Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor

Fransisca Ting[a,1], Zijing He[b,1], and Renée Baillargeon[a,1]

[a]Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820; and [b]Department of Psychology, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

Contributed by Renée Baillargeon, January 14, 2019 (sent for review October 17, 2018; reviewed by Yarrow Dunham and Katherine McAuliffe)

Adults and older children are more likely to punish a wrongdoer for a moral transgression when the victim belongs to their group. Building on these results, in violation-of-expectation experiments (n = 198), we examined whether 2.5-year-old toddlers (Exps. 1 and 2) and 1-year-old infants (Exps. 3 and 4) would selectively expect an individual in a minimal group to engage in third-party punishment (TPP) for harm to an ingroup victim. We focused on an indirect form of TPP, the withholding of help. To start, children saw a wrongdoer steal a toy from a victim while a bystander watched. Next, the wrongdoer needed assistance with a task, and the bystander either helped or hindered her. The group memberships of the wrongdoer and the victim were varied relative to that of the bystander and were marked with either novel labels (Exps. 1 and 2) or novel outfits (Exps. 3 and 4). When the victim belonged to the same group as the bystander, children expected TPP: At both ages, they detected a violation when the bystander chose to help the wrongdoer. Across experiments, this effect held whether the wrongdoer belonged to the same group as the bystander and the victim or to a different group; it was eliminated when the victim belonged to a different group than the bystander, when groups were not marked, and when either no theft occurred or the wrongdoer was unaware of the theft. Toddlers and infants thus expect individuals to refrain from helping an ingroup victim's aggressor, providing further evidence for an early-emerging expectation of ingroup support.

infancy | moral cognition | third-party punishment | ingroup support | helping

Social scientists have long argued that humans' success as a species is due in large part to their remarkable capacity for cooperation (1–3). Many different mechanisms are thought to support and sustain human cooperation, including the following. First, embedded in the basic structure of moral cognition are abstract norms of fairness, harm avoidance, ingroup support, and authority; although implemented and prioritized in different ways in different cultures, these norms still help regulate interactions within and between groups (4–10). Second, adults enforce these norms by retaliating against wrongdoers who fail to treat them as expected (11–14). Third, adults also enforce these norms by punishing wrongdoers who fail to treat others as expected (15–18); this is referred to as "third-party punishment" (TPP) and can take several forms (19–22). In direct, costly TPP, which is typically studied in the laboratory using economic games, adults willingly incur costs to engage in TPP: For example, when told they can sacrifice some of their own resources to punish a wrongdoer who has treated a victim unfairly, they often choose to do so (16–18). Outside the laboratory, direct TPP appears to be less common (23–26) and to be left largely in the hands of authority figures and other leaders (27–30). Instead, field studies often point to a less direct and less costly form of TPP: Instead of confronting wrongdoers head-on, adults avoid them, gossip about them, shun them, and refrain from helping them or from sharing resources with them (24, 31–33). In the present research, we focused on this indirect form of TPP and examined its developmental roots. In four violation-of-expectation experiments, we asked whether 2.5-y-old toddlers and 1-y-old infants would expect a bystander who had observed a wrongdoer commit a mild transgression against a victim to later refrain from helping the wrongdoer. Evidence that children detected a violation when the bystander chose to help the wrongdoer would suggest that they expected at least indirect TPP against the wrongdoer.

A second goal of our research was to explore whether considerations of group membership would modulate children's expectations about indirect TPP, assuming they had any. In adults, TPP becomes more likely when the victim of the transgression is an ingroup member (34–38). In a seminal report, Bernhard et al. (34) concluded, "We found that punishers protect ingroup victims—who suffer from a norm violation—much more than they do outgroup victims, regardless of the norm violator's group affiliation." Building on these results, we used minimal-group manipulations (6, 39, 40) to vary the group memberships of the wrongdoer and the victim relative to that of the bystander, and we examined whether children (*i*) would expect the bystander to engage in TPP when the wrongdoer's transgression was directed at an ingroup victim, irrespective of the wrongdoer's group membership, but (*ii*) would expect no TPP when the transgression was directed at an outgroup victim.

We reasoned that such results would be important for several reasons. First, they would indicate that expectations about indirect TPP are already present early in life. Second, evidence that these expectations are, from a young age, selectively limited to ingroup victims, would suggest that they are driven largely by considerations of ingroup support. Choosing not to help a wrongdoer who has harmed an ingroup victim ultimately contributes to the welfare of the victim, the punisher, and the group

Significance

Adults are more likely to punish transgressions that do not affect them when these transgressions victimize ingroup members. Such third-party punishment (TPP) often takes an indirect form, such as the withholding of help. Building on these results, we showed 2.5- and 1-year-olds scenarios involving a wrongdoer, a victim, and a bystander, and we manipulated the minimal-group memberships of the wrongdoer and the victim relative to that of the bystander. When the victim belonged to the bystander's group, children expected TPP: They detected a violation when the bystander chose to help the wrongdoer. When the victim did not belong to the bystander's group, however, children no longer expected TPP. Young children thus selectively expect indirect TPP for harm to ingroup members.

Author contributions: F.T., Z.H., and R.B. designed research; F.T. and Z.H. performed research; F.T. analyzed data; and F.T. and R.B. wrote the paper with revisions from Z.H.

Reviewers: Y.D., Yale University; and K.M., Boston College.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: fting2@illinois.edu, hezij2@mail.sysu.edu.cn, or rbaillar@illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817849116/-/DCSupplemental.

Published online March 11, 2019.

as a whole, by making clear that transgressions against any member of the group will have adverse consequences. Finally, evidence that following harm to ingroup victims, some degree of TPP is expected for outgroup as well as ingroup wrongdoers, would also point to considerations of ingroup support. As Bernhard et al. (34) argued, "punishing outsiders who harm an ingroup victim increases the general security of all ingroup members," because outsiders are then "deterred from such attacks and all ingroup members enjoy more protection." We return to these issues in *General Discussion*.

Prior Developmental Findings

How likely were the toddlers and infants in our experiments to hold expectations about TPP? And how likely were these expectations to be modulated by considerations of group membership? Two sets of findings with children age 3 y and younger were particularly germane to our research and predictions.

The first set suggested that some understanding of TPP is present in early childhood. Focusing first on negative or corrective actions, 3-y-olds who observed a harm or fairness transgression spontaneously engaged in protestations, indignant tattling, and interventions aimed at rectifying the transgression (41–44), and they also judged that a child who had committed a harm transgression deserved to be punished (45). When asked to take a treat away from either a character who had helped a protagonist or a character who had hindered or harmed the protagonist, 21-mo-olds selected the hinderer (46). Even 6-mo-olds found it unexpected (as indexed by longer looking times) when a bystander chose to hit a victim as opposed to a wrongdoer, suggesting that they viewed the wrongdoer as the appropriate target for the bystander's actions (47). Turning to positive or affiliative actions, 25- to 37-mo-olds were less likely to hand over a desired object to a hinderer as opposed to a helper, or to an unfair as opposed to a fair distributor (48–50); 21-mo-olds were less likely to give a treat to a hinderer as opposed to a helper (46); 12-mo-olds were less likely to accept treats from a hinderer as opposed to a helper, even when the hinderer offered two and the helper offered only one (51); 10- to 29-mo-olds were more likely to prefer a man who had acted positively toward a child and negatively toward an inanimate object than a man who had done the converse (52); 10- to 12-mo-olds were more likely to prefer a helper over a hinderer and expected others to show the same preference (53–55); 10-mo-olds found it unexpected when a bystander chose to give a treat to an unfair as opposed to a fair distributor (56); and 6-mo-olds found it unexpected when a bystander chose to help a wrongdoer as opposed to a victim (47). Together, these results suggested that from a young age, children possess some understanding of TPP: They are more likely to select wrongdoers as targets for negative or corrective actions, they are less likely to select wrongdoers as targets for positive or affiliative actions, and they expect others to do the same.

Although highly informative, this first set of findings still left unanswered the question of whether or under what conditions young children expect TPP. The evidence that children protest against wrongdoers, or view them as appropriate targets for punitive actions, does not mean that they generally expect individuals to deploy TPP against wrongdoers and would detect a violation if someone chose not to engage in TPP. Thus, to get at the question of whether children hold expectations about TPP, in the present research we asked if, under some conditions at least, children would view TPP as expected rather than as optional and hence would detect a violation when no TPP occurred.

The second set of findings suggested that an abstract expectation of ingroup support is present early in life (4–10, 57). Thus, 3-y-olds who heard stories about two minimal groups (marked by labels and outfits) predicted that a harmful action (e.g., stealing a block) would be more likely to be directed at an outgroup as opposed to an ingroup victim (58); 19- to 28-mo-olds expected a distributor from one group of animal puppets (e.g., monkeys) to reserve scarce resources for ingroup recipients (other monkeys) as opposed to outgroup recipients (e.g., giraffes) (59); 17-mo-olds expected an individual to provide help to another individual in need of instrumental assistance when the two belonged to the same minimal group (marked by labels), but they held no expectations about the provision of help when the two individuals belonged to different minimal groups or when their group memberships were unspecified (60); after watching nonhuman adult characters soothe baby characters, 16-mo-olds found it unexpected if one baby preferred a baby who had been soothed by a different adult (and hence presumably belonged to a different group) over a baby who had been soothed by the same adult (and hence presumably belonged to the same group) (61); after watching two groups of nonhuman characters perform distinct novel conventional actions, 7- to 12-mo-olds detected a violation if a member of one group chose to imitate the other group's conventional action (62); and 4- to 12-mo-olds expected a woman alone with a crying baby (who presumably belonged to the same group as the woman) to attempt to comfort the baby, and they found it unexpected when she ignored the baby instead (63). Together, these results suggested that an abstract expectation of ingroup support is present early in life and that mere categorization of individuals into the same minimal group is sufficient to trigger rich expectations of ingroup care and loyalty (6, 57, 58, 60).

The two sets of findings reviewed above gave credence to the possibility that considerations of ingroup support might modulate expectations about TPP from a young age. To date, the earliest evidence for such modulation comes from a report with 6-y-olds (64). Children were first assigned to one of two minimal groups (marked by labels and outfits), and then they played a child version of the TPP game (65). Children were told that a distributor had allocated six candies either fairly (3, 3) or unfairly (6, 0) between him- or herself and a recipient, and they were asked to serve as "judges" who could either allow or punish the distributor's allocation; punishment was costly in that children had to sacrifice one of their own candies to reject an allocation. Across trials, the recipient belonged to either the children's group or the other group. Like adults (34–38), children were significantly more likely to punish unfair allocations that disadvantaged ingroup recipients, and this effect held for both ingroup and outgroup distributors. In another experiment (66), 3-y-olds saw either an ingroup or an outgroup puppet harm an ingroup experimenter (e.g., destroy her drawing). Children were equally likely to protest this transgression when perpetrated by the ingroup or the outgroup puppet. Because the group membership of the victim was not varied, however, it remains unclear whether similar or different results would have been obtained with an outgroup victim.

Our research built on these various findings to examine whether 2.5-y-old toddlers (Exps. 1 and 2) and 1-y-old infants (Exps. 3 and 4) would expect an individual in a minimal group to engage in indirect TPP for a mild harm transgression against an ingroup member, but not an outgroup member.

Experiments with Toddlers

Toddlers watched live scenes in which three unfamiliar women sat at windows around three sides of a puppet-stage apparatus and played the roles of wrongdoer (right window), victim (left window), and bystander (back window). The women's group memberships were established using novel labels. In one scene, the wrongdoer stole a toy from the victim, while the bystander watched. In a later scene, the victim was absent, and the wrongdoer worked at a puzzle; the final piece was out of her reach, but within the bystander's reach (67, 68). The wrongdoer tried unsuccessfully to reach the final piece and then left. Next, the bystander picked up the piece and either brought it closer to the wrongdoer's window so that she could complete her puzzle when she returned ("help" event) or threw it away ("hinder" event).

Across experiments and conditions, we varied the group memberships of the wrongdoer and the victim, relative to that of the bystander, to explore two main questions. First, would children expect the bystander to engage in indirect TPP when the victim was an ingroup member, regardless of the wrongdoer's group membership? Second, would children expect the bystander not to engage in TPP when the victim was an outgroup member, again regardless of the wrongdoer's group membership? Positive answers to both questions would suggest that, at least for mild transgressions, early expectations about TPP are driven by a sense of ingroup support (i.e., transgressions against ingroup members are selectively expected to be punished), rather than by a broad sense of retributive justice (i.e., transgressions against all victims are universally expected to be punished).

Experiment 1. Exp. 1 examined whether 2.5-y-old toddlers would expect a bystander to engage in indirect TPP after seeing a wrongdoer harm a victim (*i*) when the victim belonged to the same group as the bystander and the wrongdoer belonged to a different group, but not (*ii*) when these affiliations were reversed.

Children were randomly assigned to one of three conditions (Fig. 1*A*) ($n = 18$ in all conditions). They sat on a parent's lap facing the apparatus and received two familiarization trials and one test trial. In the first condition, the victim belonged to the same group as the bystander, but the wrongdoer belonged to a different group (ingroup victim–outgroup wrongdoer or IV-OW condition) (Fig. 2). At the start of the first familiarization trial, the women labeled themselves in two identical rounds using the novel group labels "topids" and "jaybos"; each round started from left or right, counterbalanced across toddlers (e.g., from left: "I'm a topid!", "I'm a topid, too!", "I'm a jaybo!"). Next, while the bystander and the wrongdoer watched, the victim picked up a cup, put a block in it, closed it with a lid, and then shook it several times, causing it to rattle. The trial then ended. The second familiarization trial was identical except that the cup, block, and lid were different colors and the block was positioned across the apparatus from the victim, out of her reach (but within the wrongdoer's reach). After watching the victim reach unsuccessfully for the block, the wrongdoer stole it and left, closing the curtain at her window. The victim looked back and forth several times between her open cup and the wrongdoer's closed window, and then the trial ended.

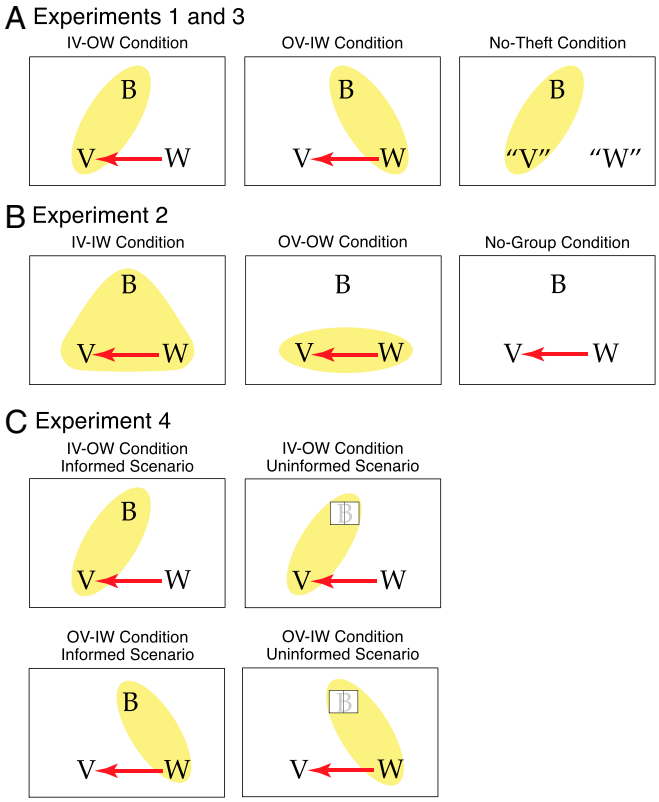

Fig. 1. Conditions in Exps. 1 and 3 (*A*), Exp. 2 (*B*), and Exp. 4 (*C*). The letters represent the bystander (B), victim (V), and wrongdoer (W); the arrows depict transgressions; the yellow overlays indicate members of the same group; and the conditions' names specify V's and W's group memberships relative to B. In *C*, the white box in each uninformed scenario signals that B was absent during the transgression (her window was closed).

In the test trial, toddlers saw either the help or the hinder event. Each event had an initial phase and a final phase, and the victim was absent in both phases (her window remained closed). During the initial phase of the help event, while the bystander watched, the wrongdoer selected puzzle pieces from a tray next to her window, one by one, and inserted them into a puzzle frame. The final piece rested across the apparatus from the wrongdoer, out of her reach (but within the bystander's reach). The wrongdoer tried unsuccessfully to reach the final piece and then left, closing her window. Next, the bystander picked up the final piece and placed it in the wrongdoer's tray so that she could complete her puzzle when she returned (60). The bystander then looked down and paused. During the final phase of the event, toddlers watched this paused scene until the trial ended. The hinder event was identical except that after picking up the final piece, the bystander dropped it out of the apparatus and then paused. Only the bystander was present in the final phase of the test trial so that children could focus on her and on the actions she had performed.

The second condition was identical except that the wrongdoer now belonged to the same group as the bystander, whereas the victim belonged to the other group (OV-IW condition). Finally, the third condition was again identical to the first except that each familiarization trial ended after the labeling: The "victim" no longer built a rattle and the "wrongdoer" no longer stole from her (no-theft condition). This control condition was included to confirm that test responses in the IV-OW condition did not simply reflect expected actions toward outgroup individuals, regardless of whether they engaged in wrongdoing.

Our predictions were as follows. First, if young children expect indirect TPP for transgressions against ingroup members, then toddlers in the IV-OW condition should look significantly longer if shown the help as opposed to the hinder event. Second, a different looking pattern was predicted in the OV-IW condition, based on prior findings that young children (*i*) view helping as expected between individuals from the same group, but as optional between individuals from different or unmarked groups, and (*ii*) view mild harm as unexpected when directed at individuals from the same group, but as permissible when directed at individuals from different or unmarked groups (46, 53–55, 58, 60, 69, 70). In line with these findings, we predicted that toddlers in the OV-IW condition would expect the bystander (*i*) to provide the ingroup wrongdoer with the help she needed to attain her goal and (*ii*) to pay little or no heed to the wrongdoer's mild harm transgression against the outgroup victim. Toddlers should therefore look significantly longer if shown the hinder as opposed to the help event. Finally, in line with the same findings reviewed above, we predicted that toddlers in the no-theft condition would view helping or hindering the outgroup "wrongdoer" as equally permissible, and would therefore look about equally at the two events. Three distinct looking patterns—longer looking at the help event, longer looking at the hinder event, and equal looking at the two events—were thus predicted across the three conditions.

Looking times during the final phase of the test trial (Fig. 3) were subjected to an ANOVA with condition (IV-OW, OV-IW, or no-theft) and event (help or hinder) as between-subject factors. The analysis yielded only a significant Condition × Event interaction,

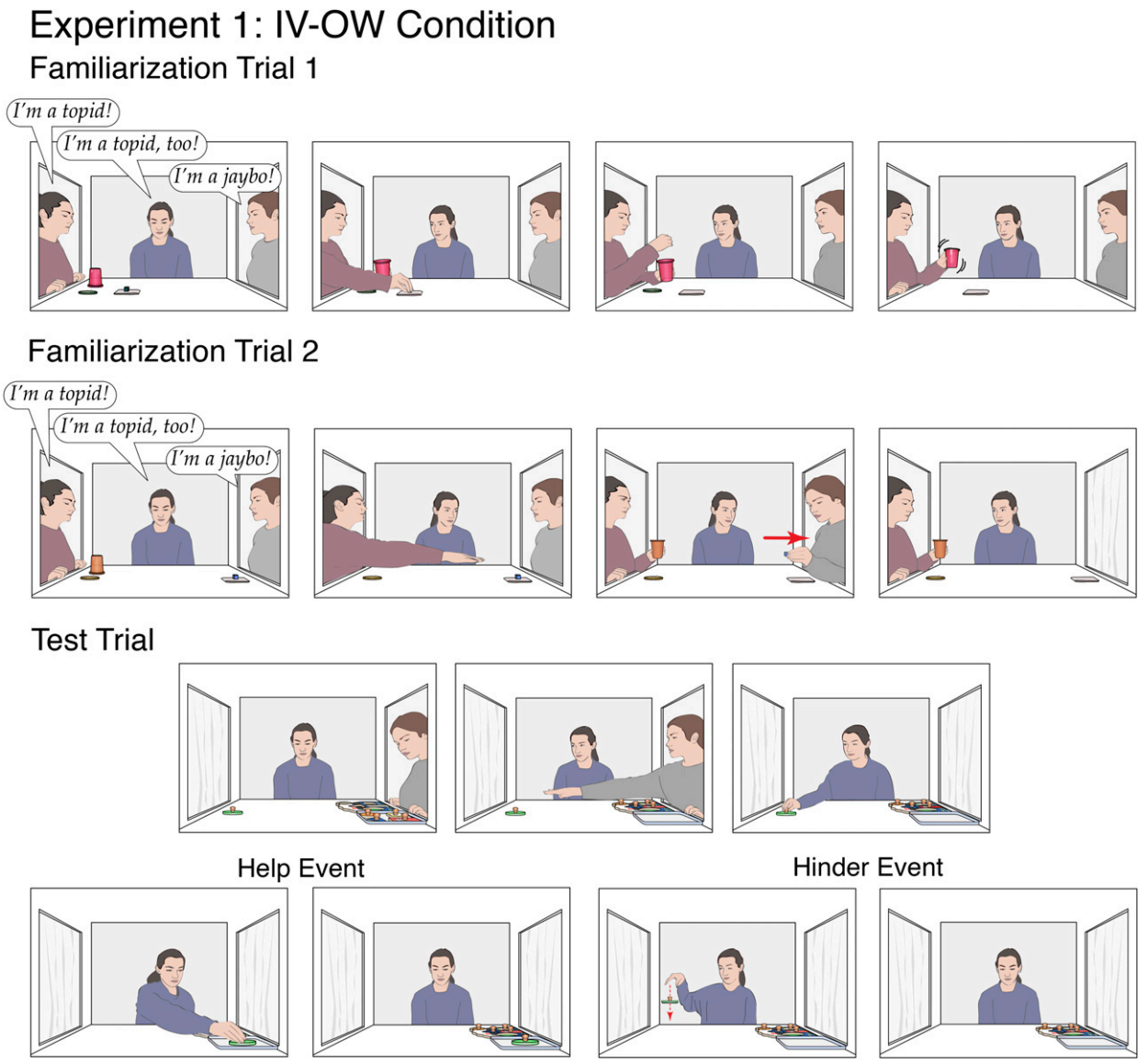

Fig. 2. Familiarization and test trials in the IV-OW condition of Exp. 1.

$F(2, 48) = 5.97$, $P = 0.005$, $\eta_p^2 = 0.199$. Planned comparisons revealed that, as predicted, toddlers in the IV-OW condition looked significantly longer if shown the help event [mean (M) = 26.43, SD = 12.32] as opposed to the hinder event (M = 14.08, SD = 3.07), $F(1, 48) = 6.48$, $P = 0.014$, Cohen's $d = 1.38$; toddlers in the OV-IW condition looked significantly longer if shown the hinder event (M = 24.40, SD = 10.59) as opposed to the help event (M = 13.17, SD = 2.27), $F(1, 48) = 5.35$, $P = 0.025$, $d = 1.47$); and toddlers in the no-theft condition looked about equally at the help (M = 21.68, SD = 10.30) and hinder (M = 23.34, SD = 15.87) events, $F(1, 48) = 0.12$, $P > 0.250$, $d = -0.12$. Nonparametric Wilcoxon rank-sum tests confirmed the results of the IV-OW ($z = 2.78$, $P = 0.005$), OV-IW ($z = -2.83$, $P = 0.005$), and no-theft ($z = 0.27$, $P > 0.250$) conditions.

When the victim belonged to the bystander's group and the wrongdoer belonged to the other group (IV-OW condition), toddlers expected the bystander to engage in indirect TPP: They detected a violation when she chose to help the wrongdoer complete her puzzle (this effect was eliminated when affiliations remained the same but no theft occurred). Conversely, when the wrongdoer belonged to the bystander's group and the victim belonged to the other group (OV-IW condition), toddlers expected no TPP: They now detected a violation when the bystander prevented the wrongdoer from completing her puzzle. Thus, like adults and older children (34–38, 64), toddlers appear to hold selective expectations about TPP: They expect TPP for a mild transgression directed at an ingroup victim, but they expect no TPP for the same transgression when directed at an outgroup victim.

Experiment 2. Exp. 2 addressed two main questions. First, would toddlers still expect indirect TPP when the bystander, victim, and wrongdoer all belonged to the same group? Positive results would rule out the possibility that toddlers expected TPP for an ingroup victim only when the wrongdoer was an outgroup member. Such results, together with those of the IV-OW condition in Exp. 1, would suggest that TPP was expected whenever the transgression was directed at an ingroup victim, regardless of the wrongdoer's group affiliation. Second, would toddlers still expect no TPP when the wrongdoer as well as the victim belonged to the other group? Here, negative results would rule out the possibility that toddlers expected TPP both (*i*) when the victim was an ingroup member and (*ii*) when the wrongdoer stole from her own

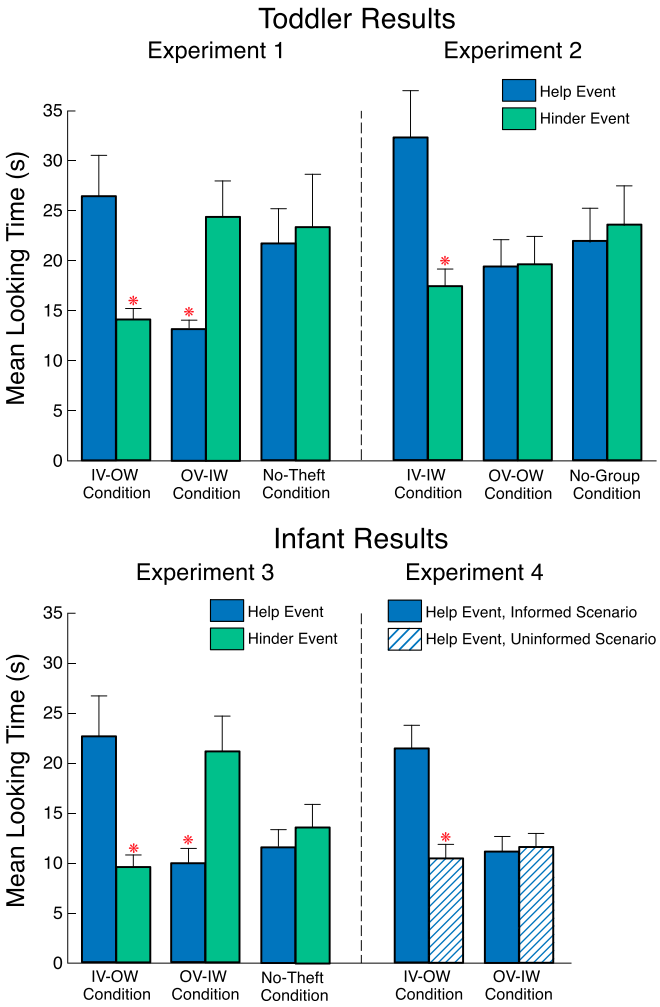

Fig. 3. Mean test looking times in Exps. 1–4, separately by condition and event. Error bars represent SEM, and an asterisk denotes a significant difference between the two events or scenarios in a condition.

group, as though this constituted a transgression more deserving of punishment. These results, together with those of the OV-IW condition in Exp. 1, would suggest that no TPP was expected when the transgression was directed at an outgroup victim, irrespective of the wrongdoer's group affiliation.

Toddlers were assigned to one of three conditions (Fig. 1*B*). The first condition was identical to the IV-OW condition of Exp. 1 except that all three experimenters now belonged to the same group (IV-IW condition). Finding that toddlers looked significantly longer if shown the help as opposed to the hinder event would confirm the result of the IV-OW condition and would indicate that TPP was expected for both ingroup and outgroup wrongdoers who stole from an ingroup victim. In the second condition, the wrongdoer and the victim both belonged to the other group (OV-OW condition). Finding that toddlers looked equally at the help and hinder events would extend the result of the OV-IW condition in Exp. 1 and show that when the transgression involved an outgroup victim, toddlers' expectations about the bystander's actions toward the wrongdoer were those typically observed for interactions within groups (OV-IW condition) and between groups (OV-OW condition). Finally, the third condition was identical to the IV-IW, condition except that in the two familiarization trials the experimenters used "I saw an *X*!" phrases (60) that provided no information about their group memberships, which therefore remained unspecified (e.g., "I saw a topid!", "I saw a topid, too!", "I saw a jaybo!"; no-group condition). This control condition served to rule out the possibility that toddlers in Exps. 1 and 2 generally expected TPP unless the victim was identified as outgroup member (i.e., that toddlers expected TPP for ingroup victims and for victims whose group memberships were unspecified, just not for outgroup victims). Finding equal looking times at the help and hinder events, in contrast to the IV-IW condition, would make clear that toddlers expected the bystander to engage in TPP only when the victim was specifically identified as an ingroup member.

Looking times in the final phase of the test trial (Fig. 3) were compared by an ANOVA with condition (IV-IW, OV-OW, or no-group) and event (help or hinder) as between-subject factors. The analysis yielded only a significant Condition × Event interaction, $F(2, 48) = 3.78$, $P = 0.030$, $\eta_p^2 = 0.136$. Planned comparisons revealed that, as predicted, toddlers in the IV-IW condition looked significantly longer if shown the help event ($M = 32.27$, SD = 13.95) as opposed to the hinder event ($M = 17.40$, SD = 4.99), $F(1, 48) = 9.98$, $P = 0.003$, $d = 1.42$; toddlers in the OV-OW condition looked equally at the help ($M = 19.39$, SD = 8.14) and hinder ($M = 19.61$, $SD = 8.27$) events, $F(1, 48) = 0.00$, $P > 0.250$, $d = -0.03$); and toddlers in the no-group condition also looked equally at the help ($M = 21.93$, SD = 10.19) and hinder ($M = 23.57$, SD = 11.84) events, $F(1, 48) = 0.12$, $P > 0.250$, $d = -0.15$). Wilcoxon rank-sum tests confirmed the results of the IV-IW ($z = 2.74$, $P = 0.006$), OV-OW ($Z = -0.26$, $P > 0.250$), and no-group ($z = -0.18$, $P > 0.250$) conditions.

Additional analyses were conducted to compare the main results of Exp. 2 to those of Exp. 1. As expected, the IV-IW condition differed significantly from the OV-IW condition [Condition × Event interaction: $F(1, 32) = 18.20$, $P < 0.001$, $\eta_p^2 = 0.363$], but not from the IV-OW condition [$F(1, 32) = 0.15$, $P > 0.250$, $\eta_p^2 = 0.005$]. Also as expected, the OV-OW condition differed significantly from both the IV-OW condition [$F(1, 32) = 4.81$, $P = 0.036$, $\eta_p^2 = 0.131$] and the OV-IW condition [$F(1, 32) = 4.33$. $P = 0.046$, $\eta_p^2 = 0.119$].

The results of Exp. 2 confirmed and extended those of Exp. 1. When the victim was an ingroup member, toddlers expected the bystander to engage in indirect TPP, and this was true whether the wrongdoer was an ingroup member (IV-IW condition) or an outgroup member (IV-OW condition). When the victim was an outgroup member, however, toddlers expected no TPP; instead, their expectations mirrored those found in previous reports (46, 53–55, 58, 60, 69, 70). Thus, when the wrongdoer was an ingroup member, toddlers expected the bystander to provide help and detected a violation when she did not (OV-IW condition); when the wrongdoer was an outgroup member, toddlers held no expectation as to whether the bystander would provide help or not (OV-OW condition); and when group affiliations were unspecified, toddlers again held no particular expectation about the bystander's actions toward the wrongdoer (no-group condition).

Experiments with Infants

In Exps. 1 and 2, 2.5-y-old toddlers expected indirect TPP for a mild transgression against an ingroup victim, but they expected no TPP when the same transgression was directed at an outgroup victim. Exps. 3 and 4 examined whether 1-y-old infants would hold similar expectations.

Experiment 3. Infants in Exp. 3 were tested using a procedure similar to that of Exp. 1 (Fig. 1*A*), with four changes designed to render the procedure suitable for these young subjects. First, given infants' limited linguistic ability, group memberships were marked by novel outfits, instead of novel labels. One outfit consisted of tiger ears, a black turtleneck, and a tiger-fur collar, and the other outfit consisted of an orange hoodie and purple eyeglasses (Fig. 4*A*). As before, across conditions, we manipulated the group memberships of the wrongdoer and the victim relative to that of the bystander.

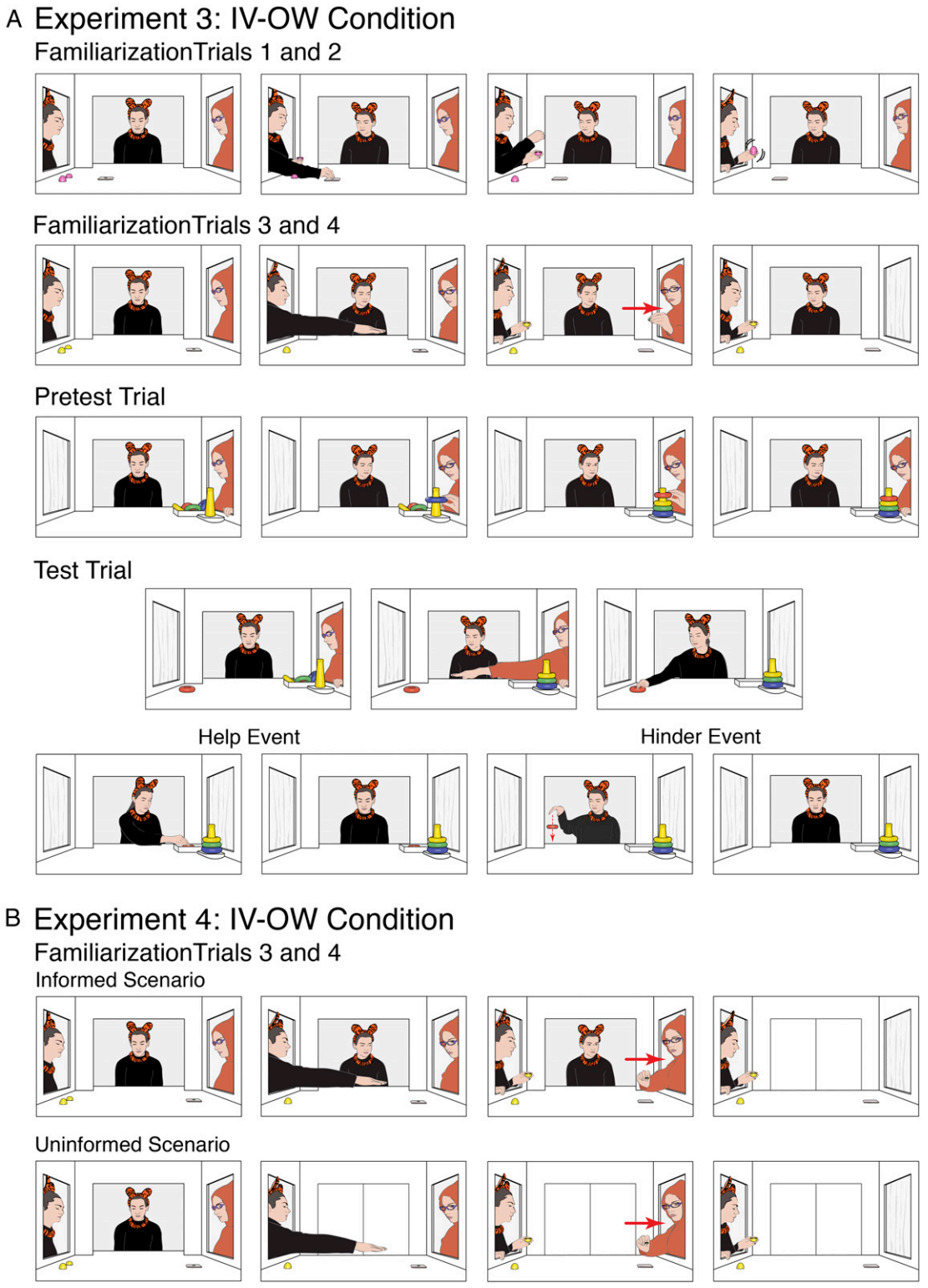

Fig. 4. Familiarization, pretest, and test trials in the IV-OW condition of Exp. 3 (*A*). In the IV-OW condition of Exp. 4 (*B*), infants saw similar events except for familiarization trials 3 and 4, which differed for the informed and uninformed scenarios.

Second, different but comparable stimuli were used: The victim now created her rattle by placing a marble in a toy egg (instead of a block in a cup), and the wrongdoer now built a tower of rings (instead of a puzzle). Third, to help infants grasp the complex events they were shown, each familiarization trial was repeated twice (infants in the no-theft condition simply saw the first familiarization trial twice). In addition, the test trial was preceded by a pretest trial designed to introduce the wrongdoer's goal. In this trial, the victim was absent (her window was closed); while the bystander watched, the wrongdoer retrieved four rings of different sizes and colors from a tray near her window and stacked them on a pole by decreasing size. In

the test trial, the fourth ring rested across the apparatus from the wrongdoer, out of her reach (but within the bystander's reach). Finally, due to the preceding changes (e.g., no labeling in the familiarization trials, multiple trials before the test trial), slightly different criteria were used to end trials (*Methods*).

Looking times during the final phase of the test trial (Fig. 3) were analyzed as in Exp. 1. The analysis yielded only a significant Condition × Event interaction, $F(2, 48) = 11.18$, $P < 0.001$, $\eta_p^2 = 0.318$. Planned comparisons revealed that as in Exp. 1, infants in the IV-OW condition looked significantly longer if shown the help event ($M = 22.59$, SD = 12.09) as opposed to the hinder event ($M = 9.54$, SD = 3.51), $F(1, 48) = 12.68$, $P < 0.001$, $d = 1.47$; infants in the OV-IW condition looked significantly longer if shown the hinder event ($M = 21.10$, SD = 10.59) as opposed to the help event ($M = 9.89$, SD = 4.23), $F(1, 48) = 9.37$, $P = 0.004$, $d = 1.39$; and infants in the no-theft condition looked about equally at the help ($M = 11.57$, SD = 5.02) and hinder ($M = 13.61$, SD = 6.95) events, $F(1, 48) = 0.31$, $P > 0.250$, $d = -0.34$. Wilcoxon rank-sum tests confirmed the results of the IV-OW ($z = 2.38$, $P = 0.017$), OV-IW ($z = -2.52$, $P = 0.012$), and no-theft ($z = -0.66$, $P > 0.250$) conditions.

Results were the same as in Exp. 1, indicating that infants (*i*) could use the women's outfits to determine group memberships and (*ii*) held selective expectations about whether the bystander would engage in indirect TPP. Thus, infants detected a violation when the bystander chose to help an outgroup wrongdoer who had harmed an ingroup victim (IV-OW condition); they detected a violation when the bystander chose not to help an ingroup wrongdoer who had harmed an outgroup victim (OV-IW condition); and they looked about equally whether the bystander chose to help or to hinder an outgroup member who had harmed no one (no-theft condition). These results thus extended downward to 1 y of age the conclusion that considerations of group membership modulate early expectations about indirect TPP.

Experiment 4. Exp. 4 had two goals. One was to confirm that infants would expect the bystander to engage in indirect TPP only when the victim belonged to her group. The other goal was to garner evidence that infants would hold this expectation only when the bystander witnessed the wrongdoer's transgression. Such evidence was essential to draw inferences about TPP: If infants responded similarly whether or not the bystander observed the transgression, it would be difficult to argue that they were forming expectations about whether the bystander would deploy TPP. Prior violation-of-expectation tasks (56, 71) indicate that infants are, indeed, sensitive to what information bystanders possess or lack about others' actions. For example, 10-mo-olds detected a violation when an informed bystander, who had observed a fair and an unfair distributor's actions, later chose to give a treat to the unfair as opposed to the fair distributor; this effect was eliminated, however, when the bystander was uninformed about the distributors' actions (56). In line with these results, infants in Exp. 4 were assigned to the same IV-OW and OV-IW conditions as in Exp. 3, with two exceptions (Fig. 1*C*). First, all infants saw the help event in the test trial. Second, in both conditions, the bystander now left the scene in the third and four familiarization trials, either after ("informed" scenario) or before ("uninformed" scenario) the wrongdoer stole from the victim (Fig. 4*B*). In each case, the bystander left the scene by closing two small doors that filled her window. In presenting infants with these informed vs. uninformed scenarios, we aimed to confirm that even our youngest subjects were reasoning about the bystander's response to the wrongdoer's transgression, and not simply forming diffuse expectations about how the wrongdoer should be treated by others.

Looking times in the final phase of the test trial (Fig. 3) were compared by an ANOVA with condition (IV-OW or OV-IW) and scenario (informed or uninformed) as between-subject factors. The analysis yielded main effects of condition, $F(1, 32) = 7.41$, $P = 0.010$, and scenario, $F(1, 32) = 9.85$, $P = 0.004$, as well as a significant Condition × Scenario interaction, $F(1, 32) = 11.54$, $P = 0.002$, $\eta_p^2 = 0.265$. Planned comparisons revealed that, as predicted, infants in the IV-OW condition looked significantly longer if shown the informed scenario ($M = 21.41$, SD = 7.05) as opposed to the uninformed scenario ($M = 10.44$, SD = 4.10), $F(1, 32) = 21.36$, $P < 0.001$, $d = 1.90$, whereas infants in the OV-IW condition looked about equally at the informed ($M = 11.14$, SD = 4.34) and uninformed ($M = 11.58$, SD = 3.99) scenarios, $F(1, 32) = 0.03$, $P > 0.250$, $d = -0.11$. Wilcoxon rank-sum tests confirmed the results of the IV-OW ($z = 3.27$, $P = 0.001$) and OV-IW ($z = -0.53$, $P > 0.250$) conditions.

Infants expected indirect TPP when the bystander witnessed a transgression against an ingroup victim (informed scenario of the IV-OW condition), but expected no TPP when this scenario was altered in one of two ways: The bystander left the scene before the transgression occurred and thus was unaware of it (uninformed scenario of the IV-OW condition), or the victim did not belong to the same group as the bystander (informed scenario of the OV-IW condition). These results confirm those of Exp. 3 and make clear that infants were forming expectations about how the bystander would respond to the transgression she had observed, rather than expectations about how the wrongdoer would be treated following her transgression.

Overall Analyses

Finally, we compared toddlers and infants' test responses to the help event ($n = 72$) when the victim was an ingroup member (IV condition) or an outgroup member (OV condition). For the toddlers, we pooled the data from the IV-OW and IV-IW conditions (IV condition) and contrasted those to the pooled data from the OV-IW and OV-OW conditions (OV condition). For the infants, we pooled the data from the IV-OW and IV-OW/informed-scenario conditions (IV condition) and contrasted those to the pooled data from the OV-IW and OV-IW/informed-scenario conditions (OV condition). These data were subjected to an ANOVA with age (2.5 or 1) and victim (IV or OV) as between-subject factors. The main effect of age was significant, $F(1, 68) = 9.48$, $P = 0.003$, $\eta_p^2 = 0.122$. Overall, toddlers looked significantly longer ($M = 22.81$, SD = 12.20) than infants ($M = 16.26$, SD = 9.35), most likely because the two age groups were tested using somewhat different, age-appropriate procedures. Critically, the main effect of victim was also significant, $F(1, 68) = 33.26$, $P < 0.001$, $\eta_p^2 = 0.329$, but the Age × Victim interaction was not, $F(1, 68) = 0.14$, $P > 0.250$, $\eta_p^2 = 0.002$, suggesting that the two age groups responded similarly in the IV and OV conditions. Planned comparisons confirmed that (*i*) toddlers looked significantly longer at the help event in the IV condition ($M = 29.35$, SD = 13.12) than in the OV condition ($M = 16.28$, SD = 6.62), $F(1, 68) = 18.85$, $P < 0.001$, $d = 1.26$, and (*ii*) infants also looked significantly longer at the help event in the IV condition ($M = 22.00$, SD = 9.62) than in the OV condition ($M = 10.52$, SD = 4.21), $F(1, 68) = 14.55$, $P < 0.001$, $d = 1.55$). Wilcoxon rank-sum tests confirmed the results with the toddlers ($z = 3.58$, $P < 0.001$) and infants ($z = 3.88$, $P < 0.001$).

General Discussion

In prior research, adults and children as young as 6 y of age were more likely to engage in TPP in response to a wrongdoer's transgression when the victim belonged to their own group, as opposed to a different group (34–38, 64). Echoing these results, we found that 2.5-y-old toddlers and 1-y-old infants expected a bystander to engage in indirect TPP following a wrongdoer's transgression when the victim belonged to the bystander's own group, but not when the victim belonged to a different group.

In four experiments, children first saw a wrongdoer steal a toy from a victim, while a bystander watched. Next, the wrongdoer

required help to complete a task because a needed object was out of reach. The bystander either brought the object closer (help event) or threw it away (hinder event). Across conditions, we varied the group memberships of the wrongdoer and the victim, relative to that of the bystander, using novel labels (toddlers) or novel outfits (infants). When the victim was an ingroup member, toddlers detected a violation in the help event, suggesting that they expected the bystander to refrain from helping someone who had harmed an ingroup member. This effect held regardless of whether the wrongdoer was an outgroup member (IV-OW condition) or an ingroup member (IV-IW condition). On the other hand, this effect was eliminated (*i*) when the victim was an outgroup member (OV-IW and OV-OW conditions), (*ii*) when group memberships were not specified (no-group condition), and (*iii*) when no theft occurred (no-theft condition). Like toddlers, infants detected a violation when the bystander chose to help someone who had harmed an ingroup member (IV-OW and IV-OW/informed-scenario). This effect was eliminated (*i*) when the victim was an outgroup member (OV-IW and OV-IW/informed scenario), (*ii*) when the bystander was absent during the wrongdoer's transgression (IV-OW/uninformed-scenario), or (*iii*) when no theft occurred (no-theft condition). Turning to the hinder event, toddlers detected a violation when the bystander chose to hinder an ingroup wrongdoer who had harmed an outgroup victim (OV-IW condition), suggesting that they expected the bystander to dismiss the wrongdoer's transgression and provide her with the help she needed. This effect was eliminated when the ingroup wrongdoer transgressed against an ingroup victim (IV-IW condition), indicating that hindering the ingroup wrongdoer then became acceptable (milder negative actions, such as ignoring the ingroup wrongdoer's need for help, would presumably also become acceptable). Finally, hindering was always acceptable when directed at someone who was not an ingroup member: Toddlers detected no violation when the bystander hindered (*i*) an outgroup wrongdoer who had harmed an ingroup member (IV-OW condition), had harmed an outgroup member (OV-OW condition), or had harmed no one (no-theft condition), or (*ii*) a wrongdoer whose group membership was unspecified (no-group condition). Like toddlers, infants detected a violation when the bystander hindered an ingroup wrongdoer who had harmed an outgroup victim (OV-IW condition), but they detected no violation when the bystander hindered an outgroup member, whether she had harmed an ingroup victim (IV-OW condition) or had harmed no one (no-theft condition). This complex array of findings is unlikely to be due to low-level factors: At each age, children saw the bystander perform exactly the same helping action or the same hindering action across conditions.

Together, our results support three main conclusions. First, they provide additional evidence that for young children (*i*) mere categorization of unfamiliar individuals into minimal groups, however it is achieved, is sufficient to elicit rich expectations about interactions within and between groups (6, 39, 58, 60, 69, 70); (*ii*) all other things being equal, helping an ingroup member in need of assistance is viewed as expected or obligatory, whereas helping a noningroup member (i.e., an outgroup member or an individual whose group membership is not specified) is viewed as optional (46, 53–55, 60, 70); and (*iii*) all other things being equal, a mild harmful action (e.g., hindering) is viewed as unacceptable when directed at an ingroup member, but as acceptable or permissible when directed at a noningroup member (46, 53–55, 58, 69, 70).

Second, our results extend these prior findings by showing how children's expectations change following a transgression against an ingroup victim. When the transgression is perpetrated by an ingroup wrongdoer, these expectations change considerably: Helping the wrongdoer becomes unacceptable, whereas hindering the wrongdoer becomes acceptable. Both of these changes—the withholding of assistance that would otherwise have been obligatory, and the infliction of mild harm that would otherwise have been unacceptable—have negative consequences for the ingroup wrongdoer. When the transgression is perpetrated by an outgroup wrongdoer, expectations also change, but less dramatically: Helping the outgroup wrongdoer, which was previously optional, now becomes unacceptable, whereas hindering the wrongdoer remains acceptable. Thus, for the outgroup wrongdoer, the main consequence of punishment is that assistance that might have been offered is now unlikely to be forthcoming. This is not to say that TPP against outgroup wrongdoers is generally less severe or less consequential than TPP against ingroup wrongdoers, and we return to this point below. All that we are suggesting here is that mild indirect TPP may impact ingroup wrongdoers more than it does outgroup wrongdoers.

Finally, our results make clear that these changes in children's expectations occur only with transgressions against ingroup victims. When the victim was not specifically identified as a member of the bystander's group, toddlers' and infants' expectations were those typically found for interactions between ingroup individuals (57–63) or between noningroup individuals (46, 53–55, 57–63, 69, 70). Thus, in the case of an ingroup wrongdoer, the bystander was expected to provide help and to refrain from hindering; in the case of a noningroup wrongdoer, however, helping and hindering were both viewed as acceptable courses of action.

The presence of selective expectations about indirect TPP in toddlers and infants gives weight to the notion that these expectations reflect an abstract principle of ingroup support, which is part of the "first draft" (8) of human moral cognition (4–10, 57, 59, 60). As noted in the Introduction, from a very young age, children's concern for ingroup support carries a rich set of expectations related to caring for ingroup members and showing them loyalty (57–63, 69). Our research extends these findings by demonstrating that just as children expect individuals to refrain from harming ingroup members, they also expect individuals to punish, at least indirectly, harm to ingroup members, whether perpetrated by ingroup or outgroup wrongdoers. Thus, beginning early in life, one key function of indirect TPP appears to be that of protecting ingroup members, by making clear that harmful actions toward them will have adverse consequences.

Future research could build on our findings in several directions. One would be to gather converging evidence for our conclusions by using indirect TPP scenarios with different transgressions (e.g., being unfair), different punitive actions (e.g., choosing not to share resources with the wrongdoer), and so on. A second direction would be to compare young children's expectations about indirect and direct TPP; in this context, it may be particularly interesting to vary the status of the punisher as a leader vs. a follower (72). Evidence that children more closely associate direct TPP with leaders and indirect TPP with followers, especially when the potential costs to the punisher are high (e.g., due to possible retaliation by the wrongdoer), would dovetail well with some of the adult findings reviewed in the Introduction (23–33).

A third research direction would be to examine whether young children would tolerate harsher punishments for outgroup wrongdoers compared with ingroup wrongdoers. Previous research on this issue with adults and older children has yielded mixed results (38), with some reports indicating harsher punishments for outgroup wrongdoers (34, 35, 37, 64, 73, 74), some indicating harsher punishments for ingroup wrongdoers (75), and some indicating similar punishments for ingroup and outgroup wrongdoers (36, 66, 76). To study this issue with young children, our puzzle scenario could be modified to involve a much harsher punishment: Imagine that instead of throwing away the final piece of the puzzle, the bystander now destroyed each and every piece of it. Without provocation, these severe harmful actions would presumably be viewed as unacceptable, even when directed at an outgroup member. Following a transgression against an ingroup victim, however, toddlers and infants might view these punitive actions as acceptable when directed at an outgroup wrongdoer, but as overly harsh when

directed at an ingroup wrongdoer. Such findings would suggest that early expectations about TPP are selective in at least two ways, both consistent with the principle of ingroup support: Children would not only expect individuals to deploy TPP for mild transgressions against ingroup but not outgroup victims (as shown in the present research), but they would also expect individuals to deploy less TPP toward ingroup as opposed to outgroup wrongdoers for the same transgression.

This last point leads to a fourth direction for future research. The present experiments involved only a mild transgression, and toddlers and infants expected the bystander to engage in indirect TPP for this transgression only when it victimized an ingroup member. When it victimized an outgroup member, the bystander was expected not to engage in TPP, and this was true even when the wrongdoer and the victim belonged to the same group, so that the wrongdoer's actions violated ingroup support (OV-OW condition). Given these results, it might be suggested that early in development, TPP always serves a consequentialist as opposed to a deontological function (77–81): Toddlers and infants expected the bystander to deploy TPP to protect her ingroup and its members (including the bystander herself) from further harm, and not as a deserved retribution for a moral transgression. However, such a suggestion might be premature. Imagine that our mild transgression was replaced with a much harsher transgression or series of transgressions. Just as nations sometimes attempt to stop or punish heinous crimes perpetrated within other nations, young children might expect TPP to be deployed for all egregious unprovoked transgressions, regardless of whether the victims are ingroup or outgroup members. If true, one interpretation of these findings might be that as transgressions become more severe, TPP takes on more deontological or retributive overtones.

In sum, the present experiments showed that toddlers and infants expected an individual to engage in indirect TPP for a mild harm transgression when the victim was specifically identified as one of the individual's ingroup members, but not otherwise. Our findings thus provide further evidence for an abstract and early-emerging expectation of ingroup support and, more generally, for the richness and subtlety of the "first draft" of human moral cognition.

Methods

Power Analysis. In the report of Jin and Baillargeon (60), which also examined early expectations about ingroup support using a between-subjects design and live minimal-group manipulations, the average Condition × Event effect size (η_p^2) across experiments was 0.19. An a priori power analysis using G*Power (82) based on this value indicated that with power set at 0.80 and α set at 0.05, the minimum number of participants required per cell (i.e., per combination of condition and event) was eight participants for 3 × 2 designs (as in Exps. 1–3) and nine participants for 2 × 2 designs (as in Exp. 4). For consistency, all four experiments used nine participants per cell.

In all experiments, each child's parent gave written informed consent, and the protocol was approved by the Institutional Review Board of the University of Illinois at Urbana–Champaign.

Exps. 1 and 2.

Participants. Participants were 108 English-speaking toddlers (55 male; *M* = 29 mo, 17 d, range = 27 mo, 26 d to 31 mo, 25 d). Another eight toddlers were excluded, five because they were distracted or inattentive, and three because their test looking times were over 3 SDs from the condition mean (one in the IV-OW condition and two in the OV-IW condition).

Apparatus. The apparatus consisted of a brightly lit display booth (201 cm high × 100 cm wide × 74 cm deep) with a large opening (57 cm × 93 cm) in its front wall; between trials, a supervisor lowered a curtain in front of this opening. Inside the apparatus, the walls were painted white and the floor was covered with a pastel adhesive paper. The victim wore a maroon shirt and the wrongdoer wore a gray shirt; each knelt at a side window (57 cm × 48 cm) with a curtain that could be drawn aside. The bystander wore a blue shirt and sat at a window (72 cm × 96 cm) in the back wall. The victim used the "topid" label, and the wrongdoer and bystander used either the "topid" or the "jaybo" label, depending on the condition. During the trials, the experimenters never made eye contact with the toddler: As the events unfolded, they looked at each other or at the objects they acted on but otherwise kept their eyes on a neutral point on the apparatus floor. Behind the experimenters, white curtains surrounded the apparatus and hid the testing room. During each session, one camera captured an image of the trials, and another camera captured an image of the toddler; the images were combined, projected onto a computer screen, and monitored by the supervisor to confirm that trials followed the prescribed scripts. Recorded sessions were also checked off-line for observer and experimenter accuracy.

Procedure. Toddlers sat on a parent's lap in front of the apparatus; parents were instructed to remain silent and close their eyes during the test trial. Two hidden observers monitored each toddler's looking behavior; observers were naïve about which event was shown in the test trial (to muffle sounds, the bottom of the final puzzle piece was covered with felt, and it was dropped into fiberfill). Looking times were computed using the primary observer's responses. Each trial began with a paused pretrial that ended when the toddler had cumulated 2 s of looking, to allow the toddler to orient to the apparatus before the trial proper began. The durations of the familiarization trials and the initial phase of the test trial were computer-controlled; the experimenters' actions followed a precise second-by-second script that lasted 18–52 s, depending on the trial. Toddlers were highly attentive during these trials and looked, on average, for 98% of each trial. The final phase of the test trial ended when toddlers either looked away for 0.5 consecutive seconds after having looked for at least 10 cumulative seconds or looked for a maximum of 50 cumulative seconds; the 10-s minimum value allowed toddlers to process the bystander's actions before the trial could end. Interobserver agreement during the final phase was calculated for 101 of 108 toddlers (only one observer was present for the other toddlers) 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 averaged 95% across both experiments.

Exps. 3 and 4.

Participants. Participants were 90 healthy term infants (45 male; *M* = 13 mo, 13 d, range = 12 mo, 6 d to 14 mo, 22 d). Another 14 infants were excluded, 10 because they were fussy or inattentive, 2 because of parental interference, and 2 because their test looking times were over three SDs from the condition mean (one in the IV-OW condition and one in the OV-IW condition).

Apparatus. The apparatus was similar to that in Exps. 1 and 2 except that the bystander's window was smaller (56 cm × 72 cm) and could be closed in Exp. 4 with two small doors (each 56 cm × 36 cm). The victim wore the tiger outfit, the wrongdoer wore the hoodie outfit, and the bystander wore whichever outfit was appropriate for the condition.

Procedure. The procedure was identical to that in Exps. 1 and 2 except as follows. First, infants received four familiarization trials (two in the no-theft condition) and one pretest trial before the test trial. The durations of the familiarization trials, pretest trial, and initial phase of the test trial were computer-controlled; in each trial, the experimenters' actions followed second-by-second scripts that lasted 19–37 s, depending on the trial. Infants were highly attentive during these trials and looked, on average, for 99% of each trial. Second, because all infants in Exp. 4 saw the help event in the test trial, the primary observer was absent from the testing room during the familiarization trials and thus was naïve about whether infants saw the informed or the uninformed scenario. Third, because infants received up to five trials before the test trial (instead of only two as in Exps. 1 and 2), the minimum value of the final phase of the test trial was decreased from 10 to 5 cumulative seconds. Interobserver agreement during the final phase was calculated for 82 of 90 infants (only one observer was present for the other infants) and averaged 97% across both experiments.

Data and Preliminary Analyses. The data from all four experiments are available in Dataset S1. Preliminary analyses of the test data revealed no significant interactions of condition and event (or scenario, Exp. 4) with subject's sex or with side of first label (Exps. 1 and 2); the data were thus collapsed across these factors in subsequent analyses.

ACKNOWLEDGMENTS. We thank Jerry DeJong and Rose Scott for helpful comments; the University of Illinois at Urbana–Champaign Infant Cognition Laboratory for their help with the data collection; graphic artist Steve Holland for producing the figures; and the families who participated in the experiments. This research was supported by a grant from the John Templeton Foundation (to R.B.).

1. Boyd R, Richerson PJ (2009) Culture and the evolution of human cooperation. *Philos Trans R Soc Lond B Biol Sci* 364:3281–3288.
2. Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425:785–791.
3. Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560–1563.
4. Baillargeon R, et al. (2015) Psychological and sociomoral reasoning in infancy. *APA Handbook of Personality and Social Psychology*, eds Borgida E, Bargh J (American Psychological Association, Washington, DC), Vol 1, pp 79–150.
5. Brewer MB (1999) The psychology of prejudice: Ingroup love and outgroup hate? *J Soc Issues* 55:429–444.
6. Dunham Y (2018) Mere membership. *Trends Cogn Sci* 22:780–793.
7. Fehr E, Schurtenberger I (2018) Normative foundations of human cooperation. *Nat Hum Behav* 2:458–468.
8. Graham J, et al. (2013) Moral foundations theory: The pragmatic validity of moral pluralism. *Adv Exp Soc Psychol* 47:55–130.
9. Rai TS, Fiske AP (2011) Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychol Rev* 118:57–75.
10. Shweder RA, Much NC, Mahapatra M, Park L (1997) The "big three" of morality (autonomy, community, and divinity) and the "big three" explanations of suffering. *Morality and Health*, eds Brandt A, Rozin P (Routledge, New York), pp 119–169.
11. Balliet D, Mulder LB, Van Lange PAM (2011) Reward, punishment, and cooperation: A meta-analysis. *Psychol Bull* 137:594–615.
12. Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Am Econ Rev* 90:980–994.
13. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
14. Raihani NJ, Thornton A, Bshary R (2012) Punishment and cooperation in nature. *Trends Ecol Evol* 27:288–295.
15. Buckholtz JW, et al. (2008) The neural correlates of third-party punishment. *Neuron* 60:930–940.
16. Fehr E, Fischbacher U (2004) Third-party punishment and social norms. *Evol Hum Behav* 25:63–87.
17. Henrich J, et al. (2010) Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327:1480–1484.
18. Henrich J, et al. (2006) Costly punishment across human societies. *Science* 312: 1767–1770.
19. Krasnow MM, Delton AW, Cosmides L, Tooby J (2015) Group cooperation without group selection: Modest punishment can recruit much cooperation. *PLoS One* 10:e0124561.
20. Nikiforakis N, Mitchell H (2014) Mixing the carrots and the sticks: Third party punishment and reward. *Exp Econ* 17:1–23.
21. Rockenbach B, Milinski M (2006) The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444:718–723.
22. Ule A, Schram A, Riedl A, Cason TN (2009) Indirect punishment and generosity toward strangers. *Science* 326:1701–1704.
23. Balafoutas L, Nikiforakis N (2012) Norm enforcement in the city: A natural field experiment. *Eur Econ Rev* 56:1773–1785.
24. Baumard N (2010) Has punishment played a role in the evolution of cooperation? A critical review. *Mind Soc* 9:171–192.
25. Guala F (2012) Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behav Brain Sci* 35:1–15.
26. Winter F, Zhang N (2018) Social norm enforcement in ethnically diverse communities. *Proc Natl Acad Sci USA* 115:2722–2727.
27. Baldassarri D, Grossman G (2011) Centralized sanctioning and legitimate authority promote cooperation in humans. *Proc Natl Acad Sci USA* 108:11023–11027.
28. Hershcovis SM, et al. (2017) Witnessing wrongdoing: The effects of observer power on incivility intervention in the workplace. *Organ Behav Hum Decis Process* 142:45–57.
29. Smith JE, et al. (2016) Leadership in mammalian societies: Emergence, distribution, power, and payoff. *Trends Ecol Evol* 31:54–66.
30. Wiessner P (2005) Norm enforcement among the Ju/'hoansi Bushmen: A case of strong reciprocity? *Hum Nat* 16:115–145.
31. Balafoutas L, Nikiforakis N, Rockenbach B (2014) Direct and indirect punishment among strangers in the field. *Proc Natl Acad Sci USA* 111:15924–15927.
32. Feinberg M, Willer R, Schultz M (2014) Gossip and ostracism promote cooperation in groups. *Psychol Sci* 25:656–664.
33. Melis AP, Semmann D (2010) How is human cooperation different? *Philos Trans R Soc Lond B Biol Sci* 365:2663–2674.
34. Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. *Nature* 442: 912–915.
35. Delton AW, Krasnow MM (2017) The psychology of deterrence explains why group membership matters for third-party punishment. *Evol Hum Behav* 38:734–743.
36. Goette L, Huffman D, Meier S (2006) The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am Econ Rev* 96:212–216.
37. Lieberman D, Linke L (2007) The effect of social category on third-party punishment. *Evol Psychol* 5:289–305.
38. McAuliffe K, Dunham Y (2016) Group bias in cooperative norm enforcement. *Philos Trans R Soc Lond B Biol Sci* 371:20150073.
39. Dunham Y, Baron AS, Carey S (2011) Consequences of "minimal" group affiliations in children. *Child Dev* 82:793–811.
40. Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social categorization and intergroup behaviour. *Eur J Soc Psychol* 1:149–178.
41. Rakoczy H, Kaufmann M, Lohse K (2016) Young children understand the normative force of standards of equal resource distribution. *J Exp Child Psychol* 150:396–403.
42. Riedl K, Jensen K, Call J, Tomasello M (2015) Restorative justice in children. *Curr Biol* 25:1731–1735.
43. Rossano F, Rakoczy H, Tomasello M (2011) Young children's understanding of violations of property rights. *Cognition* 121:219–227.
44. Vaish A, Missana M, Tomasello M (2011) Three-year-old children intervene in third-party moral transgressions. *Br J Dev Psychol* 29:124–130.
45. Smetana JG, et al. (2012) Developmental changes and individual differences in young children's moral judgments. *Child Dev* 83:683–696.
46. Hamlin JK, Wynn K, Bloom P, Mahajan N (2011) How infants and toddlers react to antisocial others. *Proc Natl Acad Sci USA* 108:19931–19936.
47. Kanakogi Y, et al. (2017) Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat Hum Behav* 1:0037.
48. Dahl A, Schuck RK, Campos JJ (2013) Do young toddlers act on their social preferences? *Dev Psychol* 49:1964–1970.
49. Surian L, Franchin L (2017) Toddlers selectively help fair agents. *Front Psychol* 8:944.
50. Vaish A, Carpenter M, Tomasello M (2010) Young children selectively avoid helping people with harmful intentions. *Child Dev* 81:1661–1669.
51. Tasimi A, Wynn K (2016) Costly rejection of wrongdoers by infants and children. *Cognition* 151:76–79.
52. Buon M, et al. (2014) Friend or foe? Early social evaluation of human interactions. *PLoS One* 9:e88612.
53. Hamlin JK, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. *Nature* 450:557–559.
54. Fawcett C, Liszkowski U (2012) Infants anticipate others' social preferences. *Infant Child Dev* 21:239–249.
55. Lee YE, Yun JE, Kim EY, Song HJ (2015) The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLoS One* 10:e0135588.
56. Meristo M, Surian L (2013) Do infants detect indirect reciprocity? *Cognition* 129: 102–113.
57. Buyukozer Dawkins M, Ting F, Stavans M, Baillargeon R, Early moral cognition: A principle-based approach. *The Cognitive Neurosciences VI*, eds Poeppel D, Mangun GR, Gazzaniga MS (MIT Press, Cambridge, MA), in press.
58. Rhodes M (2012) Naïve theories of social groups. *Child Dev* 83:1900–1916.
59. Bian L, Sloane S, Baillargeon R (2018) Infants expect ingroup support to override fairness when resources are limited. *Proc Natl Acad Sci USA* 115:2705–2710.
60. Jin KS, Baillargeon R (2017) Infants possess an abstract expectation of ingroup support. *Proc Natl Acad Sci USA* 114:8199–8204.
61. Spokes AC, Spelke ES (2017) The cradle of social knowledge: Infants' reasoning about caregiving and affiliation. *Cognition* 159:102–116.
62. Powell LJ, Spelke ES (2013) Preverbal infants expect members of social groups to act alike. *Proc Natl Acad Sci USA* 110:E3965–E3972.
63. Jin KS, Houston JL, Baillargeon R, Groh AM, Roisman GI (2018) Young infants expect an unfamiliar adult to comfort a crying baby: Evidence from a standard violation-of-expectation task and a novel infant-triggered-video task. *Cognit Psychol* 102:1–20.
64. Jordan JJ, McAuliffe K, Warneken F (2014) Development of in-group favoritism in children's third-party punishment of selfishness. *Proc Natl Acad Sci USA* 111:12710–12715.
65. McAuliffe K, Jordan JJ, Warneken F (2015) Costly third-party punishment in young children. *Cognition* 134:1–10.
66. Schmidt MFH, Rakoczy H, Tomasello M (2012) Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition* 124:325–333.
67. Köster M, Ohmer X, Nguyen TD, Kärtner J (2016) Infants understand others' needs. *Psychol Sci* 27:542–548.
68. Warneken F, Tomasello M (2007) Helping and cooperation at 14 months of age. *Infancy* 11:271–294.
69. Rhodes M, Chalik L (2013) Social categories as markers of intrinsic interpersonal obligations. *Psychol Sci* 24:999–1006.
70. Rhodes M, Hetherington C, Brink K, Wellman HM (2015) Infants' use of social partnerships to predict behavior. *Dev Sci* 18:909–916.
71. Sloane S, Baillargeon R, Premack D (2012) Do infants have a sense of fairness? *Psychol Sci* 23:196–204.
72. Margoni F, Baillargeon R, Surian L (2018) Infants distinguish between leaders and bullies. *Proc Natl Acad Sci USA* 115:E8835–E8843.
73. Schiller B, Baumgartner T, Knoch D (2014) Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evol Hum Behav* 35: 169–175.
74. Yudkin DA, Rothmund T, Twardawski M, Thalla N, Van Bavel JJ (2016) Reflexive intergroup bias in third-party punishment. *J Exp Psychol Gen* 145:1448–1459.
75. Mendoza SA, Lane SP, Amodio DM (2014) For members only: Ingroup punishment of fairness norm violations in the ultimatum game. *Soc Psychol Personal Sci* 5:662–670.
76. McAuliffe K, Dunham Y (2017) Fairness overrides group bias in children's second-party punishment. *J Exp Psychol Gen* 146:485–494.
77. Carlsmith KM, Darley JM, Robinson PH (2002) Why do we punish? Deterrence and just deserts as motives for punishment. *J Pers Soc Psychol* 83:284–299.
78. Cushman F (2015) Punishment in humans: From intuitions to institutions. *Philos Compass* 10:117–133.
79. Jordan JJ, Hoffman M, Bloom P, Rand DG (2016) Third-party punishment as a costly signal of trustworthiness. *Nature* 530:473–476.
80. Krasnow MM, Cosmides L, Pedersen EJ, Tooby J (2012) What are punishment and reputation for? *PLoS One* 7:e45662.
81. Bedau H, Kelly E (2017) Punishment. *Stanford Encyclopedia of Philosophy*, ed Zalta EN. Available at https://plato.stanford.edu/archives/win2017/entries/punishment/. Accessed October 10, 2018.
82. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39:175–191.