Children as assessors and agents of third-party punishment

Julia Marshall ✉ *and Katherine McAuliffe* ✉

Abstract | Responding to wrongdoing is a core feature of our social lives. Indeed, a central assumption of modern institutional justice systems is that transgressors should be punished. In this Review, we synthesize the developmental literature on third-party intervention to provide insight into the types of responses to transgressions that are privileged early in ontogeny. In particular, we focus on young children as both assessors and agents of third-party punishment. With respect to assessment, children have rich expectations about the pursuit of punishment and evaluate those who punish transgressors positively. With respect to agency, children punish wrongdoing even when doing so is costly, and their motives to do so are tethered to a variety of concerns (such as retribution and restoration). Our Review suggests that key concepts in modern institutional justice systems are apparent in early child development, and that third-party punishment is a signature of children's sophisticated toolkit for regulating social relationships and behaviour.

Consider a child who yells at a bully for shoving a classmate and a judge who sentences a murderer to prison. At first glance, these situations might seem quite different. For example, the former involves punishment by a child, whereas the latter involves punishment by an institutional authority with years of formal legal training. Yet despite these differences, both situations illustrate that people respond to wrongdoing by holding transgressors accountable.

Third parties can respond to transgressions in a variety of ways (FIG. 1). One particularly well studied form of intervention is third-party punishment, typically defined as the imposition of harm on a transgressor by a bystander[1–3]. This sort of punishment differs from second-party punishment wherein victims (rather than bystanders) impose harm on transgressors (BOX 1). In some cases, third-party punishment might involve the direct imposition of harm, such as when a police officer fines someone who drives over the speed limit or a parent sends their child to time-out. In experimental research studies, adults across diverse societies are willing to punish third-parties[1,4], and this behaviour is key to maintaining cooperation within societies[5–11]. In other cases, harm is imposed on transgressors indirectly by, for example, gossiping about a transgressor to a neighbour. Observational research has found that adults in Western societies often punish transgressions in indirect ways[12,13].

However, third-party intervention need not always involve direct or indirect punishment. Individuals can intervene by attempting to rectify wrongs in ways that do not necessarily involve punishment. For example, observers might recruit a third party (such as a teacher) to respond to a transgression, which often — but not always — results in punishment. Alternatively, when ordinary citizens are not in a position to punish, they might seek to hold transgressors accountable by openly acknowledging a misdeed through verbal protest, which can rise to the level of direct punishment when severe enough to cause emotional harm. Beyond these types of interventions, individuals witnessing misdeeds can engage in other forms of intervention that are less directly tied to punishment, including restorative justice (for example, compensating victims[14]), dissolving a relationship between oneself and the transgressor ('partner choice'[15]), or encouraging forgiveness[16].

Intervention by third parties, most typically in the form of punishment, represents a cornerstone of institutional justice systems. Yet, children generally encounter third-party intervention long before interacting with any formal justice systems. Indeed, children are exposed to third-party punishment (through 'time-outs' or the removal of toys) in their earliest years as witnesses to punishment or as targets of punishment themselves[17]. However, developmental research suggests that children are far more than passive witnesses to — or recipients of — third-party punishment. Rather, children can reason about contexts that call for intervention and punish others' wrongdoings themselves.

Understanding how and when children begin to think about and pursue third-party punishment in addition to other related forms of third-party intervention can shed light on the developmental, cognitive and

Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA, USA.

✉*e-mail: marshaau@bc.edu; mcaulikg@bc.edu*

https://doi.org/10.1038/s44159-022-00046-y

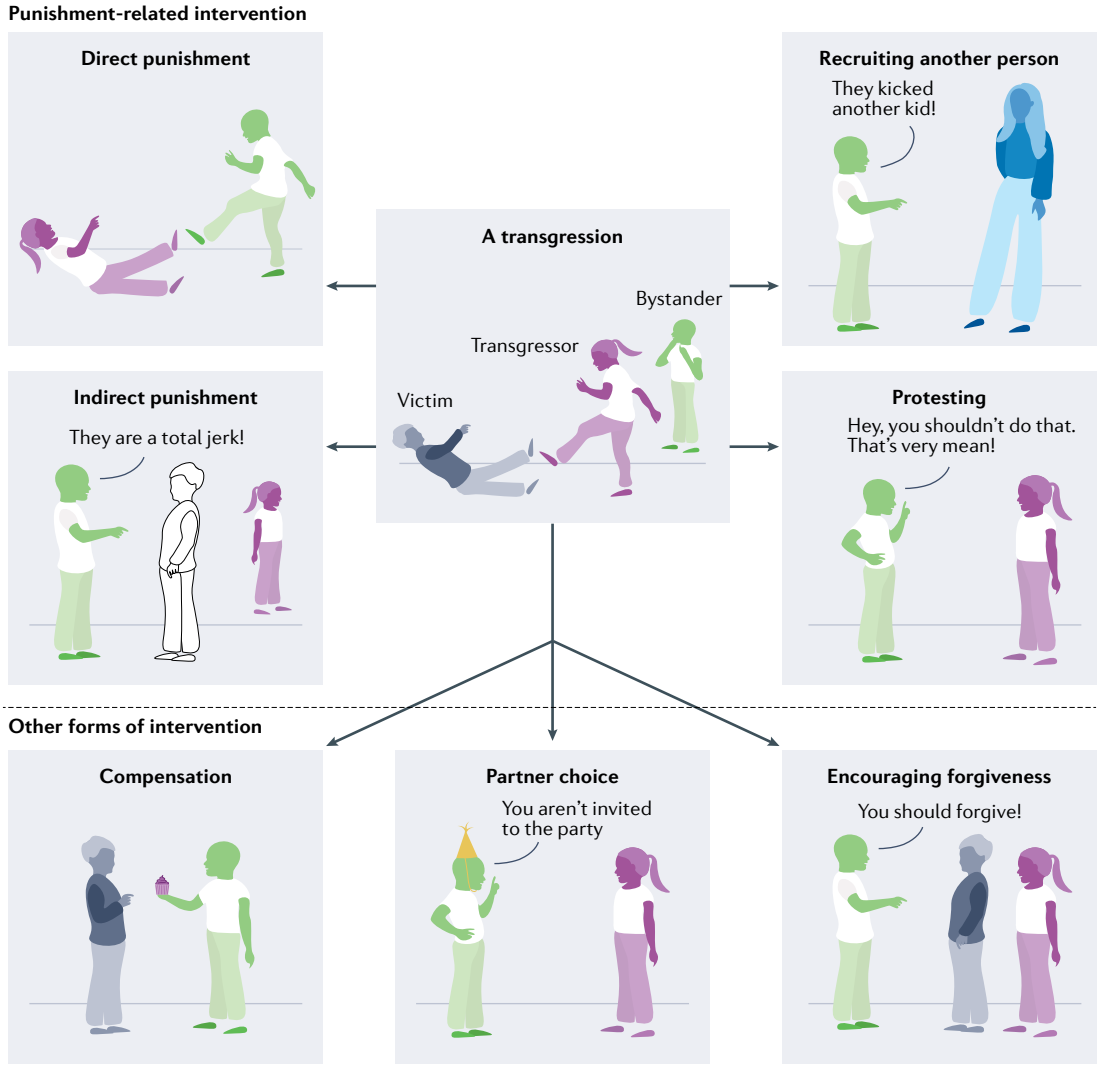

Fig. 1 | **Types of third-party intervention.** Punishment-related interventions include direct punishment, indirect punishment, recruiting another person and protesting. Other forms of intervention include compensation, partner choice and encouraging forgiveness.

motivational processes underlying modern institutions of justice. For instance, there are long-standing legal and philosophical debates regarding the moral value of punishment compared to other forms of intervention[18–21]. Although psychological data cannot adjudicate debates about the moral value of punishment, they can speak to whether people are likely to value and pursue certain forms of intervention, such as punishment, in response to transgressions. In particular, developmental psychology can provide insight into the types of intervention that are privileged early in ontogeny. For example, if children overwhelmingly favour punitive over restorative responses, that result could inform how certain interventions are maintained, promoted or discouraged in society.

In this Review, we synthesize research on children's reasoning about third-party intervention and their willingness to respond to transgressions, with a particular focus on punishment. We establish that infants and toddlers (0–2 years old) and children in both early childhood (approximately 3–5 years old) and middle-to-late childhood (approximately 6 years old to adolescence) are both assessors and agents of punishment. With respect to assessment, as early as infancy, children have rich expectations about when punishment will occur and evaluate those who punish transgressors positively. With respect to agency, children pursue the punishment of transgressors in a variety of contexts. Specifically, we discuss work showing that children in early childhood chastise wrongdoers for their misdeeds through verbal protest, recruit others to respond to transgressions, and directly punish perpetrators themselves. Next, we discuss the motivations behind children's early punishment behaviour. Together, this body of research suggests that third-party punishment is a key component of a child's toolkit for regulating social relationships and behaviour.

Children as assessors of punishment

Developmental psychologists have explored two broad questions regarding how infants, toddlers and children reason about situations involving transgressions. The first involves examining children's expectations

about bystanders who witness transgressions (FIG. 2a). Specifically, researchers have investigated who children expect to receive punishment and who children expect to pursue punishment. The second involves determining whether children evaluate punishing bystanders positively relative to non-punishing bystanders (FIG. 2b). Measuring both expectations and evaluations of punishment can provide insight into the degree to which people, even at a young age, have specific intuitions about the pursuit of punishment.

Expectations of third-party punishment. Researchers who study infants' and toddlers' expectations generally rely on non-verbal behavioural measures, such as violation-of-expectation paradigms, because children at such young ages are unable to respond to questions verbally or are inexperienced in doing so. In a typical version of the violation-of-expectation paradigm, participants are shown positive, negative or neutral social interactions among puppets, humans or animated characters. For example, participants might see a person steal a resource from another person. Next, participants watch additional scenes featuring characters interacting in positive or negative ways. For example, participants might see a person refusing to help another person. Participants' expectations are inferred based on how long they look at different scenes, under the assumption that they will look for longer at events that violate expectations[22,23]. For example, an infant looking for longer at an agent who did not punish a transgressor compared to an agent who did punish a transgressor would be taken as evidence that the infant expected the transgressor to be punished.

Studies using these violation-of-expectation paradigms have addressed two related questions concerning infants' and toddlers' reasoning about third-party punishment: whether children have expectations about who will receive punishment when it occurs, and who infants and toddlers expect to pursue punishment in response to wrongdoing.

In terms of who infants and toddlers expect to receive punishment, infants as young as six months old generally anticipate that bystanders will direct punishment toward transgressors rather than victims. For example, infants look for longer at scenes where a bystander harms a victim (an animated shape who has been pushed around) than at scenes where a bystander harms a transgressor (an animated shape who has pushed another shape around)[24]. This finding suggests that infants expect bystanders to direct punishment toward a transgressor rather than a victim. In a related line of work, 10-month-olds looked for longer at an event where a bystander gave a treat to an unfair agent than at an event where a bystander gave a treat to a fair agent[25]. This result suggests that infants expect others to withhold resources from unfair agents. Consistent with this work, a study with 13-month-olds and 15-month-olds that used longer looking times as a measure of association rather than violation of expectation found that participants differentially associated praise ("she's a good girl!") and admonishment ("she's a bad girl!") with fair and unfair individuals, respectively[26]. Together, this work suggests that children think transgressors are likely to receive punishment and unlikely to receive rewards.

Children not only expect transgressors to receive punishment but also expect bystanders who do not defend victims to receive punishment. In one set of studies[27,28], toddlers of around 21 months of age were shown two events. In one event, a transgressor harmed a victim and the bystander defended the victim by pushing the transgressor back. In the other event, a transgressor harmed a victim but the bystander did not defend the victim and instead just watched the transgression occur. A separate puppet watched these defending or non-defending events unfold in both cases. Next, participants viewed a scene where the separate onlooking puppet hit either the defending or non-defending puppet with a stick. Toddlers looked for longer when presented with the scene where the defending puppet received punishment compared to the scene where the non-defending puppet received punishment, suggesting that toddlers expect those who do not defend others to receive punishment themselves. These findings imply that by 21 months of age children expect so-called 'higher-order punishment' of non-defenders[29–32].

These studies demonstrate that infants and toddlers have specific expectations about the targets of punishment. Infants as young as six months old expect bystanders to direct punishment toward certain agents, such as transgressors[24] or those who fail to defend victims against transgressors[27,28]. However, these findings speak to whom children expect to be the targets of punishment when punishment is pursued; they do not speak to whether children expect bystanders to pursue punishment in the first place.

Research suggests that children expect bystanders only to pursue punishment in specific social situations, such as when the bystander and victim share group membership (see BOX 2). In one study[33], researchers presented 12-month-olds and 2.5-year-olds with three characters: a bystander, transgressor and victim. The transgressor and victim were always in different groups. In one condition, the bystander was in the same group as the victim; in another condition, the bystander was in

Box 1 | **Second-party punishment in children**

In addition to third-party punishment, developmental psychologists have also explored how children respond when they are the victims of transgressions, so-called 'second-party punishment'[119]. Because this work involves examining children's responses when they are victimized, studies largely investigate punishment of fairness violations rather than other violations, such as physical harm. Results show that although children willingly engage in second-party punishment, the emergence of second-party punishment differs from third-party punishment in several ways.

First, second-party punishment of selfishness might emerge before third-party punishment. In one study, children as young as four and a half years old willingly punished a selfish actor when directly victimized but were less likely to punish when someone else was the victim[120]. Second, adolescents are more likely to consider intent (whether someone divides resources unequally on purpose) when making decisions about second-party punishment compared to third-party punishment, where adolescents are less inclined to consider intent[121]. Third, desire for second-party punishment has been observed in both non-human primates[122] and young children[123], whereas third-party punishment has been observed only in humans[124] (BOX 3). Fourth, the emotional motivations underlying second-party punishment differ from the emotional antecedents of third-party punishment. Specifically, anger seems to play a part in promoting second-party punishment but not third-party punishment in 8-year-olds[125–127].

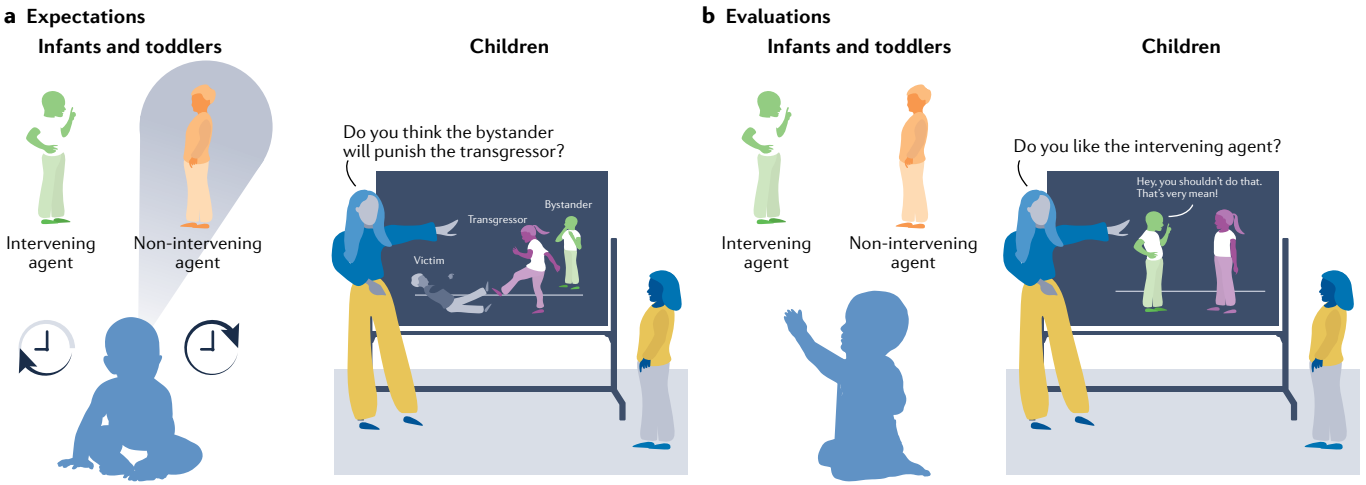

Fig. 2 | **Children as assessors of punishment. a** | Infants and toddlers look for longer at agents who do not intervene in response to a transgression, and children verbally report that transgressors will receive punishment. These findings suggest that infants, toddlers and children expect bystanders to pursue the punishment of transgressors. **b** | Infants and toddlers are more likely to reach for an intervening agent compared to a non-intervening agent, and children verbally report that they favour agents who intervene. These findings suggest that infants, toddlers and children evaluate those who intervene in response to a transgression positively.

the same group as the transgressor. Group membership was denoted by either labelling the groups with group names ('Topids' versus 'Jaybos') or by group-specific clothing (matching headbands). Participants subsequently saw scenes where the victim wanted to play with a toy and the transgressor stole that toy. Next, the transgressor was shown to need a puzzle piece that they could not reach but that the bystander could reach. In the help event, the bystander moved the puzzle piece closer to the transgressor. In the hinder event, the bystander threw the puzzle piece away. When the bystander shared group membership with the victim, both 12-month-olds and 2.5-year-olds looked longer at the help event compared to the hinder event, suggesting that children are surprised if a bystander helps a transgressor who had previously harmed an ingroup member. When the bystander shared group membership with the transgressor, participants looked longer at the hinder event compared to the help event, suggesting that children are surprised if the bystander hinders a transgressor who is in the same group as the bystander. These findings suggest that infants and toddlers hold specific expectations surrounding the group contexts in which bystanders will respond to transgressions.

Other work has examined whether infants' expectations of punishment are sensitive to social status. Indeed, children expect leaders to respond to wrongdoing but hold non-leaders to a lesser standard[34]. To establish this finding, participants saw puppet shows featuring three agents. In one condition, the protagonist puppet was established as the 'leader' (by exerting control over other puppets' behaviour or through their physical size). In another condition, the protagonist puppet was established as a non-leader. Next, all participants watched the protagonist present two blocks to a pair of puppets, one of whom stole both blocks, leaving the other with none (rather than sharing them equally). Children then watched one of two follow-up events. In one event, the protagonist took away one of the thief's blocks (intervention event). In the other event, the protagonist just looked at the blocks and did not do anything (non-intervention event). Seventeen-month-olds looked longer at the non-intervention event than the intervention event but only when the protagonist was portrayed as a leader. Furthermore, this pattern of results did not emerge when the victim explicitly acknowledged that they did not want the blocks in the first place, thereby making the stealing of blocks unproblematic. This work suggests that early in development, children expect leaders to punish transgressors by removing their resources.

Taken together, studies using looking time have revealed children's early expectations about third-party punishment. Infants as young as six months old expect bystanders to direct punishment (when it is pursued) towards specific individuals, such as transgressors[24]. Furthermore, children around one year of age expect certain individuals (such as those in a position of authority) to pursue punishment by not helping transgressors or by removing resources from transgressors[33,34]. Most of these studies have focused on infants and toddlers, but children in early-to-middle childhood also hold these same expectations. For instance, like infants and toddlers tested in violation of expectation paradigms, when explicitly asked whether bystanders will pursue punishment, 4- to 7-year-old children indicate that they expect third parties (especially authority figures) to punish transgressors[35].

Whereas soliciting younger children's expectations is constrained by their weaker language abilities, older children can not only answer questions about whether they expect transgressors to receive punishment but also whether and why transgressors should be punished. For example, foundational developmental research found that children mentioned punishment when asked to explain how individuals should respond to transgressions. This finding suggests that children believe that

third parties should punish transgressors[36]. Other foundational research found that, although children disagreed about what actions constitute transgressions, they generally agreed that transgressors should be punished[37]. More recent research finds that children as young as two years old judge that transgressors should "get in trouble" or receive punishment for their behaviour[38–43].

Furthermore, children of around four years old report that bystanders, particularly authority figures, are obligated to get transgressors in trouble for their misdeeds[35]. In addition, when given the opportunity to guide the behaviour of various dolls (a transgressor, victim and bystander), preschoolers verbally indicated that the bystander doll should punish the transgressor doll[44]. In summary, work in which children are explicitly asked about third-party intervention suggests that children expect bystanders to respond to wrongdoing in punitive ways and believe that transgressors deserve punishment[38–43].

Evaluations of third-party punishment. To assess infants' evaluations of third-party punishment, researchers sometimes use an infant-friendly task where, instead of measuring looking time towards a punishing or non-punishing agent (as in the violation of expectation paradigm), they measure which agent a child reaches for. The assumption is that children reach for the agent they prefer, providing a measure of evaluation.

The research investigating evaluations of those who respond to transgressions is considerably more limited than the work on expectations. However, one study suggests that infants and children favour agents who respond to transgressions compared to those who do not. Specifically, 6-month-olds reach toward agents who stop a transgressor from continually harming a victim compared to agents who do not intervene[24].

Beyond infancy, children of around five years of age favour those who verbally protest wrongdoing or tattle on transgressors over those who do not respond to transgressions when explicitly asked to judge bystanders[45–47]. Children evaluate direct punishment positively as well. For example, children regard those who punish others by sending a transgressor to 'time-out' as doing the right and fair thing[48,49]. Further, 8-year-olds indicate that they would prefer to live in a world with punishment than one without it, providing additional evidence that children value punishment[50]. Research on evaluations has also examined other types of third-party intervention in addition to punishment. Although this work has found that children consistently evaluate punishment positively, it also suggests that children — like adults[51] — evaluate some other forms of intervention, such as compensation, even more positively[52,53]. In sum, the few findings regarding children's evaluations of punishment (and other forms of third-party intervention) indicate that children generally evaluate those who engage in third-party punishment positively from early in development.

In summary, research from developmental science has found that starting as early as infancy, children expect bystanders to direct punishment towards specific individuals, such as transgressors[24], and expect specific bystanders (such as those in a position of authority) to pursue punishment[33,34]. Although the research is more limited, children also evaluate punishment positively, starting in early childhood[45–50,52,53]. Together, these findings demonstrate that, from early in life, children possess a set of specific beliefs about how others are likely to respond to transgressions.

Children as agents of punishment

Research on children's early-emerging expectations and evaluations of third-party intervention suggests that children actively reason about bystander intervention. However, this work does not speak to whether and when children begin to respond to wrongdoing themselves, and why children might pursue punishment.

Although children expect and positively evaluate third-party punishment, assessing third-party punishment does not require personal sacrifice. By contrast, intervening in response to transgressions typically does. For instance, bystanders who intervene often spend time and effort to do so and risk potential retaliation. These are sacrifices that children might be unwilling to incur, especially when they have not been personally victimized. Indeed, this unwillingness to accept personal cost is why many societies bestow the responsibility of punishment on authority figures[54–56]. Because of these potential costs, willingness to punish would strongly indicate a child's interest in actively responding to transgressions.

The evolution of intervention behaviour further suggests that children might not be willing to act as agents of punishment. Although adults across diverse societies willingly respond to wrongdoing, most notably via third-party punishment[1,4], this behaviour is not observed in non-human species (BOX 3). Given that third-party intervention behaviour is present across human societies yet absent in nonhuman animals, developmental research is especially relevant to understanding the origins of third-party punishment.

Box 2 | **Intergroup bias in third-party punishment**

Third-party punishment does not occur in a vacuum. Indeed, most punishment occurs within specific social contexts, such as between individuals who do or do not share similar social identities (such as gender, race, or political affiliation). Although formal notions of justice emphasize that punishment should be meted out impartially[128], psychological research with adults suggests that this is not what happens in reality[129–134]. Because bias in punishment disproportionately harms members of certain groups and their families[135], it is important to understand how children exhibit bias when punishing transgressors, and the psychological mechanisms that contribute to such bias.

To our knowledge, only a small set of studies have investigated intergroup bias in third-party punishment. One set of studies shows that 12-month-olds and 2.5-year-olds expect punishment to occur when an ingroup member is harmed but do not expect punishment to occur when an outgroup member is harmed[33]. Another set of studies show that children of around the age of four judge outgroup transgressors as more deserving of punishment than ingroup transgressors[136]. However, other research reveals that children of between 4 and 6 years old think that ingroup and outgroup transgressors deserve equal punishment regardless of the social identity of the victim[137].

Studies that have examined how intergroup bias interacts with punitive behaviour largely suggest that the identity of a transgressor shapes early third-party punishment[138]. However, the directionality of the effects is inconsistent. Some work finds that children are more likely to punish ingroup members compared to outgroup members[139], whereas other work finds that children are more likely to punish outgroup members compared to ingroup members[140]. Additional research on this topic will help to clarify the strength and directionality of group bias effects in the context of third-party punishment.

How children respond to transgressions. Developmental research has focused on children's willingness to pursue other forms of third-party intervention, such as protesting (verbally expressing disapproval of a behaviour) or recruiting a third-party to respond (tattling), in addition to investigating their punishment behaviour (FIG. 3).

When it comes to protesting, toddlers as young as three years old and children as old as eleven years old verbally protest when an agent destroys someone else's artwork[57–60], violates someone else's property rights[61], or steals someone's resources[62]. For example, they chastise transgressors for their misdeeds by saying things like, "No you're not supposed to do that!" (normative protest) or "No! Don't tear it!" (imperative protest[60]). Children as young as five years old also protest in response to those who do not adhere to prosocial norms, such as sharing resources[63]. Furthermore, 5- to 8-year-old children across small-scale and large-scale societies protest in response to those who break conventional rules, such as not playing a game properly[64]. These remarks might harm transgressors by causing embarrassment or shame (thereby characterizing such actions as more directly punitive). However, it is unclear whether children protest in an attempt to inflict such harm or rather to acknowledge wrongdoing without intending to cause embarrassment or shame. Nonetheless, children's willingness to engage in protest behaviour marks an important way that children respond to transgressions.

Box 3 | **Third-party punishment in non-human animals**

Aggression is common in non-human animals, and meets the definition of punishment when it comes at a cost to the aggressor and is in response to behaviour by a target[141]. However, unlike in human societies, where punishment occurs in both second-and third-party contexts, punishment in non-human animals seems to be restricted to second parties (BOX 1). That is, the limited research on third-party punishment in non-humans has not found evidence for it[124], suggesting that punishment by uninvolved bystanders might be unique to our species.

One example that approximates nonhuman third-party punishment is the punitive behaviour of the reef-dwelling blue-streaked cleaner wrasse (*Labroides dimidiatus*)[142]. Cleaner fish often work in male–female dyads to clean ectoparasites off 'client' fish who visit them at their so-called 'cleaning stations'. This behaviour is mutually beneficial to cleaners and clients because the cleaners get a meal while the client is cleaned. However, cleaners prefer to feed on the layer of protective mucus covering clients. Thus, eating ectoparasites represents a 'cooperative' choice by cleaners, whereas eating protective mucus represents a 'cheating' choice. This choice between cooperation and cheating resembles the Prisoner's dilemma[143]: cleaner fish must feed against their preference to cooperate yet are incentivized to cheat. In cases in which one partner (typically the female) cheats, the other (typically the male) can punish by chasing and biting the cheat, a response that promotes future cooperation. Although this behaviour does not meet the strict definition of third-party punishment (because a cooperative partner is affected by the cheat as they both lose access to the client), it is an excellent example of punishment in nonhuman animals in the domain of cooperation.

Another study more directly tested third-party punishment of theft in chimpanzees (*Pan troglodytes*)[124]. Chimpanzees demonstrate second-party punishment. Specifically, they are more likely to punish in a condition where a conspecific steals their food (theft condition) than in a condition where the experimenter removes their food and gives it to a conspecific (outcome disparity condition), even though in both cases the victim loses food[122]. These findings suggest that chimpanzees do punish, although so far this response has been observed only in the context of theft, not unfairness[144,145]. Building on the finding that chimpanzees punish theft when they are the victims, researchers tested whether a third-party observer chimpanzee would punish a thief who had stolen food from a victim[124]. Third-party chimpanzees were no more likely to intervene in the theft condition than in the control conditions, suggesting that chimpanzees do not punish as third parties.

In terms of recruiting another individual to respond to transgressions, observational and experimental data reveal that by five years of age children willingly tattle on their peers[65], although they are more likely to do so when they are the victim of the transgression as opposed to a witness[66]. One possible explanation for tattling is that children worry they will be blamed for the transgression and speak up out of self-interest. For example, children might tattle on a transgressor to ensure others know that they are not responsible for the misdeed. However, children will tattle on their peers even when they cannot possibly be blamed, suggesting there may be a less selfish motive for tattling. For example, perhaps children want an adult to punish the transgressor or help the victim[67]. However, which intervention children hope the adult will pursue in response to tattling is unclear from the existing research. Nevertheless, this work suggests that children recognize others' misbehaviour and are willing to intervene even when not directly affected.

Although protesting and tattling might result in punishment of transgressions down the line, these behaviours might not immediately impose costs on the transgressor. To establish when and in what contexts children act as punishers themselves, researchers construct scenarios that give participants the option to remove positive resources (such as candy or time with a fun activity[68,69]) or to allocate negative resources (such as bad-tasting candies[70]) to a transgressor (FIG. 3). Sometimes, but not always, the decision to punish is associated with a cost to the participant themselves[68], such as the participant sacrificing their own candies in exchange for a transgressor being punished.

Results show that children will punish a wide range of transgressions. For example, when prompted to remove a resource from either a hindering puppet or a helpful puppet, 19-month-olds prefer to remove a resource from the hindering puppet[71]. Similarly, 3-year-olds will remove stolen goods from a thief[72], and 4-year-olds will take valued stickers from someone who did not contribute to a public good (a free-rider), even if doing so requires giving up their own stickers[73]. Children around four years of age will punish those who ruin other people's artwork by removing positive resources (the opportunity to participate in a fun activity[69]) or allocating negative resources (bad-tasting treats) to the transgressor[70]. Further, children around five years of age will punish those who engage in disloyal behaviour (refusing to help a fellow teammate)[74]. In sum, children are willing to engage in third-party punishment in response to various transgressions, including hindering behaviour, theft, free-riding, property destruction and disloyalty.

Children are also willing to punish as third-parties in response to fairness norm violations (such as when someone does not share with others[75]). For example, 6-year-olds will sacrifice a valued resource (for example a tasty candy) to ensure that a selfish individual does not receive the same resource, although children younger than six years old typically do not[68,76–78]. Unlike studies on other types of transgressions (such as property destruction), research investigating children's willingness to punish selfishness has included children from diverse societies rather than exclusively

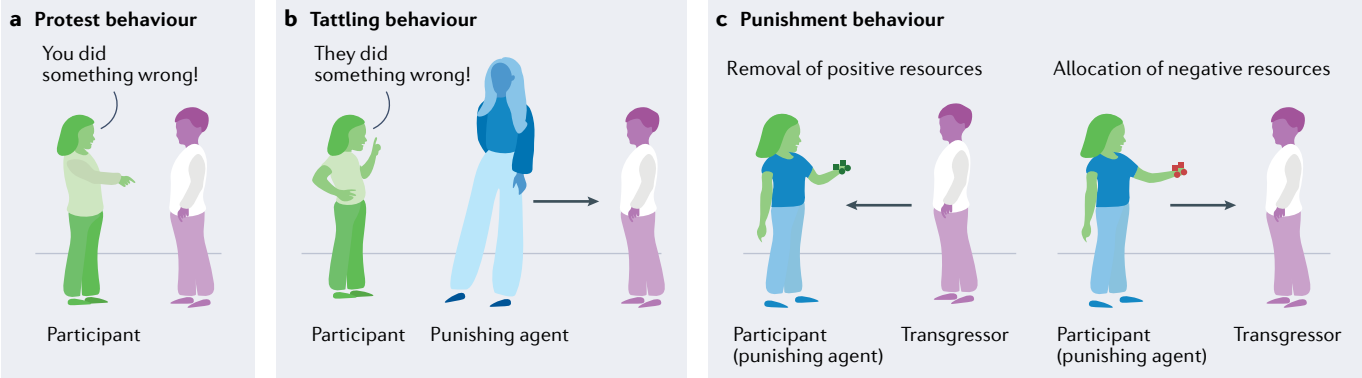

Fig. 3 | **Children as agents of punishment.** Children intervene against transgressions by verbally expressing disapproval (protesting behaviour; part **a**), reporting another person's misdeeds to a person of authority (tattling; part **b**), or third-party punishment behaviour (either removing positive rewards or allocating negative resources; part **c**).

Western children (who are typically the focus of developmental work; see REF.[79]). Children across various countries — including Argentina, Ecuador, Germany, India and the USA — exhibit a willingness to sacrifice their own resources to ensure that selfish others receive fewer resources. However, the precise age at which this willingness emerges varies across societies and, in some cases, does not present until middle childhood (around 9 years of age[80]). Because children younger than six years old do not systematically punish selfishness but children from middle childhood onward do, it is possible that cultural norms help to shape the emergence of punishment behaviours[80].

A strength of the work reviewed here is the use of experimental methods in well controlled laboratory settings, allowing researchers to clearly test whether children are willing to punish transgressors under specific conditions. However, these tasks might overestimate children's willingness to punish in non-experimental settings because children in everyday life are generally not given such obvious opportunities to punish transgressors. Experiments with children are also limited by the kinds of interactions that are ecologically valid for the age group. As such, the types of third-party punishment typically studied in young children (for example, protesting and tattling) are quite different from the more severe forms of punishment enacted by institutional systems (such as prison time). Nevertheless, the reviewed work on third-party punishment in children collectively illustrates how a core feature of modern institutional justice systems (a willingness to punish third-parties) resembles behaviours that emerge in early childhood.

Why children respond to transgressions. In considering why children are willing to serve as agents of punishment, it is useful to draw on philosophical discussions of punitive motives. Legal philosophers have long discussed why individuals pursue punishment of transgressors from a moral perspective[81]. Specifically, both consequentialist and retributive motives might underlie punitive behaviour. On the one hand, some argue that punishment should be grounded in consequentialist principles: transgressors should be punished not because they deserve punishment but because punishment can effectively deter the transgressor and others from future misdeeds[82–84]. On the other hand, some argue that punishment should be grounded in retributive principles: transgressors should be punished not because of any positive consequences stemming from punishment but because transgressors deserve punishment in proportion to their crimes[85–88].

Although developmental data cannot speak to the moral value of different punitive motives, it sheds light on the presence or absence of these motives in childhood. In turn, researchers can explore the extent to which children punish because they want to inflict deserved suffering, deter bad behaviour, or some combination of the two. Indeed, starting at around five years of age, children, like adults[89–95], are driven by multiple motives, including those that are both retributive and consequentialist[96,97]. Children make personal sacrifices (like giving up time to be spent playing a fun game) to punish transgressors, and they do so even when there is no clear personal or social benefit[96,97]. However, children more frequently punish when they believe punishment will teach the transgressor a lesson[96,97], either to specifically deter the transgressor or establish general norms about inappropriate behaviour for onlookers. Furthermore, children act more fairly rather than selfishly after being punished by a third-party for selfishness[98] and are more likely to cooperate in a social dilemma when faced with the threat of third-party punishment than when punishment is not possible[99]. These findings indicate that children are tuned to punishment's capacity to shape individual behaviour.

These studies on punitive motives are consistent with the idea that children's willingness to punish transgressors is sensitive to both retributive and consequentialist concerns. These findings connect to larger societal and philosophical discussions regarding whether punishment is unjust and whether society should abandon its punitive orientation and shift toward a more restorative system[18–21] that de-emphasizes deserved punishment and instead focuses on attending to victims' needs. In this way, restorative justice tends to take the focus away from the transgressor and reorients justice efforts toward the victim in the form of emotional and monetary compensation.

Some developmental research has found that children's intervention behaviour is sensitive to restorative justice concerns. For example, when given the opportunity to intervene in response to theft, participants as young as three years old preferred a victim-oriented option (removing a stolen item from a transgressor and returning it to the victim) rather than a self-oriented option (removing a stolen item from a transgressor and keeping it for themselves)[72,100]. In another study, 5- to 6-year-old children learned about a transgressor who stole a resource from a victim[101]. Participants were given the option to punish the transgressor without restoring the stolen resource to the victim, or to return the stolen resource to the victim without removing additional resources from the transgressor. Children tended to prefer the latter option, which only mildly punished the transgressor by removing the stolen resource and simultaneously restored justice by returning the stolen resource to its rightful owner. Finally, when children between 5 and 9 years old were given the option to punish a selfish individual by taking away their resources, they punished in ways that were broadly consistent with rectifying inequality[77]. Specifically, participants in this study preferred the restorative option that resulted in equal resources for the transgressor and victim to either removing all resources from both actors or to inflicting maximum punishment on the transgressor. These results suggest that children's willingness to intervene in response to transgressions is, at least in part, informed by restorative justice considerations.

These studies suggest that children take restorative justice concerns into account when responding to transgressions. However, they do not tell us whether children prefer restorative measures that affect only victims and not transgressors over more punitive ones that affect only transgressors but not victims because punishment and restoration are never entirely deconfounded in these studies[72,77,101]. That is, a purely restorative option (such as compensating victims without affecting the transgressor) is never pitted against a purely punitive option (such as removing resources from transgressors without affecting the victim). If participants are interested in restorative justice, they can only pursue it through punitive means that require removing resources from the transgressor.

Few studies have addressed whether children prefer victim-focused restoration to transgressor-focused punishment when restoration affects only victims. However, one study found that children favour punishment over restoration in the form of compensation[102]. Specifically, 6- to 9-year-old children were presented with a selfish agent who did not want to share resources with a victim and were given two intervention options: participants could sacrifice one of their resources either to remove the selfish agent's resources without affecting the victim's resources (punishment option) or to compensate the victim for their losses without affecting the transgressor's resources (restoration option); participants were also given the opportunity to do nothing. Children preferred to invest in punishing the selfish agent rather than compensating the victim or doing nothing, suggesting that children see more value in inflicting harm on transgressors than restoring justice to victims when such options are pitted against one another[102].

In sum, children's early intervention behaviour appears to be sensitive to various factors that ultimately have strong connections to long-standing complex legal and philosophical issues, including notions of retribution, consequentialism and restoration[103]. For example, children are especially interested in punishing transgressors to inflict suffering (retribution) and to deter bad behaviour (consequentialism[96,97]). Additionally, children will pursue intervention options that involve an element of punishment while also restoring justice to the victim[72,77,101]. Further, when restoration affects only victims and does not affect transgressors, children prefer a more punitive option to a more restorative option when both options are available[102]. This finding provides evidence that, although children are concerned with restorative justice, they are especially interested in ensuring that transgressors are punished for their misdeeds.

Summary and future directions

Research demonstrates that key building blocks of our modern institutions of justice, such as third-party punishment, are evident in early child development. As illustrated in FIG. 4, infants and toddlers (0 to 2 years old) have clear intuitions about who is likely to be punished[24–28] and who is likely to punish[33,34]. Similarly, children in both early childhood (approximately 3–5 years old) and middle-to-late childhood (approximately 6 years old to adolescence) exhibit expectations consistent with the notion that certain individuals, particularly people in a position of authority, will and should intervene in response to transgressions[35–43]. Beyond expectations, children also positively evaluate those who intervene to prevent harm[24] and — once they reach an age where they can verbalize their opinions — they explicitly evaluate punishers positively[45–49,52,53].

In terms of whether children will pursue third-party punishment themselves, the studies reviewed here suggest that intervention in response to transgressions occurs early in development. Children spontaneously protest misbehaviour[57–64] and report others' misdeeds by tattling[65–67]. Furthermore, children will punish a variety of transgressions, including fairness violations (for example, not sharing resources[68]), property violations (for example, ripping up someone else's artwork[69]), free-riding (for example, not contributing to a group project[73]), and disloyalty (for example, refusing to help a group member[74]). Importantly, this penchant for punishment emerges even when punishment requires sacrifice[68], across cultures[80], and when the option to compensate victims is also available[102].

Although ample research from developmental science demonstrates that infants, toddlers and children reason about punishment as assessors and act as agents of punishment themselves, this work has limitations. First, most of the studies reviewed here rely on samples from children residing in Western, educated, industrialized, rich and democratic (WEIRD[104]) societies, with some notable exceptions[64,80]. Children from diverse populations must be recruited in future research to broaden

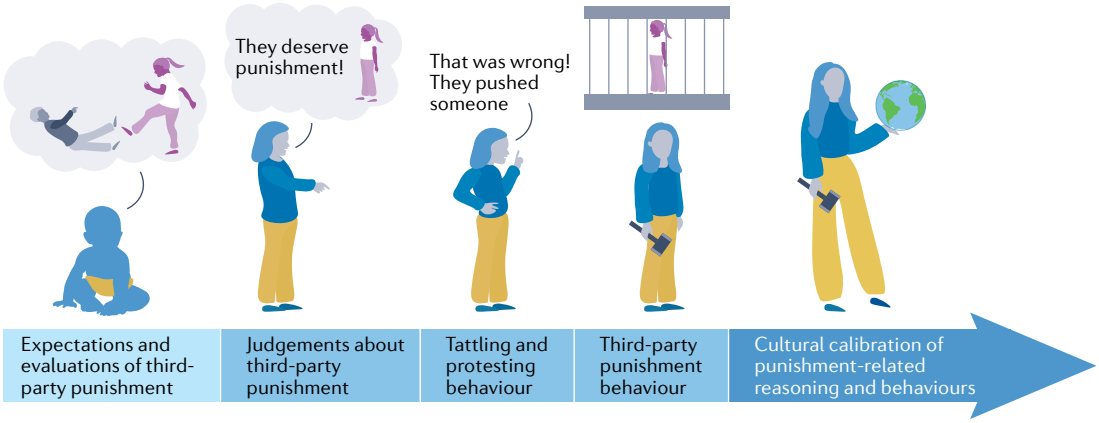

Fig. 4 | **Development of third-party punishment behaviour.** Infants, toddlers and children expect bystanders, especially authority figures such as teachers, to punish transgressors, and value bystanders who intervene in response to transgressors. At around middle childhood, children explicitly acknowledge that transgressors deserve punishment, spontaneously call out misdeeds via verbal protest and tattling behaviour, and will engage with transgressors via third-party punishment. Finally, social and cultural learning shape expectations and evaluations of punishment in addition to the manifestation of third-party punishment behaviour. The emergence of each specified capability is additive; for example, children still maintain expectations and evaluations of third-party punishment even when they begin protesting wrongdoing.

our understanding of children's early reasoning about punishment and their punitive behaviour.

Second, although the literature reviewed here speaks to how third-party intervention takes shape across broad age ranges, it cannot adequately home in on exact ages at which certain conceptualizations or behavioural inclinations emerge. For instance, children tend to intervene against property destruction at around four years of age[69], a few years before intervening against unfairness, which emerges at around 6 years of age, or even later, depending on cultural context[80]. These findings suggest that children's willingness to intervene might be influenced by transgression type (or perhaps transgression severity[105]). Thus, the transgression types being studied might influence inferences about the apparent age at which third-party punishment emerges. This issue of developmental precision is further exacerbated by arbitrary a priori sampling decisions (for example, a study might test 5-year-olds and 8-year-olds) that influence which ages are associated with which behaviours. That is, the ages associated with behaviours are limited to the ages recruited and tested, rather than reflecting the ages associated with behaviours following comprehensive sampling and testing of all possible ages. For these reasons, the current research is best suited to speak to broad developmental patterns in the emergence of conceptualizations of punishment and punishment behaviour itself. We encourage researchers to consider testing wide age ranges in future work to generate a clearer developmental picture of when intervention-related judgements and behaviours develop.

Third, the types of transgressions presented to children in developmental studies are — for obvious ethical and practical reasons — substantially less severe than crimes that the criminal justice system typically confronts. Although these differences are challenging to address from an experimental standpoint, research on how children reason about legal punishment finds that children do understand the terms 'jail' and 'prison'[106–108]. Future research should continue to integrate real-world questions about incarceration and punishment to better tether conversations about institutional punishment to early intuitions and behaviours in development.

Fourth, most research on punishment involves substantial experimental scaffolding, meaning that children's decisions to punish are almost always elicited by experimenters in laboratory contexts rather than naturalistically observed in the world. In an attempt to understand children's more spontaneous punishment behaviour, one study investigated whether 4-year-olds to 6-year-olds punish transgressors in a more open-ended task that allowed participants to freely engage with a transgressor in a variety of ways, only one of which was punitive (hitting). Interestingly, participants in this more open-ended task were not particularly punitive, suggesting that children may be less inclined to exhibit third-party punishment in less structured tasks[109]. None of the work reviewed here — with the exception of one study[66] which examined children's tattling in ecologically valid contexts — can speak to children's willingness to punish transgressors in non-experimental settings. On this note, there is debate in the adult literature about the extent to which adult third-party punishment behaviour truly reflects a willingness to engage in peer punishment in everyday life[12,13,110,111] (see REF.[112] for a review). Many of the same concerns raised in this adult work apply to developmental research. However, at a minimum, developmental work suggests that, regardless of whether children would actually punish their peers spontaneously in the real world, they will often do so when asked directly, reflecting an early willingness to engage in third-party punishment when doing so is an option.

Finally, there is a rich and lively debate about why third-party punishment evolved in humans. According to one view, third-party punishment evolved within communities to support group norm adherence[113,114]. Another view is that third-party punishment evolved to confer immediate reputational benefits to punishers by

advertising punishers as valuable social partners[115–117]. Yet another possibility is that third-party punishment evolved as a byproduct of an appetite for second-party punishment[118] (BOX 1). According to this perspective, empathy for victims results in a desire to punish transgressors in third-party contexts. Similarly, institutional third-party punishment might be a proxy for second-party punishment insofar as institutional actors identify with the victim and act on their behalf to rectify injustices. Although these possibilities raise interesting questions about the nature of punishment, the work reviewed here is broadly consistent with each of these possibilities and cannot adjudicate between them. Future research could better engage with these debates by examining how capacities such as empathy and perspective-taking shape the pursuit of third-party punishment. As we argue here, developmental research can help provide a more complete picture of why people engage in third-party punishment by shedding light on when and how punishment is expressed in its earliest forms.

Published online: 04 April 2022

1. Fehr, E. & Fischbacher, U. Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004).
2. Raihani, N. J., Thornton, A. & Bshary, R. Punishment and cooperation in nature. *Trends Ecol. Evol.* **27**, 288–295 (2012).
3. Vidmar, N. & Miller, D. T. Social psychological processes underlying attitudes toward legal punishment. *Law Soc. Rev.* **14**, 565–602 (1980).
4. Henrich, J. et al. Costly punishment across human societies. *Science* **312**, 1767–1770 (2006).
5. Balliet, D., Mulder, L. B. & Van Lange, P. A. Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594–615 (2011).
6. Boyd, R., Gintis, H. & Bowles, S. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **328**, 617–620 (2010).
7. Carpenter, J. & Matthews, P. H. What norms trigger punishment? *Exp. Econ.* **1**, 272–288 (2009).
8. Cushman, F. Punishment in humans: from intuitions to institutions. *Phil. Compass* **10**, 117–133 (2015).
9. Gächter, S., Renner, E. & Sefton, M. The long-run benefits of punishment. *Science* **322**, 1510–1510 (2008).
10. Yamagishi, T. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116 (1986).
11. Henrich, J. & Muthukrishna, M. The origins and psychology of human cooperation. *Annu. Rev. Psychol.* **72**, 207–240 (2021).
12. Hofmann, W., Brandt, M. J., Wisneski, D. C., Rockenbach, B. & Skitka, L. J. Moral punishment in everyday life. *Pers. Soc. Psychol. B* **44**, 1697–1711 (2018).
13. Molho, C., Tybur, J. M., Van Lange, P. A. & Balliet, D. Direct and indirect punishment of norm violations in daily life. *Nat. Commun.* **11**, 1–9 (2020).
14. Heffner, J. & FeldmanHall, O. Why we don't always punish: preferences for non-punitive responses to moral violations. *Sci. Rep.* **9**, 1–13 (2019).
15. Martin, J. W. & Cushman, F. To punish or to leave: distinct cognitive processes underlie partner control and partner choice behaviors. *PLoS ONE* **10**, e0125193 (2015).
16. McCullough, M. E. Forgiveness: who does it and how do they do it? *Curr. Dir. Psychol. Sci.* **10**, 194–197 (2001).
17. Theunissen, M. H., Vogels, A. G. & Reijneveld, S. A. Punishment and reward in parental discipline for children aged 5 to 6 years: prevalence and groups at risk. *Acad. Pediatr.* **15**, 96–102 (2015).
18. Johnstone, G. in *Restorative Justice: Ideas, Values, Debates* (Routledge, 2013).
19. Marshall, T. F. in *Restorative Justice: An Overview* (Home Office, 1999).
20. Van Ness, D. & Strong, K. H. in *Restoring Justice: An Introduction to Restorative Justice* (Routledge, 2014).
21. Wiessner, P. The role of third parties in norm enforcement in customary courts among the Enga of Papua New Guinea. *Proc. Natl Acad. Sci. USA* **117**, 32320–32328 (2020).
22. Baillargeon, R. Innate ideas revisited: for a principle of persistence in infants' physical reasoning. *Perspect. Psychol. Sci.* **3**, 2–13 (2008).
23. Sim, Z. L. & Xu, F. Infants preferentially approach and explore the unexpected. *Br. J. Dev. Psychol.* **35**, 596–608 (2017).
24. Kanakogi, Y. et al. Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat. Hum. Behav.* **1**, 1–7 (2017).
25. Meristo, M. & Surian, L. Do infants detect indirect reciprocity? *Cognition* **129**, 102–113 (2013).
26. DesChamps, T. D., Eason, A. E. & Sommerville, J. A. Infants associate praise and admonishment with fair and unfair individuals. *Infancy* **21**, 478–504 (2016).
27. Geraci, A. & Surian, L. Toddlers' expectations of third-party punishments and rewards following an act of aggression. *Aggress. Behav.* **47**, 521–529 (2021).
28. Geraci, A. Toddlers' expectations of corporal third-party punishments against the non-defender puppet. *J. Exp. Child. Psychol.* **210**, 105199 (2021).
29. Fu, T., Ji, Y., Kamei, K. & Putterman, L. Punishment can support cooperation even when punishable. *Econ. Lett.* **154**, 84–87 (2017).
30. Henrich, J. & Boyd, R. Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* **208**, 79–89 (2001).
31. Kiyonari, T. & Barclay, P. Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment. *J. Pers. Soc. Psychol.* **95**, 826–842 (2008).
32. Martin, J. W., Jordan, J. J., Rand, D. G. & Cushman, F. When do we punish people who don't? *Cognition* **193**, 104040 (2019).
33. Ting, F., He, Z. & Baillargeon, R. Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proc. Natl Acad. Sci. USA* **116**, 6025–6034 (2019).
34. Stavans, M. & Baillargeon, R. Infants expect leaders to right wrongs. *Proc. Natl Acad. Sci. USA* **116**, 16292–16301 (2019).
35. Marshall, J., Mermin-Bunnell, K. & Bloom, P. Developing judgments about peers' obligation to intervene. *Cognition* **201**, 104215 (2020).
36. Piaget, J. *The Moral Judgment of the Child* (Routledge, 2013).
37. Kohlberg, L. & Kramer, R. Continuities and discontinuities in childhood and adult moral development. *Hum. Dev.* **12**, 93–120 (1969).
38. Barchard, K. & Atkins, C. Children's decisions about naughtiness and punishment: dominance of expiatory punishments. *J. Res. Child. Educ.* **5**, 109–115 (1991).
39. Cushman, F., Sheketoff, R., Wharton, S. & Carey, S. The development of intent-based moral judgment. *Cognition* **127**, 6–21 (2013).
40. Killen, M., Mulvey, K. L., Richardson, C., Jampol, N. & Woodward, A. The accidental transgressor: morally-relevant theory of mind. *Cognition* **119**, 197–215 (2011).
41. Van de Vondervoort, J. W. & Hamlin, J. K. Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *J. Exp. Child. Psychol.* **164**, 136–151 (2017).
42. Smetana, J. G. Preschool children's conceptions of moral and social rules. *Child. Dev.* **52**, 1333–1336 (1981).
43. Smith, C. E. & Warneken, F. Children's reasoning about distributive and retributive justice across development. *Dev. Psychol.* **52**, 613–628 (2016).
44. Kenward, B. & Östh, T. Enactment of third-party punishment by 4-year-olds. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2012.00373 (2012).
45. Loke, I. C., Heyman, G. D., Forgie, J., McCarthy, A. & Lee, K. Children's moral evaluations of reporting the transgressions of peers: age differences in evaluations of tattling. *Dev. Psychol.* **47**, 1757–1762 (2011).
46. Loke, I., Heyman, G. D., Itakura, S., Toriyama, R. & Lee, K. Japanese and American children's moral evaluations of reporting on transgressions. *Dev. Psychol.* **50**, 1520–1531 (2014).
47. Vaish, A., Herrmann, E., Markmann, C. & Tomasello, M. Preschoolers value those who sanction non-cooperators. *Cognition* **153**, 43–51 (2016).
48. Vittrup, B. & Holden, G. W. Children's assessments of corporal punishment and other disciplinary practices: the role of age, race, SES, and exposure to spanking. *J. Appl. Dev. Psychol.* **31**, 211–220 (2010).
49. Catron, T. F. & Masters, J. C. Mothers' and children's conceptualizations of corporal punishment. *Child. Dev.* **64**, 1815–1828 (1993).
50. Bregant, J., Shaw, A. & Kinzler, K. D. Intuitive jurisprudence: early reasoning about the functions of punishment. *J. Empir. Legal Stud.* **13**, 693–717 (2016).
51. Dhaliwal, N. A., Patil, I. & Cushman, F. Reputational and cooperative benefits of third-party compensation. *Organ. Behav. Hum. Dec.* **164**, 27–51 (2021).
52. Lee, Y. E. & Warneken, F. Children's evaluations of third-party responses to unfairness: children prefer helping over punishment. *Cognition* **205**, 104374 (2020).
53. Liu, X., Yang, X. & Wu, Z. To punish or to restore: how children evaluate victims' responses to immorality. Front. Psychol. 12, 696160 (2021).
54. Gross, J., Méder, Z. Z., Okamoto-Barth, S. & Riedl, A. Building the Leviathan — voluntary centralisation of punishment power sustains cooperation in humans. *Sci. Rep.* **6**, 1–9 (2016).
55. Hilbe, C., Traulsen, A., Röhl, T. & Milinski, M. Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. *Proc. Natl Acad. Sci. USA* **111**, 752–756 (2014).
56. Pfattheicher, S., Boehm, R. & Kesberg, R. The advantage of democratic peer punishment in sustaining cooperation within groups. *J. Behav. Decis. Mak.* **31**, 562–571 (2018).
57. Schmidt, M. F. & Tomasello, M. Young children enforce social norms. *Curr. Dir. Psychol. Sci.* **21**, 232–236 (2012).
58. Schmidt, M. F., Rakoczy, H. & Tomasello, M. Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition* **124**, 325–333 (2012).
59. Heyman, G. D., Loke, I. C. & Lee, K. Children spontaneously police adults' transgressions. *J. Exp. Child. Psychol.* **150**, 155–164 (2016).
60. Vaish, A., Missana, M. & Tomasello, M. Three-year-old children intervene in third-party moral transgressions. *Br. J. Dev. Psychol.* **29**, 124–130 (2011).
61. Rossano, F., Rakoczy, H. & Tomasello, M. Young children's understanding of violations of property rights. *Cognition* **121**, 219–227 (2011).
62. Josephs, M., Kushnir, T., Gräfenhain, M. & Rakoczy, H. Children protest moral and conventional violations more when they believe actions are freely chosen. *J. Exp. Child. Psychol.* **141**, 247–255 (2016).
63. Friedrich, J. P. & Schmidt, M. F. Preschoolers agree to and enforce prosocial, but not selfish, sharing norms. *J. Exp. Child. Psychol.* **214**, 105303 (2022).
64. Kanngiesser, P. et al. Children across societies enforce conventional norms but in culturally variable ways. *Proc. Natl Acad. Sci. USA* **119**, e2112521118 (2022).
65. Misch, A., Over, H. & Carpenter, M. The whistleblower's dilemma in young children: when loyalty trumps other moral concerns. *Front. Psychol.* **9**, 250 (2018).
66. Ingram, G. P. & Bering, J. M. Children's tattling: the reporting of everyday norm violations in preschool settings. *Child. Dev.* **81**, 945–957 (2010).
67. Yucel, M. & Vaish, A. Young children tattle to enforce moral norms. *Soc. Dev.* **27**, 924–936 (2018).
68. McAuliffe, K., Jordan, J. J. & Warneken, F. Costly third-party punishment in young children. *Cognition* **134**, 1–10 (2015).
69. Yudkin, D. A., Van Bavel, J. J. & Rhodes, M. Young children police group members at personal cost. *J. Exp. Psychol. Gen.* **149**, 182–191 (2020).

70. Kenward, B. & Östh, T. Five-year-olds punish antisocial adults. *Aggress. Behav.* **41**, 413–420 (2015).
71. 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).
72. Riedl, K., Jensen, K., Call, J. & Tomasello, M. Restorative justice in children. *Curr. Biol.* **25**, 1731–1735 (2015).
73. 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).
74. Arini, R., Wiggs, L. & Kenward, B. Moral duty and equalization concerns motivate children's third-party punishment. *Dev. Sci.* **57**, 1325–1341 (2021).
75. McAuliffe, K., Blake, P. R., Steinbeis, N. & Warneken, F. The developmental foundations of human fairness. *Nat. Hum. Behav.* **1**, 1–9 (2017).
76. Ziv, T., Whiteman, J. D. & Sommerville, J. A. Toddlers' interventions toward fair and unfair individuals. *Cognition* **214**, 104781 (2021).
77. Lee, Y. & Warneken, F. Does third-party punishment in children aim at equality? *Dev. Psychol.* https://doi.org/10.1037/dev0001331 (2022).
78. Salali, G. D., Juda, M. & Henrich, J. Transmission and development of costly punishment in children. *Evol. Hum. Behav.* **36**, 86–94 (2015).
79. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child. Psychol.* **162**, 31–38 (2017).
80. House, B. R. et al. Social norms and cultural diversity in the development of third-party punishment. *Proc. R. Soc. B* **287**, 20192794 (2020).
81. Michael, M. A. Utilitarianism and retributivism: what's the difference? *Am. Phil. Quart.* **29**, 173–182 (1992).
82. Amarasekara, K. & Bagaric, M. The errors of retributivism. *Melb. Univ. Law Rev.* **24**, 124–189 (2000).
83. Bentham, J. in *What is Justice? Classic and Contemporary Readings* (eds Soloman, R. C. & Murphy, M. C.) 215–220 (Oxford Univ. Press, 2000).
84. Christopher, R. L. Deterring retributivism: the injustice of just punishment. *Northwest. Univ. Law Rev.* **96**, 843 (2001).
85. Kant, I. in *Why Punish? How Much? A Reader on Punishment* (ed. Tonry, M. H.) 31–36 (Oxford Univ. Press, 2011).
86. Berman, M. N. Rehabilitating retributivism. *Law. Phil.* **32**, 83–10 (2013).
87. Kershnar, S. A defense of retributivism. *Int. J. Appl. Phil.* **14**, 97–117 (2000).
88. Moore, M. S. Justifying retributivism. *Isr. Law Rev.* **27**, 15–49 (1993).
89. 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).
90. Crockett, M. J., Özdemir, Y. & Fehr, E. The value of vengeance and the demand for deterrence. *J. Exp. Psychol. Gen.* **143**, 2279–2286 (2014).
91. Goodwin, G. P. & Gromet, D. M. Punishment. *Wiley Interdiscip. Rev. Cogn. Sci.* **5**, 561–572 (2014).
92. 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).
93. Nadelhoffer, T., Heshmati, S., Kaplan, D. & Nichols, S. Folk retributivism and the communication confound. *Econ. Phil.* **29**, 235–261 (2013).
94. Nikiforakis, N. & Normann, H. T. A comparative statics analysis of punishment in public-good experiments. *Exp. Econ.* **11**, 358–369 (2008).
95. Ouss, A. & Peysakhovich, A. When punishment doesn't pay: cold glow and decisions to punish. *J. Law Econ.* **58**, 625–655 (2015).
96. Marshall, J., Yudkin, D. A. & Crockett, M. J. Children punish third parties to satisfy both consequentialist and retributive motives. *Nat. Hum. Behav.* **5**, 361–368 (2021).
97. Twardawski, M. & Hilbig, B. E. The motivational basis of third-party punishment in children. *PLoS One* **15**, e0241919 (2020).
98. Martin, J. W., Martin, S. & McAuliffe, K. Third-party punishment promotes fairness in children. *Dev. Psychol.* **57**, 927–939 (2021).
99. Lergetporer, P., Angerer, S., Glätzle-Rützler, D. & Sutter, M. Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. *Proc. Natl Acad. Sci. USA* **111**, 6916–6921 (2014).
100. Van de Vondervoort, J. W. & Hamlin, J. K. Young children remedy second-and third-party ownership violations. *Trends Cogn. Sci.* **19**, 490–491 (2015).
101. Yang, X., Wu, Z. & Dunham, Y. Children's restorative justice in an intergroup context. *Soc. Dev.* **30**, 663–683 (2021).
102. McAuliffe, K. & Dunham, Y. Children favor punishment over restoration. Dev. Sci. 24, e13093 (2021).
103. Caruso, G. D. *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (Cambridge Univ. Press, 2021).
104. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
105. Yucel, M., Drell, M. B., Jaswal, V. K. & Vaish, A. Young children do not perceive distributional fairness as a moral norm. *Dev. Psychol.* (in the press).
106. Dunlea, J. P. & Heiphetz, L. Children's and adults' understanding of punishment and the criminal justice system. *J. Exp. Soc. Psychol.* **87**, 103913 (2020).
107. Dunlea, J. P. & Heiphetz, L. Children's and adults' views of punishment as a path to redemption. *Child. Dev.* **92**, e393–e415 (2021).
108. Dunlea, J. P. & Heiphetz, L. Moral psychology as a necessary bridge between social cognition and law. *Soc. Cogn.* **39**, 183–199 (2021).
109. Marshall, J., Gollwitzer, A., Wynn, K. & Bloom, P. The development of corporal third-party punishment. *Cognition* **190**, 221–229 (2019).
110. Pedersen, E. J., McAuliffe, W. H. & McCullough, M. E. The unresponsive avenger: more evidence that disinterested third parties do not punish altruistically. *J. Exp. Psychol. Gen.* **147**, 514–544 (2018).
111. Pedersen, E. J. et al. When and why do third parties punish outside of the lab? A cross-cultural recall study. *Soc. Psychol. Pers. Sci.* **11**, 846–853 (2020).
112. Guala, F. Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* **35**, 1–15 (2010).
113. Boyd, R. & Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **13**, 171–195 (1992).
114. Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
115. Jordan, J. J. & Rand, D. G. Signaling when no one is watching: a reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *J. Pers. Soc. Psychol.* **118**, 57–88 (2020).
116. Jordan, J. J., Hoffman, M., Bloom, P. & Rand, D. G. Third-party punishment as a costly signal of trustworthiness. *Nature* **530**, 473–476 (2016).
117. Barclay, P. Reputational benefits for altruistic punishment. *Evol. Hum. Behav.* **27**, 325–344 (2006).
118. Bloom, P. *Just Babies: The Origins of Good and Evil* (Crown Publishers, 2013).
119. Sally, D. & Hill, E. The development of interpersonal strategy: autism, theory-of-mind, cooperation and fairness. *J. Econ. Psychol.* **27**, 73–97 (2006).
120. Bernhard, R. M., Martin, J. W. & Warneken, F. Why do children punish? Fair outcomes matter more than intent in children's second-and third-party punishment. *J. Exp. Child. Psychol.* **200**, 104909 (2020).
121. Gummerum, M. & Chu, M. T. Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior. *Cognition* **133**, 97–103 (2014).
122. Jensen, K., Call, J. & Tomasello, M. Chimpanzees are vengeful but not spiteful. *Proc. Natl Acad. Sci. USA* **104**, 13046–13050 (2007).
123. Mendes, N., Steinbeis, N., Bueno-Guerra, N., Call, J. & Singer, T. Preschool children and chimpanzees incur costs to watch punishment of antisocial others. *Nat. Hum. Behav.* **2**, 45–51 (2018).
124. Riedl, K., Jensen, K., Call, J. & Tomasello, M. No third-party punishment in chimpanzees. *Proc. Natl Acad. Sci. USA* **109**, 14824–14829 (2012).
125. Gummerum, M., Lopez-Perez, B., Van Dijk, E. & Van Dillen, L. F. When punishment is emotion-driven: children's, adolescents', and adults' costly punishment of unfair allocations. *Soc. Dev.* **29**, 126–142 (2020).
126. Gummerum, M., Lopez-Perez, B., Van Dijk, E. & Van Dillen, L. F. Ire and punishment: incidental anger and costly punishment in children, adolescents, and adults. *J. Exp. Child. Psychol.* **218**, 105376 (2022).
127. Marshall., J. & McAuliffe, K. in *The Oxford Handbook of Evolution and Emotions* (ed. Al-Shawaf, L. & Shackelford, T. K.) (Oxford Univ. Press, in the press).
128. Spohn, C. *How Do Judges Decide?: The Search for Fairness and Justice in Punishment* (Sage, 2008).
129. Baumgartner, T., Götte, L., Gügler, R. & Fehr, E. The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Hum. Brain Mapp.* **33**, 1452–1469 (2012).
130. Brunson, R. K. & Miller, J. Young black men and urban policing in the United States. *Br. J. Criminol.* **46**, 613–640 (2006).
131. Delton, A. W. & Krasnow, M. M. The psychology of deterrence explains why group membership matters for third-party punishment. *Evol. Hum. Behav.* **38**, 734–743 (2017).
132. Guo, R., Ding, J. & Wu, Z. How intergroup relation moderates group bias in third-party punishment. *Acta Psychol.* **205**, 103055 (2020).
133. Lieberman, D. & Linke, L. The effect of social category on third party punishment. *Evol. Psychol.* **5**, 147470490700500203 (2007).
134. Okonofua, J. A. & Eberhardt, J. L. Two strikes: race and the disciplining of young students. *Psychol. Sci.* **26**, 617–624 (2015).
135. Rucker, J. M. & Richeson, J. A. Toward an understanding of structural racism: implications for criminal justice. *Science* **374**, 286–290 (2021).
136. Chapman, M. S., May, K. E., Scofield, J., DeCoster, J. & Bui, C. Does group membership affect children's judgments of social transgressions? *J. Exp. Child. Psychol.* **189**, 104695 (2020).
137. Schuhmacher, N. & Kärtner, J. Preschoolers prefer in-group to out-group members, but equally condemn their immoral acts. *Soc. Dev.* **28**, 1074–1094 (2019).
138. McAuliffe, K. & Dunham, Y. Group bias in cooperative norm enforcement. *Phil. Trans. R. Soc. Lond. B* **371**, 20150073 (2016).
139. Gonzalez-Gadea, M. L., Dominguez, A. & Petroni, A. Children's group biases in third-party punishment are guided by norms-focused behaviors. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/pxyw4 (2020).
140. 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).
141. Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
142. Raihani, N. J., Grutter, A. S. & Bshary, R. Punishers benefit from third-party punishment in fish. *Science* **327**, 171–171 (2010).
143. Bshary, R., Grutter, A. S., Willener, A. S. & Leimar, O. Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature* **455**, 964–966 (2008).
144. Jensen, K., Call, J. & Tomasello, M. Chimpanzees are rational maximizers in an ultimatum game. *Science* **318**, 107–109 (2007).
145. Proctor, D., Williamson, R. A., de Waal, F. B. & Brosnan, S. F. Chimpanzees play the ultimatum game. *Proc. Natl Acad. Sci. USA* **110**, 2070–2075 (2013).

Acknowledgements
The authors would like to thank the members of the Cooperation Lab, F. Warneken, Y. Lee and F. Ting for their feedback on this manuscript.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Psychology thanks Maria Gonzalez-Gadea and the other, anonymous, reviewers for their contribution to the peer review of this work.

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