = 6.54, s.d.$_{age}$ = 0.86; 25 females), and 46 in the baseline control condition (M_{age} = 6.63, s.d.$_{age}$ = 0.88; 17 females). Notably, we did not test four-year-olds, because they had difficulties understanding the experiment, as evidenced by them representing the majority age group failing the comprehension checks in Study 1.

We aimed to equally recruit participants within each age group within each condition. However, the sample ultimately comprised mostly seven-year-olds (N = 57), primarily because more children of this age signed up at the schools where we tested. Forty-two six-year-olds and 39 five-year-olds also participated. Forty-three participants were tested at schools and summer camps in Connecticut, 41 participants were tested in the lab at Yale University, 35 participants were tested at a private school in Atlanta, and 19 participants were tested at a private school in Manhattan. Because we did not exclusively test children in the lab, as we did in Study 1, we were not able to collect parental demographic data. Several additional children (N = 17) were tested but excluded for various reasons—see the Supplementary Information for the full information.

Materials and procedure. Introduction of the iPad and costly self-report validation. The introduction of the iPad and the costly self-report validation items were the same as in Study 1. As in Study 1, we asked two questions about whether the participant liked playing on the iPad and whether they wanted to play more. See the Supplementary Information for the full information.

Antisocial (or neutral) child introduction. The introduction of the antisocial (or neutral) child was largely similar to Study 1. We made two main changes. First, rather than including only one Jessica character and one John character, we had two Jessica characters and two John characters to ensure that the effects in Study 1 were not a result of the specific stimuli we had used. Which character the participants learned about was randomized across participants.

Second, we altered one slight ambiguity in Study 1's script. Specifically, in Study 1, we said, "When Betty had to go to the bathroom, I'm going to show you a video of what Jessica did to Betty's artwork." In Study 2, we said, "When Betty was done drawing, she had to go to the bathroom, so she asked Jessica to look at her drawing while she went. I'm going to show you a video of what happened next." We made this change because we thought it may have been confusing in Study 1 to say that Jessica did something to Betty's drawing when she simply looked at it in the baseline control condition.

Locked box and open box introduction. The locked and open box introduction was largely similar to Study 1. However, rather than stating that Jessica would learn a lesson in the communicative condition, we changed the language to: "Jessica could learn a lesson." For the projected sadness question, we used a six-point scale whereby, after children indicated that Jessica would be sad if she could not play on the iPad, we asked, "How sad will she be? A tiny bit sad, very, very sad, or somewhere in between?" The question was accompanied by a laminated card depicting five sad faces increasing in size. The design of these questions then rendered a six-point Likert-style question in which 1 signified responding 'no' and 6 signified 'very, very sad'.

Punishment decision. The punishment decision was also as in Study 1, but we did not include the continuous punishment measure because the results did not differ for the continuous and binary measures in Study 1.

Follow-up questions. We asked several follow-up questions after the participants made their punishment decisions. We provide the methods here only for the ones presented in the main text; see the Supplementary Information for the additional questions.

Re-offending question. We asked the participants to recall what happened in the video they had seen earlier to verify the children had not forgotten.

The experimenter then asked, "So, if Jessica comes back tomorrow, do you think Jessica will do that again? Yes or no?" We also asked, "Why do you think that?"

General punitive attitudes. Inspired by previous work[15], we developed four questions, two of which measured consequentialist beliefs and two of which measured retributive ones: (1) "Do you think people who do bad things should be punished because they deserve to feel sad?" (retributive), (2) "Do you think people who do bad things should be punished because it's the right thing to do?" (retributive), (3) "Do you think people who do bad things should be punished because they need to learn a lesson?" (consequentialist) and (4) "Do you think people who do bad things should be punished because they should change their behavior?" (consequentialist). We asked these questions in a randomized order. For each of these questions, we asked 'yes' or 'no'. If the participants said 'yes', we asked, "How much do you think that? A teeny bit, a lot, or somewhere in between?" These questions were accompanied by a laminated sheet featuring five circles increasing in size. This structure rendered a six-point scale where 1 represented 'no' and 6 represented 'a lot'.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data related to these studies are publicly available on OSF at https://osf.io/ht7j6/.

Code availability

Most analyses were conducted in SPSS and using freely available packages in the R environment for statistical computing. All syntax and code are available at https://osf.io/ht7j6/.

Received: 14 October 2019; Accepted: 21 September 2020;
Published online: 23 November 2020

References

1. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
2. Henrich, J. et al. Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484 (2010).
3. Vidmar, N. & Miller, D. T. Social psychological processes underlying attitudes toward legal punishment. *Law Soc. Rev.* **14**, 565–602 (1980).
4. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388 (1982).
5. Fehr, E. & Fischbacher, U. Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004).
6. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
7. Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535 (2003).
8. Balliet, D., Mulder, L. B. & Van Lange, P. A. Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594–615 (2011).
9. Mathew, S. & Boyd, R. Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl Acad. Sci. USA* **108**, 11375–11380 (2011).
10. Kant, I. in *Why Punish? How Much? A Reader on Punishment* (ed. Tonry, M. H.) 31–36 (Oxford Univ. Press, 2011).
11. Bentham, J. in *What Is Justice? Classic and Contemporary Readings* (eds Soloman, R. C. & Murphy, M. C.) 215–220 (Oxford Univ. Press, 2000).
12. Funk, F., McGeer, V. & Gollwitzer, M. Get the message: punishment is satisfying if the transgressor responds to its communicative intent. *Pers. Soc. Psychol. Bull.* **40**, 986–997 (2014).
13. Baron, J. in *Psychological Perspectives on Justice: Theory and Applications* (eds. Mellers, B. & Baron, J.) 109–137 (Cambridge Univ. Press, 1993).
14. Carlsmith, K. M., Darley, J. M. & Robinson, P. H. Why do we punish? Deterrence and just deserts as motives for punishment. *J. Pers. Soc. Psychol.* **83**, 284–299 (2002).
15. Crockett, M. J., Özdemir, Y. & Fehr, E. The value of vengeance and the demand for deterrence. *J. Exp. Psychol. Gen.* **143**, 2279–2286 (2014).
16. Goodwin, G. P. & Gromet, D. M. Punishment. *WIREs Cogn. Sci.* **5**, 561–572 (2014).
17. Keller, L. B., Oswald, M. E., Stucki, I. & Gollwitzer, M. A closer look at an eye for an eye: laypersons' punishment decisions are primarily driven by retributive motives. *Soc. Justice Res.* **23**, 99–116 (2010).
18. Nadelhoffer, T., Heshmati, S., Kaplan, D. & Nichols, S. Folk retributivism and the communication confound. *Econ. Philos.* **29**, 235–261 (2013).
19. Ouss, A. & Peysakhovich. When punishment doesn't pay. *J. Law Econ.* **58**, 625–655 (2015).
20. Hamlin, J. K., Wynn, K., Bloom, P. & Mahajan, N. How infants and toddlers react to antisocial others. *Proc. Natl Acad. Sci. USA* **108**, 19931–19936 (2011).
21. Vaish, A., Missana, M. & Tomasello, M. Three-year-old children intervene in third-party moral transgressions. *Br. J. Dev. Psychol.* **29**, 124–130 (2011).
22. Heyman, G. D., Loke, I. C. & Lee, K. Children spontaneously police adults' transgressions. *J. Exp. Child Psychol.* **150**, 155–164 (2016).
23. Yucel, M. & Vaish, A. Young children tattle to enforce moral norms. *Soc. Dev.* **27**, 924–936 (2018).
24. House, B. R. et al. Social norms and cultural diversity in the development of third-party punishment. *Proc. R. Soc. B* **287**, 20192794 (2020).
25. Jordan, J. J., McAuliffe, K. & Warneken, F. Development of in-group favoritism in children's third-party punishment of selfishness. *Proc. Natl Acad. Sci. USA* **111**, 12710–12715 (2014).
26. McAuliffe, K., Jordan, J. J. & Warneken, F. Costly third-party punishment in young children. *Cognition* **134**, 1–10 (2015).
27. Yang, F., Choi, Y. J., Misch, A., Yang, X. & Dunham, Y. In defense of the commons: young children negatively evaluate and sanction free riders. *Psychol. Sci.* **29**, 1598–1611 (2018).
28. Yudkin, D. A., Van Bavel, J. J. & Rhodes, M. Young children police in-group members at personal cost. *J. Exp. Psychol. Gen.* **149**, 182–191 (2019).
29. Riedl, K., Jensen, K., Call, J. & Tomasello, M. Restorative justice in children. *Curr. Biol.* **25**, 1731–1735 (2015).
30. Kanakogi, Y. et al. Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat. Hum. Behav.* **1**, 0037 (2017).
31. Jordan, J. J., Hoffman, M., Bloom, P. & Rand, D. G. Third-party punishment as a costly signal of trustworthiness. *Nature* **530**, 473–476 (2016).
32. Twardawski, M., Tang, K. T. & Hilbig, B. E. Is it all about retribution? The flexibility of punishment goals. *Soc. Justice Res.* **33**, 195–218 (2020).
33. Bommarito, N. Virtuous and vicious anger. *J. Ethics Soc. Philos.* **11**, 1–27 (2017).
34. Flanagan, O. *The Geography of Morals: Varieties of Moral Possibility* (Oxford Univ. Press, 2016).
35. Leboeuf, C. in *The Moral Psychology of* Anger (eds Cherry, M. & Flanagan, O.) 15–30 (Rowman & Littlefield International, 2017).
36. Nussbaum, M. C. *Anger and Forgiveness: Resentment, Generosity, and Justice* (Oxford Univ. Press, 2016).
37. Silvermint, D. Rage and virtuous resistance. *J. Polit. Philos.* **25**, 461–486 (2017).
38. Tessman, L. *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (Oxford Univ. Press, 2005).
39. Srinivasan, A. The aptness of anger. *J. Polit. Philos.* **26**, 123–144 (2018).
40. FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J. & Phelps, E. A. Fairness violations elicit greater punishment on behalf of another than for oneself. *Nat. Commun.* **5**, 5306 (2014).
41. Heffner, J. & FeldmanHall, O. Why we don't always punish: preferences for non-punitive responses to moral violations. *Sci. Rep.* **9**, 13219 (2019).
42. Miller, D. T. & McCann, C. D. Children's reactions to the perpetrators and victims of injustices. *Child Dev.* **50**, 861–868 (1979).

Acknowledgements

We thank the Crockett Lab and the Mind and Development Lab for their valuable feedback in the design and methodology of the present studies. We also thank A. Buck, A. Gollwitzer, S. Hollander, C. Johnson, E. Mahaffey, S. Minnillo, A. Morra, I. Munday, A. Sacchi, C. Seita and C. Welsh for assistance with the data collection. Finally, we thank the generous and wonderful parents, children and schools who helped us with this project.

Author contributions

J.M., D.A.Y. and M.J.C. developed the study concept and study design. J.M. collected the data and performed the analyses. J.M., D.A.Y. and M.J.C. drafted the manuscript and provided critical revisions. All authors approved the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00975-9.

Correspondence and requests for materials should be addressed to J.M. or M.J.C.

Peer review information Primary handling editor: Charlotte Payne.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

nature research

Corresponding author(s): Julia Marshall & Molly Crockett

Last updated by author(s): May 13, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All available is on Open Science Framework: https://osf.io/ht7j6/?view_only=9daf7fe7b8ef40a5bb4fa77c07a0fcf4
Data analysis	All data analyses were conducted on either SPSS or R. All code for these analyses are on OSF.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data related to these studies are publicly available on Open Science Framework at https://osf.io/ht7j6/?view_only=9daf7fe7b8ef40a5bb4fa77c07a0fcf4.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The studies involved quantitative research methodologies.
Research sample	The sample included children ranging in age from 4 to 7 years of age (in Study 1) and children ranging in age from 5 to 7 years of age (in Study 2). We recruited a developmental sample because our research question pertained to the early emergence of retributive behavior. The demographics of the sample are included in the methods. In general, the children in Study 1 were from the New Haven area. The sample from Study 2 involved a more geographically diverse sample, as some participants were recruited in Atlanta, New York, New Haven, and other areas around Connecticut.
Sampling strategy	Both samples were convenience samples. The sample size determinations are explained in detail in our pre-registrations. For Study 1, we set a rule to either end data collection after two months or after we reached ~144 participants--the 144 number was determined from a power analysis in which we aimed to have 85% power to detect a medium effect size of condition. The sample size for Study 2 was based off of a power analysis in which we had 80% power to replicate the smallest effect documented in Study 1 (the difference between control and non-communicative punishment).
Data collection	The data was all recorded via a video camera and coded by research assistants who were blind to experimental hypotheses. It was impossible to fully blind the experimenters considering they had to read different scripts for different conditions. In Study 1, we were able to switch experimenters at the time of the decision to minimize any task demands or experimenter effects. We were not able to do so in Study 2 because we tested many of the participants in schools and were therefore subject to different testing conditions in those situations. Still we document retributive behavior in both studies. We also had several experimenters during the course of data collection, further mitigating any potential experimenter effects.
Timing	For Study 1, as described in the Method section, we tested children in the communicative and non-communicative conditions in the summer of 2018. Data collection for the control condition occurred in December 2018-February 2019 after the other conditions were completed. Importantly, in Study 2, all conditions were tested collectively (randomized across participants, between-subjects) from March 2019 to August 2019.
Data exclusions	In Study 1, 24 children were excluded, as described on pp. 24 in the main manuscript. These exclusions were as a result of comprehension check failures that were pre-registered. Importantly, as explained in the mansucript, the results do not meaningfully change if we include all participants. The same was true in Study 2--we had to exclude 17 children for reasons outlined in our preregistration. These are described in the Supplemental Materials.
Non-participation	No participants began the experiment and dropped out.
Randomization	Participants were randomly assigned to condition. Prior to beginning Study 1 and 2, we created a list from a random number generator to determine which condition participants would be subject to.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	For Study 1, participants included families who have previously signed up for the Yale Mind & Development lab database. In general, these are families who reside in the New Haven area. For Study 2, we designed the experiment such that it was more portable and were therefore able to test participants outside of the lab. Here we sent permission forms home to families attending schools for which we have a collaboration to see whether they would want their child to participate in the study. In general, we have high rates of consent form return and do not suspect any evidence of a problematic self-selection bias in recruitment. Nonetheless, we note the limitations of our sample in the main manuscript in the penultimate paragraph on pp. 13.
Ethics oversight	Yale Human Subjects Committee (Protocol number: 1302011578)

Note that full information on the approval of the study protocol must also be provided in the manuscript.