Classification: Social Sciences, Psychological and Cognitive Sciences

Title (135): Toddlers Draw Broad Negative Inferences from Wrongdoers’ Moral Violations

Short title: Early Inferences from Moral Violations

Fransisca Ting\textsuperscript{a,b,1}, and Renée Baillargeon\textsuperscript{a,1}

\textsuperscript{a}Department of Psychology
University of Illinois at Urbana-Champaign
603 East Daniel Street
Champaign, IL 61820
USA

\textsuperscript{b}Current affiliation:
Department of Psychological and Brain Sciences
Boston University
64 Cummington Mall
Boston, MA 02215
USA

\textsuperscript{1}To whom correspondence may be addressed: fting@bu.edu or rbaillar@illinois.edu.

Author contributions: F.T. and R.B. designed the research; F.T. performed the research and analyzed the data; and F.T. and R.B. wrote the paper.

The authors declare no conflict of interest.
Significance Statement (120 words)

According to prior research, toddlers have expectations related to several moral principles. We found that after a wrongdoer violated one principle when interacting with an individual, 25-month-old toddlers did not find it unexpected if the wrongdoer violated a different principle when interacting with other individuals. Toddlers did, however, find it unexpected when the wrongdoer next behaved generously toward another individual. These results make clear that a complete account of early moral cognition must specify not only (a) how moral principles guide early expectations about social interactions, but also (b) how violations of these expectations affect children’s evaluations of wrongdoers’ moral characters, and (c) how these evaluations, in turn, support broad inferences about wrongdoers’ likely behavior in new contexts.
Abstract

By two years of age, children already possess expectations related to several different moral principles. Building on these results, we asked whether children who observed a wrongdoer violate a principle would draw broad negative inferences from this violation about how the wrongdoer was likely to behave in other contexts. In four violation-of-expectation experiments, 25-month-old toddlers \((N = 152)\) first saw a wrongdoer harm a protagonist. When toddlers judged the wrongdoer’s behavior to violate the principle of ingroup support or harm avoidance, they did not find it unexpected if she next violated the principle of fairness by dividing resources unfairly between two other protagonists (Experiments 2 and 3), but they did find it unexpected if she next acted generously toward another protagonist by giving her most of a resource to be shared between them (Experiment 4). When toddlers did not construe the wrongdoer’s harmful behavior as a moral violation, these responses reversed: They found it unexpected if the wrongdoer next acted unfairly (Experiment 1), but not if she next acted generously (Experiment 4). Detecting a moral violation thus lowered toddlers’ assessment of the wrongdoer’s moral character and brought down their expectations concerning the likelihood that she would perform (a) obligatory actions required by other principles or (b) supererogatory or virtuous actions not required by the principles. Together, these findings expand our understanding of how young children evaluate others’ moral characters and reveal how these evaluations, in turn, enable children to form sophisticated expectations about others’ behavior in new contexts.

Keywords: infancy, moral violations, moral character, obligatory actions, supererogatory actions
Social scientists have long been interested in uncovering the basic structure of human moral cognition. An influential proposal has been that a small set of universal moral principles contributes to this structure (1-12). Although researchers disagree widely on the nature, number, and contents of these principles, common candidates include fairness, harm avoidance, ingroup support, and authority. General assumptions about these principles are that they evolved during the millions of years our ancestors lived in small groups of hunter-gatherers, where survival depended on cooperation; that they are normative and determine what actions are obligatory, forbidden, and permissible in specific contexts; and that they are implemented, stressed, and rank-ordered differently in different cultures, resulting in the diverse moral landscape that exists in the world today.

Over the past 10 years, developmental psychology has produced new evidence for this principle-based approach to moral cognition by showing that children under 2.5 years of age (henceforth young children) already possess abstract expectations related to each of the four candidate principles (13-26; for recent reviews, see 27,28). With respect to fairness, for example, 4–28-month-olds expected an individual to divide windfall resources equally between two similar potential recipients (13-15,18,22,23); 21-month-olds expected an individual to reward a worker who had done an assigned chore but not a slacker who had done no work (23); and 17-month-olds expected a resource acquired by two workers to be shared according to the amount of effort each had exerted (26). Similarly, with respect to ingroup support, 13–29-month-olds expected an individual to provide help to an ingroup member in need of instrumental assistance (16,25), but to refrain from helping an ingroup wrongdoer who had harmed an ingroup victim (21,25); during an episode of conflict between two protagonists from different groups, 9–18-month-olds found it unexpected if a member of one group chose to help the protagonist from the
other group, instead of the protagonist from its own group (20); and after two groups performed distinct novel conventional actions, 7–12-month-olds found it unexpected if a member of one group chose to imitate the other group’s conventional action (19).

Building on these and other findings of early moral expectations, the present research asked the following question: How closely linked are the different moral principles in the first years of life? Do young children construe the principles as separate, unconnected constraints on individuals’ actions, or do they view them, at least to some extent, as interconnected? To get at these questions, we first showed 25-month-old toddlers a scene in which a wrongdoer interacted with a protagonist and violated a principle. Next, the same wrongdoer interacted with new protagonists in a new scene that called for the application of a different principle. At issue was whether toddlers would expect the wrongdoer to adhere to this second principle. On the one hand, evidence that they did would suggest that they construed the different principles as independent, unconnected constraints on social interactions: After observing a wrongdoer violate one principle, they still expected the wrongdoer to adhere to another principle, as though the transgression they had witnessed had no bearing on their expectations concerning other principles. On the other hand, evidence that toddlers did not find it unexpected if the wrongdoer, having violated the first principle, now violated the second one as well, would suggest that they did view the different principles as interconnected and drew broad negative inferences from the initial transgression they witnessed.

What mechanism might support such inferences? One hypothesis was suggested by evidence that beginning at a young age, children not only attempt to predict how individuals will act toward others but also evaluate their actions, as though assessing their moral characters (29-34; for relevant research with older children and adults, see e.g., 35-40). For example, when
faced with two distributors, one who had divided resources fairly and one who had done so unfairly, 13–17-month-olds preferred the fair over the unfair distributor (29-31), 13–15-month-olds associated admonishment (e.g., “She’s a bad girl!”) with the unfair distributor (32), and 10–25-month-olds either gave or expected others to give a treat to the fair as opposed to the unfair distributor (33-34). These results suggested that upon observing a fairness violation, children evaluated the wrongdoer’s moral character unfavorably, and this evaluation drove their affiliative and punitive attitudes toward the wrongdoer: They were less likely to select the wrongdoer as a target for an affiliative action, they were more likely to select the wrongdoer as a target for a punitive action, and they expected others to do the same. We reasoned that if children were able to use their evaluation of a wrongdoer’s moral character to also consider how the wrongdoer might behave in a new situation, then observing a moral violation might lead them to draw broad negative inferences from this violation.

To flesh out this idea, we speculated that when encountering a novel individual, young children endow her with a good moral character by default (i.e., everyone is assumed to have a good character until proven otherwise; 41-43); they view it as highly likely that she will perform the obligatory actions (or avoid the forbidden actions) specified by the various moral principles; and they find it unexpected if she violates one of the principles, as noted earlier (13-26). In turn, observing such a violation causes children to lower their assessment of the individual’s character and brings down their estimate of the likelihood that she will perform other obligatory actions, from highly likely to only somewhat likely. In this scheme, all of the principles thus bear on the common evaluative dimension of good/bad character, making possible broad negative inferences from any one violation. Specifically, children should not find it unexpected if an individual who has violated one principle—and thus has revealed herself to be of questionable character—next
violates another principle.

To evaluate these speculations, we conducted a series of violation-of-expectation experiments in which 25-month-old toddlers first saw a wrongdoer harm a protagonist. Next, toddlers saw the wrongdoer divide resources either fairly or unfairly between two new protagonists. At issue was whether toddlers, when they perceived the harm committed by the wrongdoer as a moral violation, would look about equally whether she acted fairly or unfairly, suggesting that they did not find it unexpected when she chose to act unfairly. Such evidence would suggest that by two years of age, some degree of interconnection is already present among the different moral principles: Children realize that a wrongdoer who violates one principle may violate others as well.

The Present Research

According to a recent proposal (28), at least two principles, harm avoidance and ingroup support, contribute to young children’s judgments about whether a harmful action constitutes a moral violation. Harm avoidance applies to interactions among members of the same moral circle (e.g., humans) and sets broad limits on the amount of harm a member can inflict on other members. Ingroup support applies to interactions among members of the same social group within a moral circle (e.g., sports team) and sets much stricter limits on the amount of harm that can be inflicted on ingroup members. Recent developmental findings are consistent with this proposal. First, when an individual A harms another individual B, and there is nothing to suggest that A and B belong to the same moral circle (e.g., they are different non-verbal, non-human characters), young children give no evidence that they view A’s behavior as a moral violation, even if it involves substantial unprovoked harm; for example, 6–12-month-olds detected no violation when A repeatedly hit B or knocked B down a steep hill, causing it to roll end-over-end
to the bottom of the hill (44-47). Second, when there are unambiguous cues that A and B belong to the same social group, even *slight* unprovoked harm is perceived as a violation; for example, 13–29-month-olds detected a violation when A ignored B’s need for instrumental assistance or threw an object B needed on the floor (16,25,28). Finally, when there are cues that A and B belong to the same moral circle, but no cues that they belong to the same social group, results are intermediate between those described above, with substantial but not slight unprovoked harm being perceived as a violation (16,25,28). We built on these results in deciding what harmful actions to use in our research.

In Experiment 1, which served as a baseline, toddlers were first introduced to two groups of puppets marked by different kinds (dogs vs. rabbits), different labels spoken in female voices by the puppets (“I’m a dog!” vs. “I’m a rabbit!”), and different outfits (dresses vs. skirts). We expected that toddlers would view the dogs and the rabbits (a) as members of the same moral circle, based on their shared capacity for human speech (much like Bert, Ernie, Big Bird, and other verbal puppets on the television show Sesame Street, who are akin to “honorary humans”), and also (b) as members of two distinct social groups, based on the multiple cues provided (i.e., different kinds, labels, and outfits; 13,16,21,25). Next, toddlers saw a puppet from one group (the wrongdoer; e.g., a dog) direct a single, slightly harmful action at a puppet from the other group (the victim; a rabbit). More specifically, the wrongdoer knocked down a tower built by the victim, flipped over a puzzle filled by the victim, or crumpled a drawing made by the victim. Finally, toddlers saw the wrongdoer divide two items either equally (*fair* event) or unequally (*unfair* event) between two other puppets (recipients). Both recipients belonged to either the same group as the wrongdoer (two dogs; *ingroup-recipients* condition) or the other group (two rabbits; *outgroup-recipients* condition). Based on the findings reviewed above, toddlers should
perceive the wrongdoer’s slightly harmful action as *permissible*, because it did not violate the restrictions imposed by the harm-avoidance principle on unprovoked harm to outgroup members. Consequently, toddlers should still expect the wrongdoer to divide the two items equally between the two ingroup or outgroup recipients, in accordance with the fairness principle, and they should find it unexpected when she did not, as in prior research on early sensitivity to fairness (13-15,18,22-23). (Prior findings also suggest that infants who do not construe a wrongdoer’s harmful actions as a moral violation may still draw broad negative inferences about the wrongdoer’s moral character if these actions are repeatedly contrasted with the helpful, compassionate actions of an altruist (48); we return to these findings in the Discussion of Experiment 1).

Experiment 2 was identical except that the wrongdoer now harmed an ingroup puppet, instead of an outgroup puppet. Based on the findings reviewed above, toddlers should now perceive the wrongdoer’s slightly harmful action as *forbidden*, because it violated the strict limits imposed by the ingroup-support principle on unprovoked harm to ingroup members (16,25,28). We reasoned that if by two years of age children already construe the different moral principles as interconnected and recognize that a wrongdoer who fails to adhere to one principle may also fail to adhere to others, then toddlers should no longer find it unexpected when the wrongdoer divided the two items unequally between the two recipients, in violation of the fairness principle. Moreover, finding the same results whether the potential recipients were ingroup members (*ingroup-recipients* condition) or outgroup members (*outgroup-recipients* condition) not only would support the conclusion that toddlers drew broad negative inferences about the wrongdoer but also would rule out alternative interpretations positing narrower inferences (e.g., perhaps the wrongdoer had a conflictual relationship with her ingroup but otherwise treated others as morally
Experiment 3 was identical to the ingroup-recipients conditions of Experiments 1 and 2 with one exception: Instead of directing a single slightly harmful action at the victim, the wrongdoer now directed three such actions at the victim, in successive trials (tower, puzzle, and drawing). If toddlers judged that all three actions together amounted to substantial harm, then predictions for when the wrongdoer harmed an outgroup victim (outgroup-victim condition) should differ from those in Experiment 1. Because substantial harm to an outgroup member violated the harm-avoidance principle, toddlers might no longer find it unexpected if a wrongdoer who had committed such harm subsequently treated ingroup recipients unfairly. In contrast, predictions for when the wrongdoer harmed an ingroup victim (ingroup-victim condition) should be the same as in Experiment 2. If one slightly harmful action was sufficient to violate the ingroup-support principle, then three such actions should do so as well, with the same consequences for toddlers’ expectations: In either case, toddlers should not find it unexpected when the wrongdoer next treated ingroup recipients unfairly.

Finally, Experiment 4, which we introduce later, sought to address a possible alternative interpretation of the results of Experiments 2 and 3, and also began to examine additional expectations toddlers might form about individuals with a bad moral character (see Table S1 for a summary of all experiments and conditions).

**Experiment 1**

Experiment 1 served as a baseline and tested whether 25-month-olds who first saw a wrongdoer direct a slightly harmful action at an outgroup victim (an action toddlers should not view as a moral violation) would still expect the wrongdoer to act fairly when dividing resources between two ingroup recipients (ingroup-recipients condition) or two outgroup recipients
(outgroup-recipients condition). In each condition, English-speaking toddlers sat on a parent’s lap facing a large puppet-stage apparatus, and they received one familiarization trial, one harm trial, and two test trials (Fig. 1). In total, eight puppets were used in Experiments 1–3: There were four identical brown dogs who wore dresses in different colors and patterns (D1-D4), and four identical beige rabbits who wore skirts in different solid colors (R1-R4). Each toddler saw a subset of four puppets that included a wrongdoer (D1 or R1), a victim (D2 or R2) and two recipients (D3 and D4 or R3 and R4). In the familiarization trial, all four puppets were present and occupied windows around the three sides of the apparatus; the wrongdoer stood at the left window, the victim stood at the right window, and the two recipients stood at the ends of a wide window in the back wall. All of the windows were filled with curtains, to hide the assistants who operated the puppets. Whether D1 or R1 served as the wrongdoer was counterbalanced within each experiment; for ease of communication, we use D1 in our descriptions.

In the ingroup-recipients condition of Experiment 1, the familiarization trial involved wrongdoer D1, outgroup victim R2, and ingroup recipients D3 and D4. The trial was computer-controlled, followed a second-by-second script, and lasted 12 s. The four puppets labeled themselves in turn, one group at a time, starting from the left or the right (counterbalanced; e.g., D1: “I’m a dog!”; D3: “I’m a dog, too!”; D4: “I’m a dog, too!”; R2, “I’m a rabbit!”). Each puppet tilted from side to side as she spoke, to make clear who was speaking. The trial ended after the last puppet spoke.

The harm trial was computer-controlled, lasted 40 s, and depicted a slightly harmful action in which D1 destroyed a tower, puzzle, or drawing (counterbalanced) produced by R2. In the tower event, R2 was initially alone and stood next to a tray filled with five discs of different colors and sizes; the discs were designed to be stacked by increasing sizes, and the largest disc
had a toy bear attached to its top. After R2 stacked the first two discs, one at a time, D1 entered, and the two puppets greeted each other (“Hi!”). Next, while D1 watched, R2 stacked the remaining discs, returned to her window, and clapped twice while admiring her tower (“Yay!”). D1 then approached the tower and hit it, causing it to fall over. She then returned to her window, and the trial ended. The puzzle event was similar, with the following exceptions: R2 had a tray filled with six puzzle pieces, each with a round knob, and she inserted them one by one into a wooden frame; after she completed her puzzle and admired it (“Yay!”), D1 grabbed its top edge, pulled it toward herself, and flipped it over, causing all of the pieces to fall out (the top edge of the puzzle rested on small supports, making it easy for D1 to grab it). Finally, the drawing event was also similar to the preceding events, with the following exceptions: R2 had a marker and a coloring page (the page rested on a clipboard for easy coloring); after R2 completed her drawing and admired it (“Yay!”), D1 crumpled it and tossed it back on the clipboard.

Each test trial had an initial phase and a final phase. The initial phase was computer-controlled, lasted 26 s, and depicted the fair or unfair event, ending with a paused scene; during the final phase, which was infant-controlled, toddlers watched this scene until the trial ended (see Methods). At the start of the initial phase, only D3 and D4 were present and tilted from side to side in unison. D3 stood at the left end of the back window, as before, and D4 now stood next to her; in front of each puppet was a placemat. Next, D1 entered, carrying a tray with two toys (yellow blocks or purple balls, counterbalanced). She set down the tray and announced, “I have toys!” and D3 and D4 responded enthusiastically, “Yay, yay!” in different voices. D1 placed one toy on one of the placemats (counterbalanced), and then she placed the other toy either on the same placemat (unfair event) or on the other placemat (fair event). Next, D1 picked up her empty tray and left the scene, and D3 and D4 looked down at their placemats and paused until
the trial ended. The order of the two test events was counterbalanced.

The *outgroup-recipients* condition was identical except that the two puppets in the back window in the familiarization and test trials were R3 and R4, instead of D3 and D4.

We reasoned that if toddlers (a) viewed the four puppets as members of the same moral circle, based on their shared capacity for human speech, (b) assigned the dog puppets and the rabbit puppets to two distinct social groups, based on the multiple cues provided, and (c) viewed the wrongdoer’s single slightly harmful action as permissible (i.e., it did not violate the principle of harm avoidance, and the principle of ingroup support did not apply as the wrongdoer’s harmful action was directed at an outgroup victim), then toddlers should still expect the wrongdoer to adhere to fairness when distributing the toys. Toddlers in both conditions should thus look significantly longer at the unfair than at the fair event, as in prior research on early sensitivity to fairness.

Toddlers (*N* = 32) were highly attentive during the familiarization and harm trials and the initial phases of the test trials (which were all computer-controlled); across conditions, they looked, on average, for 99% of the trials. Looking times during the final phases of the test trials (Fig. 2) were analyzed using an ANOVA with Condition (ingroup- or outgroup-recipients) as a between-subjects factor and Event (unfair or fair) as a within-subject factor. The main effect of Event was significant, *F*(1, 30) = 21.55, *p* < .001, η² = .42, but the other effects were not, both *F*(1, 30) ≤ 0.02, *p* ≥ .904, η² = .00. Planned comparisons revealed that (a) toddlers in the ingroup-recipients condition looked significantly longer at the unfair (*M* = 24.58, *SD* = 15.20) than at the fair (*M* = 14.16, *SD* = 7.95) event, *F*(1, 30) = 10.23, *p* = .003, Cohen’s *d* = 0.86, with 12/16 toddlers showing this pattern, and (b) toddlers in the outgroup-recipients condition also looked significantly longer at the unfair (*M* = 24.48, *SD* = 11.34) than at the fair (*M* = 13.51, *SD* = 7.34).
= 7.06) event, $F(1, 30) = 11.34, p = .002, d = 1.16$, with 14/16 toddlers showing this pattern. Nonparametric Wilcoxon signed-rank tests confirmed the results of both conditions (with $Z = 2.56, p = .010$, in each condition).

After observing a wrongdoer direct a slightly harmful action at an outgroup victim (i.e., destroy the victim’s tower, puzzle, or drawing), toddlers still expected the wrongdoer to divide windfall resources fairly between two ingroup or outgroup recipients. These results suggested that toddlers did not perceive the wrongdoer’s harmful action as a moral violation, and hence that they drew no particular inferences from this action about her moral character. As such, these results provided an important baseline: They confirmed, with the present stimuli, events, and procedure, that young children tend to view a slightly harmful action directed at an outgroup victim as permissible (16,25,28), and that they expect a distributor (who to their knowledge has performed no moral violation) to treat two similar recipients fairly, whether they belong to the distributor’s group or not (13-15,18,22-23).

Nevertheless, the main finding of Experiment 1—that toddlers still expected a wrongdoer to treat two recipients fairly after she had inflicted permissible harm on a victim—might, at first sight, appear inconsistent with prior findings from infant experiments conducted with the help/hinder paradigm (44,46,48,49-53). In these experiments, 3–21-month-olds watched events involving three different non-verbal, non-human protagonists, A, B, and C; no cues signaled that they belonged to the same social group or the same moral circle. The events included a help and a hinder event, which were repeated across trials. In the help event, A gave C the instrumental assistance it needed to achieve its goal (e.g., helped C reach the top of a steep hill it could not reach on its own); in the hinder event, B interfered with C’s goal (e.g., knocked C down to the bottom of the hill). Across ages, infants looked equally at the two events, indicating that they did
not construe the hinder event as a moral violation (had they done so, they would have looked significantly longer at that event than at the help event). In the absence of cues signaling that B and C belonged to the same social group or the same moral circle, no harm restrictions applied, so infants detected no violation in the hinder event. Nevertheless, infants provided robust evidence that they evaluated A’s action favorably and B’s action unfavorably and assessed their moral characters accordingly. First, these evaluations influenced infants’ affiliative attitudes: When asked to choose between A and B, 3–11-month-olds preferred A (44, 49-51), and 10–12-month-olds expected C to do the same (44,46,52). Second, these evaluations influenced infants’ punitive attitudes: 21-month-olds chose B when asked to take away a treat, but they chose A when asked to give a treat (53). Third, and most relevant for present purposes, these evaluations led infants to draw broad inferences about A’s and B’s moral characters. In an experiment by Surian and colleagues with 15-month-olds (48), the help and hinder events were followed by a single test event in which A or B divided two items either equally (fair event) or unequally (unfair event) between two potential recipients. When A was the distributor, infants looked significantly longer if shown the unfair as opposed to the fair event; when B was the distributor, however, infants looked equally at the two events, as though they did not find this unfair distribution unexpected, given B’s bad character.

How can we reconcile these last results (infants did not find it unexpected when a wrongdoer who had committed permissible harm acted unfairly) with the results of Experiment 1 (toddlers found it unexpected when a wrongdoer who had committed permissible harm acted unfairly)? Salient differences between the two experiments included the participants’ age (15 vs. 25 months) and the number and nature of the wrongdoer’s harmful actions (the hinderer bumped the climber down the hill twice per trial for a total of four bumps; the wrongdoer destroyed one
item produced by the victim). However, we suspect that the key difference between the two experiments has to do with the contrast between the hinderer’s harmful actions and the helper’s compassionate, altruistic actions. Across trials, as in previous experiments using the help/hinder paradigm (44,46,49,50,52), the hinderer consistently chose to hinder, whereas the helper consistently chose to help, even though this action was not obligatory (there were no cues to indicate that the helper and the climber belonged to the same group). This repeated contrast may have induced infants to compare the helper’s and hinderer’s actions and to draw inferences about their respective moral characters, even though they detected no moral violation in the hinderer’s actions. In the present research, however, no such contrast was available: Toddlers saw only a wrongdoer who directed one or more harmful actions at an ingroup or an outgroup member. Our research examined whether toddlers would draw broad negative inferences about the wrongdoer’s moral character when they construed these harmful actions as a moral violation, but would draw no such inferences otherwise. The results of Experiment 1 supported this last prediction (toddlers still expected the wrongdoer to act fairly following a harmful action they deemed permissible), setting the scene for the following experiments.

**Experiment 2**

Experiment 2 was identical to Experiment 1 with one exception: R2 was replaced with D2, so that D1’s slightly harmful action was now directed at an ingroup as opposed to an outgroup victim (Fig. 1). We reasoned that if (a) toddlers now perceived D1’s action as an ingroup-support violation, as in prior research (16,25,28), and (b) this violation lowered their assessment of D1’s moral character and their estimate of the likelihood that she would perform other obligatory actions, then toddlers should no longer find it unexpected when D1 acted unfairly in the test trials, and they should therefore look equally at the unfair and fair events.
Moreover, obtaining this negative result in the outgroup-recipients condition as well as in the ingroup-recipients condition would rule out alternative interpretations of the latter condition (e.g., perhaps D1 had a conflictual relationship with her ingroup members and ignored moral principles when interacting with them, but still treated others as morally expected). Finding that following D1’s slightly harmful action toward D2, toddlers no longer expected D1 to act fairly when dealing with either ingroup members D3 and D4 or outgroup members R3 and R3 would suggest that they drew broad negative inferences about D1’s moral character upon observing her initial violation.

Toddlers (N = 32) were highly attentive during the familiarization and harm trials and the initial phases of the test trials, looking, on average, for 99% of the trials. Looking times during the final phases of the test trials (Fig. 2) were analyzed as before. The main effect of Event was not significant, $F(1, 30) = 0.50, p = .484, \eta_p^2 = 0.02$, nor were the other effects, both $Fs(1, 30) \leq 2.12, p \geq .156, \eta_p^2 \leq 0.07$. Planned comparisons revealed that (a) toddlers in the ingroup-recipients condition looked equally at the unfair ($M = 17.49, SD = 7.79$) and fair ($M = 16.66, SD = 6.55$) events, $F(1, 30) = .06, p = .810, d = 0.12$, with 8/16 toddlers looking longer at the unfair event; and (b) toddlers in the outgroup-recipients condition also looked equally at the unfair ($M = 22.26, SD = 10.91$) and fair ($M = 19.65, SD = 13.90$) events, $F(1, 30) = 0.58, p = .453, d = 0.21$, with 9/16 toddlers looking longer at the unfair event. Wilcoxon signed-rank tests confirmed the results of the ingroup-recipients ($Z = 0.31, p = .756$) and outgroup-recipients ($Z = 0.52, p = .605$) conditions.

In another ANOVA, we compared the responses of toddlers in Experiment 1, who saw an outgroup-victim/1-harm scenario, to those of toddlers in Experiment 2, who saw an ingroup-victim/1-harm scenario. The analysis yielded a significant main effect of Event, $F(1, 60) = 13.77$,
$p < .001, \eta^2_p = 0.19$, and a significant Victim $\times$ Event interaction, $F(1, 60) = 7.20, p = .009, \eta^2_p = 0.11$, confirming that toddlers drew different inferences from D1’s slightly harmful action depending on whether it was directed at an outgroup victim (Experiment 1) or an ingroup victim (Experiment 2).

Experiment 1 showed that after observing a wrongdoer direct a slightly harmful action at an outgroup victim, toddlers still expected the wrongdoer to divide windfall resources fairly between two ingroup or outgroup recipients. In contrast, Experiment 2 showed that when this same action was directed at an ingroup victim, thereby violating the ingroup-support principle, toddlers no longer found it unexpected if the wrongdoer next treated ingroup or outgroup recipients unfairly. These negative results support a characterization of early moral cognition that posits some degree of interconnection among the different moral principles: Toddlers assumed that a wrongdoer who violated the ingroup-support principle might also violate the fairness principle, suggesting that her initial violation led them to draw broad negative inferences about her moral character and about the likelihood that she would perform obligatory actions dictated by other principles.

**Experiment 3**

In Experiments 1 and 2, the wrongdoer (D1) directed a single slightly harmful action at a victim; in Experiment 3, the wrongdoer committed three such actions, in successive harm trials (i.e., destroyed the victim’s tower, puzzle, and drawing, in that order). For half of the toddlers, these actions were directed at an outgroup victim (R2), as in Experiment 1 (outgroup-victim condition); for the other toddlers, these actions were directed at an ingroup victim (D2), as in Experiment 2 (ingroup-victim condition). In both conditions, the three harm trials were followed by the same test trials as in the ingroup-recipients conditions of Experiments 1 and 2, with the
wrongdoer dividing two toys between two ingroup members (D3 and D4).

We reasoned that in the ingroup-victim condition, results should be the same as in Experiment 2: If a single slightly harmful action directed at an ingroup victim was sufficient for toddlers to detect an ingroup-support violation and, consequently, to no longer expect the wrongdoer to act fairly, then three such actions should, a fortiori, have the same effect. In the outgroup-victim condition, however, results might differ from those in Experiment 1. If toddlers took all three harmful actions to amount to substantial harm, then they should judge these actions to violate the harm-avoidance principle (recall that this principle sets limits on the amount of harm that can be inflicted on other members of a moral circle). Furthermore, if, as in Experiment 2, this violation lowered toddlers’ evaluation of the wrongdoer’s moral character, thereby bringing down their estimate of the likelihood that she would perform other obligatory actions, then toddlers should no longer find it unexpected when the wrongdoer acted unfairly in the test trials. Toddlers in both the ingroup- and outgroup-victim conditions were thus expected to look equally at the fair and unfair events.

The procedure of Experiment 3 was identical to that in the preceding experiments except that toddlers now received three harm trials. As before, toddlers (N = 32) were highly attentive during the familiarization and harm trials and the initial phases of the test trials, looking, on average, for 99% of the trials. Looking times during the final phases of the test trials (Fig. 2) were analyzed using an ANOVA with Condition (outgroup- or ingroup-victim) as a between-subjects factor and Event (unfair or fair) as a within-subject factor. The main effect of Event was not significant, $F(1, 30) = 0.26, p = .613, \eta^2_p = 0.01$, nor were the other effects, both $F$s($1, 30) \leq 1.49, p \geq .231, \eta^2_p \leq 0.05$. Planned comparisons confirmed that (a) toddlers in the outgroup-victim condition looked equally at the unfair ($M = 18.27, SD = 9.88$) and fair ($M = 21.04, SD =$
11.24) events, \( F(1, 30) = 0.55, \ p = .466, \ d = -0.26 \), with 6/16 toddlers looking longer at the unfair event; and (b) toddlers in the ingroup-victim scenario similarly looked equally at the unfair (\( M = 16.73, \ SD = 9.09 \)) and fair (\( M = 16.66, \ SD = 10.28 \)) events, \( F(1, 30) = 0.00, \ p = .987, \ d = 0.01 \), with 9/16 toddlers looking longer at the unfair event. Wilcoxon signed-rank tests confirmed the results of the outgroup-victim (\( Z = 0.63, \ p = .532 \)) and ingroup-victim (\( Z = 0.41, \ p = .679 \)) conditions.

In additional ANOVAs, we compared test responses across Experiments 1–3, focusing on harm to outgroup and ingroup victims separately. For conditions involving harm to an outgroup victim, we first compared the outgroup-victim condition of Experiment 3 (three harmful actions) to the ingroup-recipients condition of Experiment 1 (one harmful action; recall that both of these conditions involved ingroup recipients). This ANOVA (\( N = 32 \)) used Harm (3 or 1) as a between-subjects factor and Event (fair or unfair) as a within-subject factor. The Harm \( \times \) Event interaction was significant, \( F(1, 30) = 7.43, \ p = .011, \ \eta^2_p = 0.20 \), and the same result was found when the outgroup-recipients condition in Experiment 1 was substituted for the ingroup-recipients condition, \( F(1, 30) = 7.24, \ p = .012, \ \eta^2_p = 0.19 \). Turning to conditions involving harm to an ingroup victim, we first compared the ingroup-victim condition of Experiment 3 (three harmful actions) to the ingroup-recipients condition of Experiment 2 (one harmful action). The Harm \( \times \) Event interaction was not significant, \( F(1, 30) = 0.03, \ p = .873, \ \eta^2_p = 0.00 \), and the same result was found when the outgroup-recipients condition in Experiment 2 was substituted for the ingroup-recipients condition, \( F(1, 30) = 0.23, \ p = .638, \ \eta^2_p = 0.01 \).

Together, the data of Experiments 1–3 supported three conclusions. First, when the wrongdoer directed one or more slightly harmful actions at an ingroup victim, toddlers judged the wrongdoer’s behavior to violate the ingroup-support principle, and they did not find it
unexpected if she next violated the fairness principle when dealing with other protagonists. Second, when the wrongdoer directed three slightly harmful actions at an outgroup victim, toddlers judged the wrongdoer’s behavior to violate the harm-avoidance principle, and they again did not find it unexpected if she next acted unfairly. Finally, when the wrongdoer directed a single slightly harmful action at an outgroup victim, toddlers did not perceive this action as a moral violation, and they found it unexpected if the wrongdoer next acted unfairly. These results suggest that by 2 years of age, toddlers already draw broad negative inferences from the moral violations they observe, suggesting that some degree of interconnection is already present among the different principles.

**Experiment 4**

We have suggested that toddlers in Experiments 2 and 3 looked equally at the fair and unfair events they were shown because (a) they perceived the wrongdoer’s harmful actions as a violation of the ingroup-support or harm-avoidance principle; (b) this violation lowered their assessment of her moral character; and (c) this unfavorable assessment, in turn, led them to not find it unexpected when she chose to violate the fairness principle. However, our results were open to another, subtly different interpretation: It could be that upon detecting the wrongdoer’s moral violation in the harm trial(s), toddlers simply concluded that her actions were unpredictable, causing them to refrain from forming any further expectations about her behavior. Experiment 4 sought to rule out this alternative interpretation.

To do so, we needed a situation in which toddlers would find it unexpected if a wrongdoer who first violated a moral principle next performed a particular action. We turned to supererogatory actions, which are virtuous actions that go beyond what is required by the moral principles and that signal excellences of character such as compassion, generosity, and bravery.
Previous research suggests that whereas young children view individuals of good character as highly likely to perform obligatory actions, as was discussed previously, they view them as only somewhat likely to perform supererogatory actions (16, 24, 25, 28). For example (16), 17-month-old infants found it unexpected when an unfamiliar individual chose to ignore an ingroup member in need of instrumental assistance, suggesting that they judged helping the ingroup member to be an obligatory action (dictated by the ingroup-support principle) and had assumed the individual was highly likely to perform that action. In contrast, infants did not find it unexpected if the individual chose to ignore an outgroup member with the same need, suggesting that they judged helping the outgroup member to be a supererogatory, compassionate action and had assumed the individual was only somewhat likely to perform that action. We speculated that if observing a wrongdoer violate a moral principle brought down toddlers’ expectations not only about the likelihood that the wrongdoer would perform other obligatory actions (from highly likely to somewhat likely), as shown in Experiments 2 and 3, but also about the likelihood that the wrongdoer would perform supererogatory actions (from somewhat likely to very unlikely), then toddlers should find it unexpected if the wrongdoer next performed a supererogatory action.

In line with these speculations, Experiment 4 asked whether toddlers who first saw a wrongdoer direct harmful actions at an ingroup member would find it unexpected if the wrongdoer next gave another ingroup member most of a resource to be shared between them (a supererogatory, generous action). Evidence that toddlers found the wrongdoer’s generosity unexpected would rule out the suggestion that children in Experiments 2 and 3 merely viewed the wrongdoer as unpredictable and refrained from forming expectations about her behavior.

Toddlers in the ingroup-victim/3-harms condition first received the same three harm trials as in the ingroup-victim condition of Experiment 3. Next, they received a single test trial in
which they saw either a generous-wrongdoer or a generous-ingroup event. Each event had a computer-controlled initial phase that lasted 22 s, followed by an infant-controlled final phase. At the start of the initial phase in the generous-wrongdoer event, only wrongdoer D1 and ingroup member D4 were present, in the same places as in the familiarization trial. They tilted left and right until they were joined by D3, who entered the apparatus (in her usual place, between D1 and D4) carrying a tray with six identical toys (red blocks). D3 set the tray down on the apparatus floor, announced, “I have toys!” and D1 and D4 both answered excitedly, “Yay, yay!” Next, D3 let go of the tray, opened her arms, and said, “Go ahead!”, as though inviting D1 and D4 to share the toys. She then exited the apparatus. As soon as she left, D1 quickly lunged over the tray, picked up the toy in the farthest corner, placed it in front of herself, and then pushed the remaining five toys in front of D4. The two puppets then looked down and paused, and toddlers watched this scene until the trial ended. The generous-ingroup event was identical except that D4 lunged over the tray, took the toy in the farthest corner, and gave the remaining five toys to D1.

Before we can outline our predictions, three comments are in order. First, because D3 and D4 were not present in the three harm trials when D1 destroyed D2’s tower, puzzle, and drawing, they should be uninformed or ignorant about D1’s harmful actions; toddlers should therefore not expect D3 and D4 to adopt a punitive attitude toward D1 (23,56). Second, while young children generally expect a distributor to divide a resource equally between two similar recipients (13-15,18,22,23), they recognize that an individual who is sharing a resource with an ingroup member may elect to give away more than an equal share (24,26); toddlers should thus view generosity to an ingroup member as morally permissible. Third, Experiment 4 used a between-subjects design, instead of a within-subject design, to avoid interpretive issues that might arise
across test trials. We were concerned that after toddlers saw one puppet give away five of the six toys in the first trial, seeing the other puppet do the same in the second trial would be open to multiple interpretations (e.g., she might want to imitate her ingroup member, she might want to achieve an equal distribution of the toys across trials, or she might want to reciprocate by also acting generously). To avoid these ambiguities, each toddler received a single test trial. Moving on to our predictions, we reasoned that if toddlers viewed ingroup member D4 as somewhat likely to act generously, but wrongdoer D1 as very unlikely to do so, then they should look significantly longer if shown the generous-wrongdoer as opposed to the generous-ingroup event.

Toddlers (N = 32) were highly attentive during the familiarization and harm trials and the initial phase of the test trial, looking, on average, for 98% of the trials. Looking times in the final phase of the test trial were subjected to an ANOVA with Event (generous-wrongdoer or generous-ingroup) as a between-subjects factor. This effect was significant, $F(1, 30) = 9.85$, $p = .004$, $d = 1.13$, indicating that toddlers looked reliably longer if shown the generous-wrongdoer event ($M = 23.16$, $SD = 10.89$) as opposed to the generous-ingroup event ($M = 12.89$, $SD = 6.89$) event. A non-parametric Wilcoxon rank-sum test confirmed this result ($Z = 2.85$, $p = .004$).

After they saw the wrongdoer harm an ingroup victim three times, toddlers found it unexpected if she next acted generously toward another ingroup member. To provide evidence that this response arose only when the wrongdoer’s generosity came on the heels of a moral violation, additional toddlers were tested in a final, identical condition except that the wrongdoer now harmed an outgroup victim once, as in Experiment 1 (outgroup-victim/1-harm condition). If toddlers again viewed this single harmful action as permissible and drew no broad negative inferences from it, then they should not find it unexpected if the wrongdoer next chose to act generously. Toddlers ($N = 24$) were highly attentive during the familiarization and harm trials
and the initial phase of the test trial, looking, on average, for 99% of the trials. Looking times during the final phase of the test trial were analyzed as above. The main effect of Event was not significant, \( F(1, 22) = 0.11, p = .741, d = -0.14 \), indicating that toddlers looked about equally whether they were shown the generous-wrongdoer event (\( M = 21.87, SD = 9.85 \)) or the generous-ingroup event (\( M = 23.27, SD = 10.65 \)). A Wilcoxon rank-sum test confirmed this result (\( Z = 0.46, p = .644 \)). Finally, to compare the two conditions of Experiment 4, we conducted an ANOVA similar to that above but with Condition (ingroup-victim/3-harms or outgroup-victim/1-harm) as an additional between-subjects factor. The Condition \( \times \) Event interaction was significant, \( F(1, 52) = 4.97, p = .030, \eta_p^2 = 0.09 \), indicating that toddlers found the wrongdoer’s generosity unexpected when it was preceded by harm they construed as a moral violation (ingroup-victim/3-harms condition), but not otherwise (outgroup-victim/1-harm condition).

The results of Experiment 4 thus complemented those of Experiments 1–3 and supported their interpretations. When toddlers judged a wrongdoer’s harmful behavior to be a moral violation, this caused them to lower their assessment of her moral character and their expectations concerning the likelihood that she would perform other obligatory actions (from highly likely to somewhat likely) or supererogatory actions (from somewhat likely to very unlikely). As a result, toddlers did not find it unexpected if she next acted unfairly (Experiments 2 and 3), but they did find it unexpected if she next acted generously (Experiment 4). However, when toddlers did not view the wrongdoer’s harmful behavior as a moral violation, they apparently drew no negative inferences from her behavior and showed the reverse responses: They found it unexpected if she next acted unfairly (Experiment 1), but not if she next acted generously (Experiment 4).
General Discussion

Research over the past decade has revealed that young children already hold expectations related to various moral principles, including ingroup support, harm avoidance, and fairness (13-16,18-23,25-26,29-34,48). Building on these results, the present research asked whether children who observe a wrongdoer violate one of these principles draw broad negative inferences that affect their expectations about the wrongdoer’s likely behavior in new situations.

Experiment 1 served as a baseline and showed that after a wrongdoer directed a single slightly harmful action at an outgroup victim (an action young children typically do not perceive as a moral violation), 25-month-old toddlers still expected the wrongdoer to act fairly when dividing windfall resources between two ingroup or outgroup recipients. In Experiment 2, the same harmful action was directed at an ingroup victim, violating the principle of ingroup support, and toddlers no longer found it unexpected if the wrongdoer next acted unfairly toward ingroup or outgroup recipients. In Experiment 3, three harmful actions were shown. When these were directed at an ingroup victim, toddlers again detected an ingroup-support violation and, consistent with the results of Experiment 2, did not find it unexpected if the wrongdoer next treated ingroup recipients unfairly. When these harmful actions were directed at an outgroup victim, toddlers judged these actions to amount to substantial harm, thereby violating the harm-avoidance principle, and they no longer found it unexpected if the wrongdoer next treated ingroup recipients unfairly. Experiments 2 and 3 thus provided converging evidence that toddlers did not find it unexpected if a wrongdoer who violated one principle next violated another. Complementing these results, Experiment 4 showed that following three harmful actions to an ingroup victim, toddlers again detected an ingroup-support violation and found it unexpected if the wrongdoer next generously shared a resource with another ingroup member, by giving her
most of the resource. Finally, in line with the results of Experiment 1, toddlers did not find the wrongdoer’s generosity unexpected if she first directed a single slightly harmful action at an outgroup victim. Together, the results of Experiments 1–4 thus indicated that after observing a wrongdoer violate the ingroup-support or harm-avoidance principle when interacting with one protagonist, toddlers did not find it unexpected if she next violated the fairness principle when interacting with other protagonists, but they did find it unexpected if she next behaved generously toward another protagonist.

Our results provide new evidence for several prior findings in early moral cognition. In particular, they confirm that (a) toddlers use available cues to assign unfamiliar individuals to social groups (13,16,19-21,25,29); (b) all other things being equal, toddlers expect individuals to divide windfall resources fairly between ingroup or outgroup recipients (in accordance with fairness; 13-15,18,22,23), to refrain from harming ingroup members (in accordance with ingroup support; 16,25,28), and to comply with restrictions on harm to outgroup members (in accordance with harm avoidance; 28); and (c) toddlers do not expect individuals who are unaware of harm inflicted by a wrongdoer (because they were absent when it occurred) to adopt a punitive attitude toward the wrongdoer (25,56).

In addition, our results break new ground in two important ways: by showing that by 2 years of age, young children already draw broad negative inferences from the moral violations they observe, and by suggesting a mechanism that explains the basis of these inferences. All other things being equal, when young children encounter a new individual, they assume that she possesses a good character, that she is highly likely to perform obligatory actions that are dictated by the moral principles, but that she is only somewhat likely to perform supererogatory actions that go beyond what is required by the principles and reveal excellences of character such
as compassion, generosity, or bravery. If the individual then fails to perform an obligatory action, this violation causes children to lower their evaluation of her moral character, which in turn brings down their expectations concerning the likelihood that she will perform other obligatory actions (now viewed as only somewhat likely) as well as supererogatory actions (now viewed as very unlikely). Our results thus make clear that a complete account of early moral cognition must include not only the different moral principles that help children predict individuals’ actions, but also a basic understanding of moral character that includes concepts of obligatory and supererogatory actions and considers the likelihood of these actions for good vs. bad individuals.

Future research can build on our results in several directions. First, it will be important to confirm our results using a variety of scenarios that weave together different obligatory and supererogatory actions. For example, recall that young children look significantly longer if an individual ignores as opposed to helps an ingroup member in need of assistance, but look about equally if the individual ignores or helps an outgroup member with the same need (16). The present research suggests that if children first saw the individual divide resources unfairly between two similar recipients, both looking patterns would reverse. Children should now look about equally whether the individual helped, ignored, or even hindered an ingroup member in need, and they should look significantly longer if the individual helped as opposed to ignored or hindered an outgroup member in need. In other words, following a wrongdoer’s fairness violation, children should view an obligatory action dictated by the ingroup-support principle as only somewhat likely, and they should view a supererogatory, compassionate action toward an outgroup member as very unlikely. Exploring these issues with children in the first and second years of life should also help shed light on the infant roots of character evaluation.

Second, future research can explore whether observing a wrongdoer commit more or
more varied moral violations would lead children to lower their expectations concerning the likelihood that the wrongdoer would perform other obligatory actions, from very likely to very unlikely (as opposed to somewhat likely). In Experiment 3, for example, after toddlers saw the wrongdoer violate the principle of harm avoidance by directing three harmful actions at an outgroup victim, they tended to look equally whether she acted fairly or unfairly toward ingroup members. What if the wrongdoer first committed a more severe violation, by directing more harmful actions at the victim or by harming a larger number of victims (35)? Would children now expect her to act unfairly in the test trials, and hence would they look significantly longer if she acted fairly instead? Another approach would be to show children a wrongdoer who violated two different principles. For example, what if a wrongdoer failed to distribute resources fairly between two outgroup members and then harmed another outgroup member by destroying her tower, puzzle, and drawing? Would children infer that a wrongdoer who violated two principles (fairness and harm avoidance) was very likely to also violate a third (e.g., ingroup support or authority)? For instance, would children look significantly longer if the wrongdoer chose to help, as opposed to ignore or hinder, an ingroup member in need of assistance?

Third, future studies may explore whether toddlers restrict the inferences they make about a wrongdoer to only the moral domain, or would toddlers make very broad inferences beyond the domain of morality. For example, prior results suggest that infants expect an agent to behave rationally: that is, for an agent to act consistently with their mental states (e.g., beliefs, stable preferences, knowledge, goals) and for the agent to expend the least amount of resources to achieve their goal (xx). Thus, infants expect an agent to reach for their preferred toy over a non-preferred toy, and to take the shortest path instead of a longer one to reach their target object (xx). Would toddlers still expect a wrongdoer who violated a moral principle to follow the
principle of rationality? Would toddlers no longer find it unexpected if a wrongdoer chose a non-preferred toy over a preferred toy, or if they took a tortuous path over a short one to achieve their goal? Evidence for inferences made beyond the moral domain might suggest that toddlers’ evaluation of an individual’s moral character is central to how toddlers view the core personhood of this individual (xx).

Finally, another important direction for future research will be to explore what information might lead children to reverse their negative evaluations of wrongdoers, and (absent such information) how stable these negative evaluations tend to be over time. For example, would genuine signs of contrition by the wrongdoer, perhaps accompanied by explicit forgiveness from the victim, tend to lead children to revert to basic expectations about the wrongdoer’s likely behavior in new contexts (57)? Furthermore, would the simple passage of time tend to have a beneficial effect on children’s negative evaluations? Prior research suggests that older children are likely to remember nice or virtuous characters even after a week’s delay (58). What if a similar delay was imposed between the initial trials depicting a wrongdoer’s moral violation and the test trials assessing her willingness to perform other obligatory actions or supererogatory actions? Would children tend to revert back to baseline expectations over time, as though no violation had occurred?

Conclusion

The present research showed that by 2 years of age, children draw broad negative inferences from the moral violations they observe. After seeing a wrongdoer violate the principle of ingroup support or harm avoidance when interacting with a protagonist, toddlers did not find if unexpected if the wrongdoer next acted unfairly toward other protagonists, but they did find it unexpected if she next acted generously toward another protagonist. These findings expand our
understanding of how young children evaluate others’ moral characters and demonstrate how these evaluations, in turn, enable children to form sophisticated and nuanced expectations about others’ behavior in new contexts.

Methods

**Power Analyses.** Experiments 1–3 each had a 2 (Conditions) × 2 (Events) within-subject design, with Event as a repeated factor. To estimate the appropriate sample size of each experiment, we relied on a prior report on early sensitivity to fairness that also used a within-subject design (23). A 2 (Conditions) × 2 (Events) ANOVA, with a sample of 32, yielded a Condition × Event effect size ($\eta_p^2$) of 0.14. A G*Power analysis (59) based on this value, with alpha set at .05 and power set at .80, suggested that the minimum total sample size per experiment was at least 22. We tested 32 participants in each experiment, with 16 per condition, as in this prior report. Experiment 4 had a 2 (Conditions) × 2 (Events) between-subjects design, and here we relied on a prior report on early sensitivity to ingroup support that also used a between-subjects design (16). A series of 2 (Conditions) × 2 (Events) ANOVAs, each with a sample of 32, yielded an average Condition × Event effect size ($\eta_p^2$) of 0.19. A G*Power analysis based on this value suggested that the minimum total sample size in Experiment 4 was at least 36. We tested 56 participants, 32 in the ingroup-victim/3-harms condition and 24 in the final, outgroup-victim/1-harm condition (we had planned to again test 32 but data collection was shut down by the COVID-19 pandemic).

**Participants.** Participants were 152 English-speaking toddlers (75 male; M = 25 months, 5 days, range = 21;18–29,10). Another 9 toddlers were excluded, 5 because they were fussy (4) or distracted (1), 2 because they experienced parental interference, and 2 because their test looking time was over 3 SDs from the condition mean (both were in the ingroup-victim/3-harms
condition of Experiment 4 and saw the generous-ingroup event). In each condition of Experiments 1–3, about half of the toddlers saw the unfair event first, and the others saw the fair event first. In the ingroup-victim/3-harms condition of Experiment 4, 17 toddlers saw the generous-wrongdoer event and 15 saw the generous-ingroup event; in the outgroup-victim/1-harm condition, 12 toddlers saw each event. In all four experiments, each toddler’s parent gave written informed consent, and the protocol was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign.

**Apparatus.** The apparatus consisted of a brightly lit display booth (201 cm high × 101 cm wide × 58 cm deep) with a large opening (56 × 95 cm) in its front wall; between trials, a supervisor lowered a curtain in front of this opening. Inside the apparatus, the floor and side walls were white and the back wall was blue. The wrongdoer and the victim were introduced into the apparatus through side windows (each 51 × 38 cm and filled with a white fringe curtain), and the other protagonists were introduced through a back window (18 × 61 cm and filled with stretchy blue fabric divided by a horizontal slit). Four dog puppets and four rabbit puppets were used across experiments. The dogs (each about 24 × 24 × 18 cm at the largest points) were brown with a white circular patch around the left eye, and they wore different patterned dresses; the wrongdoer and the victim wore a blue dress with scattered dots and a cream dress with overlapping dots (counterbalanced), and the other two protagonists wore a red dress with small flowers (left dog) and a white dress with small lines (right dog). The rabbits (each about 24 × 20 × 18 cm) were beige with a white lower face and stomach, and they wore solid color skirts and matching hair clips; the wrongdoer and the victim wore purple and green skirts (counterbalanced), and the other two protagonists wore yellow (left rabbit) and pink (right rabbit) skirts. The puppets were operated by three female assistants; one (in a white shirt) knelt at the
left window and operated the wrongdoer, one (also in a white shirt) knelt at the right window and operated the victim, and one (in a blue shirt) sat at the back window and operated the other two protagonists. Behind the assistants, floor-to-ceiling white (side assistants) or blue (back assistant) curtains hid the testing room from view. In the familiarization trial, when the two protagonists at the back spoke in turn, the back assistant used a high voice for one protagonist and a normal voice for the other; in the test trials of Experiments 1–3, when the two protagonists cheered together, the back assistant and the supervisor spoke in unison.

Stimuli in the harm trials included a tower and a clear rectangular tray with five colorful discs; a wooden puzzle depicting a street scene and a white rectangular tray with six puzzle pieces; and a green marker and a beige rectangular clipboard with a coloring page depicting a flower (the clipboard was affixed to the apparatus floor by Velcro strips, to make it easier for the wrongdoer to take the drawing). Stimuli in the test trials of Experiments 1–3 included two rectangular placemats with a granite-pattern adhesive paper, a beige rectangular tray, and either two yellow blocks or two purple balls (with flat bottoms to prevent rolling); the blocks were used with the dog recipients and the balls with the rabbit recipients, to stand out better against their clothes. Stimuli in Experiment 4 included six small red blocks and an edgeless rectangular tray that was covered with a wood-pattern adhesive paper and had a thin vertical handle at the back. For better control as events unfolded, the tray had a thin metal plate under its adhesive paper, the blocks had magnets that kept them in place as the tray was moved, and the tray was deposited on Velcro strips on the apparatus floor. The blocks were placed on the tray in two staggered rows of three, to make it easy for toddlers to see how many there were.

During each testing session, two cameras captured images of the toddler and events; the two images were combined, projected onto a monitor located behind the apparatus, and checked
by the supervisor to confirm that the events followed the prescribed scripts. Recorded sessions were also checked off-line for experimenter accuracy.

**Procedure.** Toddlers sat on a parent’s lap in front of the apparatus; parents were instructed to close their eyes in the test trial(s) and to remain silent and neutral in all trials. During a test trial, the toddler’s looking behavior was monitored by two hidden observers who were naïve about which event was shown in the trial; looking times were computed using the primary observer’s responses. During familiarization and harm trials, the primary observer was absent from the testing room and was thus also naïve about which puppets were present and how many harmful actions the wrongdoer inflicted on the victim.

The final phase of a test trial ended when toddlers either (a) looked away for 1 consecutive second after having looked for at least 5 cumulative seconds or (b) looked for a maximum of 45 cumulative seconds. The 5-s minimum-look gave toddlers more time to process the events shown in the initial phase before the final phase could end; when actions stop and a paused scene begins, toddlers sometimes look away briefly, and a minimum-look allows them to return to the scene and continue processing it. Consistent with this analysis, across all 248 test trials in the present research, toddlers took 5.22 s (SD = 0.78), on average, to complete the 5-s minimum-look. Finally, inter-observer agreement during the final phase of a test trial was calculated by dividing the number of 100-ms intervals in which the two observers agreed by the total number of intervals in the final phase. Agreement in the two test trials of Experiments 1-3 was calculated for 91/96 toddlers (only one observer was present for the other toddlers) and averaged 96% per trial; agreement in the test trial of Experiment 4 was calculated for 55/56 toddlers and averaged 95% per trial.

**Data and Preliminary Analyses.** The data from all four experiments are available in
Dataset S1 in the Supplementary Online Material, along with preliminary analyses of the test data in each experiment.

Acknowledgments

This research was supported by a grant from the John Templeton Foundation to R.B. We thank the UIUC Infant Cognition Lab for their help with the data collection; graphic artist Steve Holland for producing the figures; and the families who participated in the experiments.

References


15. Enright EA, Gweon H, Sommerville JA (2017). ‘To the victor the spoils’: Infants expect resources to align with dominance structures. *Cognition* 164:8-21


**Figures**
Fig. 1. Familiarization, harm, and test trials in the ingroup-recipient conditions of Experiment 1 (left panel) and Experiment 2 (right panel); half the children saw a dog wrongdoer (shown here), and half saw a rabbit wrongdoer. Toddlers in the outgroup-recipients condition (not shown) received identical test trials except that the potential recipients were outgroup members. In
Experiment 3, toddlers in the outgroup-victim and ingroup-victim conditions were tested using the same procedure as in the ingroup-recipients conditions of Experiments 1 and 2, respectively, except that they received all three harm trials, in the order depicted.
Experiment 1: Outgroup−Victim / 1−Harm

Experiment 2: Ingroup−Victim / 1−Harm

Experiment 3: 3−Harms

Experiment 4
Fig. 2. Mean test looking times ($N = 152$), separately by condition and event. In Experiments 1–3, which used a within-subject design, connected dots represent individual toddlers, with 16 per condition. Red diamonds indicate means, boxes represent the interquartile ranges between the first and third quartiles, and horizontal lines inside boxes indicate medians. In Experiment 4, which used a between-subjects design, each dot represents an individual toddler; dots are jittered horizontally to improve visibility. In the ingroup-victim/3-harms condition, 17 toddlers saw the generous-wrongdoer event and 15 saw the generous-ingroup event; in the outgroup-victim/1-harm condition, 12 toddlers saw each event.
Fig. 3. Test trial in Experiment 4. Toddlers in the ingroup-victim/3-harms condition first received the same familiarization and harm trials as in the ingroup-victim condition of Experiment 3; toddlers in the outgroup-victim/1-harm condition first received the same familiarization and harm trials as in Experiment 1.
Supplementary Information

A. Summary of Conditions in Experiments 1-4

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>N</th>
<th>Victim</th>
<th>Harm(s)</th>
<th>Test Event(s)</th>
<th>Recipient(s)</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ingroup-recipients</td>
<td>16</td>
<td>outgroup</td>
<td>1</td>
<td>fair and unfair</td>
<td>ingroup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>outgroup-recipients</td>
<td>16</td>
<td>outgroup</td>
<td>1</td>
<td>fair and unfair</td>
<td>outgroup</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ingroup-recipients</td>
<td>16</td>
<td>ingroup</td>
<td>1</td>
<td>fair and unfair</td>
<td>ingroup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>outgroup-recipients</td>
<td>16</td>
<td>ingroup</td>
<td>1</td>
<td>fair and unfair</td>
<td>outgroup</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>outgroup-victim</td>
<td>16</td>
<td>outgroup</td>
<td>3</td>
<td>fair and unfair</td>
<td>ingroup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ingroup-victim</td>
<td>16</td>
<td>ingroup</td>
<td>3</td>
<td>fair and unfair</td>
<td>ingroup</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ingroup-victim/3-harms</td>
<td>32</td>
<td>ingroup</td>
<td>3</td>
<td>generous-wrongdoer or generous-ingroup</td>
<td>ingroup</td>
<td></td>
</tr>
<tr>
<td></td>
<td>outgroup-victim/1-harm</td>
<td>24</td>
<td>outgroup</td>
<td>1</td>
<td>generous-wrongdoer or generous-ingroup</td>
<td>ingroup</td>
<td></td>
</tr>
</tbody>
</table>

Table S1. For each condition of each experiment, the table lists the sample size, affiliation of the victim in the harm event(s), number of harm events shown, test event(s) shown, and affiliation of the recipient(s) in the test event(s); it also provides an example of the four characters used in the condition (with a dog wrongdoer).

B. Preliminary analyses in Experiments 1-4

Experiments 1-3. In each experiment, preliminary analyses of the test data examined whether there were significant interactions of condition and event with the following counterbalancing factors: order in which the test events were presented, participants’ sex, harm event (in Experiments 1-2), wrongdoer kind, side of the first label in the familiarization trial, and side of the first recipient in the test trials. As can be seen in Table S2, these interactions were non-significant in all three experiments, with one exception: There was a significant interaction of condition and event with side of first recipient in Experiment 1, $F(1, 28) = 9.84, p = .004, \eta_p^2 = 0.26$. Follow-up ANOVAS indicated that in the outgroup-recipients condition, toddlers responded similarly to the fair and unfair test events whether the left or the right puppet was the first recipient (Event x Side of first recipient interaction: $F(1,14) = 1.20, p = .292, \eta_p^2 = 0.08$). In the ingroup-recipients condition, however, this interaction was significant, $F(1, 14) = 17.49, p < .001, \eta_p^2 = .56$. Follow-up contrasts indicated that when the right recipient received the first toy, toddlers looked significantly longer at the unfair event (M = 33.14, SD = 12.61) than at the fair event (M = 15.91, SD = 9.06), $F(1, 14) = 41.97, p < .001, \eta_p^2 = 0.75$. However, when the left recipient received the first toy, toddlers looked equally at the unfair (M = 10.32, SD = 4.10) and fair (M = 11.25, SD = 5.06) events, $F(1, 14) = 0.07, p = .790, \eta_p^2 = 0.01$. This last result is most likely due to sampling variability as (a) only six toddlers in this condition saw the left recipient receive the first toy, so that this effect was based on a very small number of participants, (b) no effect of side of first recipient was found in Experiments 2 and 3, and (c) no effect of side of first
recipient has been reported in previous studies on early sensitivity to fairness using distributions to two similar recipients (1-2). For all three experiments, the data were therefore collapsed across these counterbalancing factors in subsequent analyses.

**Experiment 4.** As can be seen in Table S2, preliminary analyses of the test data in the ingroup-victim/3-harms and outgroup-victim/1-harm conditions revealed no significant interaction of event with participant’s sex, harm event (outgroup-victim/1-harm condition only), wrongdoer kind, and side of the first label in the familiarization trial. The data were therefore collapsed across these counterbalancing factors in subsequent analyses.
<table>
<thead>
<tr>
<th>Exp 1</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond x Event x Order (unfair/fair or fair/unfair)</td>
<td>1.52</td>
<td>0.228</td>
<td>0.05</td>
</tr>
<tr>
<td>Participant's sex (M or F)</td>
<td>0.05</td>
<td>0.833</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Harm event (tower, puzzle, or drawing)</td>
<td>0.96</td>
<td>0.397</td>
<td>0.07</td>
</tr>
<tr>
<td>Wrongdoer kind (dog or rabbit)</td>
<td>0.02</td>
<td>0.896</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Side of first label in familiarization (L or R)</td>
<td>0.39</td>
<td>0.536</td>
<td>0.01</td>
</tr>
<tr>
<td>Side of first recipient in test (L or R)</td>
<td>9.84</td>
<td>0.004</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exp 2</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond x Event x Order (unfair/fair or fair/unfair)</td>
<td>0.04</td>
<td>0.841</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Participant's sex (M or F)</td>
<td>1.91</td>
<td>0.178</td>
<td>0.06</td>
</tr>
<tr>
<td>Harm event (tower, puzzle or drawing)</td>
<td>2.25</td>
<td>0.126</td>
<td>0.15</td>
</tr>
<tr>
<td>Wrongdoer kind (dog or rabbit)</td>
<td>0.27</td>
<td>0.607</td>
<td>0.01</td>
</tr>
<tr>
<td>Side of first label in familiarization (L or R)</td>
<td>0.08</td>
<td>0.786</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Side of first recipient in test (L or R)</td>
<td>0.34</td>
<td>0.566</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exp 3</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond x Event x Order (unfair/fair or fair/unfair)</td>
<td>0.50</td>
<td>0.486</td>
<td>0.02</td>
</tr>
<tr>
<td>Participant's sex (M or F)</td>
<td>1.72</td>
<td>0.200</td>
<td>0.06</td>
</tr>
<tr>
<td>Wrongdoer kind (dog or rabbit)</td>
<td>0.64</td>
<td>0.430</td>
<td>0.02</td>
</tr>
<tr>
<td>Side of first label in familiarization (L or R)</td>
<td>3.69</td>
<td>0.065</td>
<td>0.12</td>
</tr>
<tr>
<td>Side of first recipient in test (L or R)</td>
<td>1.11</td>
<td>0.302</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exp 4, Ingroup-victim/3-harms condition</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event x Participant's sex (M or F)</td>
<td>2.56</td>
<td>0.121</td>
<td>0.08</td>
</tr>
<tr>
<td>Wrongdoer kind (dog or rabbit)</td>
<td>2.03</td>
<td>0.165</td>
<td>0.07</td>
</tr>
<tr>
<td>Side of first label in familiarization (L or R)</td>
<td>1.00</td>
<td>0.325</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exp 4, Outgroup-victim/1-harm condition</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event x Participant's sex (M or F)</td>
<td>0.35</td>
<td>0.561</td>
<td>0.02</td>
</tr>
<tr>
<td>Harm event (tower, puzzle, or drawing)</td>
<td>1.39</td>
<td>0.276</td>
<td>0.13</td>
</tr>
<tr>
<td>Wrongdoer kind (dog or rabbit)</td>
<td>0.12</td>
<td>0.730</td>
<td>0.01</td>
</tr>
<tr>
<td>Side of first label in familiarization (L or R)</td>
<td>0.28</td>
<td>0.603</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Table S2.** Summary of results from the preliminary analyses of Experiments 1-4.
C. References
