

Young Children's Responses to Guilt Displays

Amrisha Vaish, Malinda Carpenter, and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology

Displaying guilt after a transgression serves to appease the victim and other group members, restore interpersonal relationships, and indicate the transgressors' awareness of and desire to conform to the group's norms. We investigated whether and when young children are sensitive to these functions of guilt displays. In Study 1, after 4- and 5-year-old children watched videos of transgressors either displaying guilt (without explicitly apologizing) or not displaying guilt, 5-year-olds appropriately inferred that the victim would be madder at the unremorseful transgressor and would prefer the remorseful transgressor. They also said that they would prefer to interact with the remorseful transgressor, judged the unremorseful transgressor to be meaner, and, in a distribution of resources task, gave more resources to the remorseful transgressor. The 4-year-olds did not draw any of these inferences and distributed the resources equally. However, Study 2 showed that 4-year-olds were able to draw appropriate inferences about transgressors who explicitly apologized versus those who did not apologize. Thus, 4-year-olds seem to know the appeasement functions that explicit apologies serve but only when children have reached the age of 5 years do they seem to grasp the emotions that apologies stand for, namely, guilt and remorse, and the appeasement functions that displaying these emotions serve.

Keywords: guilt, appeasement, social emotions, moral emotions

Guilt is the aversive emotion that follows the realization that one has harmed another person (Nelissen & Zeelenberg, 2009). Like other self-conscious emotions, guilt likely plays a central role in socialization and adherence to conventions, norms, and morals. The aversive feeling of guilt, or the anticipation thereof, is a powerful mechanism that prevents individuals from transgressing and motivates individuals to rectify their transgressions (Hoffman, 1982; Keltner, 1995). Empirical work with adults provides support for this function of guilt (e.g., Ketelaar & Au, 2003), and work with young children also suggests a link between the experience of guilt and moral development (e.g., Kochanska, DeVet, Goldman, Murray, & Putnam, 1994; Kochanska, Gross, Lin, & Nichols, 2002; Zahn-Waxler & Kochanska, 1990).

Why, though, do transgressors display and verbalize their feelings of guilt to others? A prevailing view is that guilt displays serve appeasement functions.¹ Thus, the display of guilt shows

others that the transgressor is also suffering, which evokes sympathy, concern, and forgiveness, and thus reduces the likelihood of punishment (Keltner & Anderson, 2000; Leary, Landel, & Patton, 1996). Guilt displays also signal to others that the transgressor is aware of and committed to the norms of the group, and so the transgression is not reflective of the transgressor's personality, character, or ability. They may furthermore serve as a promise of more acceptable conduct in the future (Castelfranchi & Poggi, 1990; Goffman, 1967; Keltner, Young, & Buswell, 1997; Leary et al., 1996; Nelissen & Zeelenberg, 2009). A transgressor who displays guilt is thus more likely to be seen as self-policing, dependable, and cooperative (Darby & Schlenker, 1989). Empirical research has shown that adults indeed are sensitive to these social and interpersonal functions of displaying guilt (e.g., Ohbuchi, Kameda, & Agarie, 1989; O'Malley & Greenberg, 1983).

The ontogenetic emergence of sensitivity to guilt displays is not well known, but there is a vast literature on the development of emotion comprehension more generally (see, e.g., Saarni, Campos, Camras, & Witherington, 2006; Vaish, Grossmann, & Woodward, 2008). This literature suggests that the perception and categorization of some basic emotions such as happiness, fear, sadness, and anger are evident in the first few months of life (see, e.g., Nelson,

This article was published Online First July 18, 2011.

Amrisha Vaish, Malinda Carpenter, and Michael Tomasello, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

This research was partially supported by a grant from the DFG Excellence Cluster 302 (Languages of Emotions) in cooperation with the Free University Berlin. We thank all the day care centers and children for their friendly cooperation; Manuela Missana, Elvira Portner, Eva Siegert, and Manja Teich for their hard work in creating the videos; Eva Siegert for assistance with data collection; Irka Wetzig for transcribing the data; Maria Baumeister for help with coding; and Tobias Grossmann for helpful discussions.

Correspondence concerning this article should be addressed to Amrisha Vaish, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103, Germany. E-mail: vaish@eva.mpg.de

¹ Note that displays of other self-conscious emotions (e.g., embarrassment and shame) are also thought to serve appeasement functions (see, e.g., Keltner et al., 1997). For instance, adults report high levels of affiliative emotions such as amusement and sympathy in response to others' embarrassment (Keltner et al., 1997; Miller, 1987). Adults also help individuals who previously displayed an appropriate amount of embarrassment more than individuals who displayed too much embarrassment or none at all (Levin & Arluke, 1982; Semin & Manstead, 1981). Guilt displays are thus not alone in serving appeasement functions. In this article, however, we focused on guilt displays alone.

1987, for a review). By around 12–14 months of age, infants are able to use others' displays of basic emotions to guide their own behavior (e.g., Mumme, Fernald, & Herrera, 1996) but not yet to predict others' behavior (Vaish & Woodward, 2010). This latter ability emerges by around the age of 18 months (Repacholi & Gopnik, 1997). Thus, the understanding of basic emotions is well underway by the second year of life.

By contrast, the understanding of more complex social emotions such as embarrassment, shame, and guilt emerges somewhat later (e.g., Brody & Harrison, 1987; Harris, Olthof, Terwogt, & Hardman, 1987; Wiggers & van Lieshout, 1985). For instance, preschoolers and third- and fourth-graders perform better at matching tasks involving basic emotions such as fear and anger than those involving complex emotions such as guilt, jealousy, and pride (Brody & Harrison, 1987). Still, by middle childhood, children have developed some understanding of the causes and consequences of complex emotions. With regard to guilt more specifically, by around the age of 7 years, children are aware of what kinds of situations elicit guilt versus other social emotions (Harris et al., 1987). Children around this age also know that a person who feels guilt would and should apologize, formulate good intentions, and mend or substitute damaged objects (Berti, Garattoni, & Venturini, 2000). However, it remains unclear whether children understand the appeasement functions of guilt displays. Do they make appropriate inferences about the effects that guilt displays have on the victim? Do they themselves prefer to interact with transgressors who display guilt (and are thus more dependable and cooperative group members) than those who display no guilt? Finally, do they cooperate more with (by distributing more resources to) a transgressor who displays guilt than one who does not? These are the questions addressed in the present article.

Similar questions have been asked in the related work on children's understanding of apologies, which are admissions of blameworthiness and regret and thus function as a stand-in for guilt. This work shows that young children do understand and draw appropriate inferences from apologies. For example, apologetic actors are blamed less, forgiven more, liked more, seen as more remorseful, and punished less by 6-year-old children (Darby & Schlenker, 1982, 1989). In addition, children around the age of 4–5 years regard situations in which an actor apologizes as better and more just than ones in which the actor is unapologetic (Irwin & Moore, 1971; Wellman, Larkey, & Somerville, 1979). Also around this age, children attribute negative feelings to an apologetic transgressor and improved feelings to a victim who has received an apology, although they do not yet seem to make inferences about the transgressor's moral character on the basis of whether he or she apologized or not (Smith, Chen, & Harris, 2010). Although this work on the understanding of apologies is interesting and important, it should be noted that children are regularly taught and prompted to apologize by their caregivers and teachers (see Smith et al., 2010). Their ability to make judgments about transgressors who say they are sorry or who are described as having apologized may thus rely rather superficially on the use of key words such as *sorry* or *apologize*. It remains unclear how deeply children understand and what they can infer from the emotions behind apologies, namely, guilt and remorse. In the first study in the current article, we thus examined 4- and 5-year-old children's responses to displays of guilt without the confounding effects of apologies. In the second study, we used the same method as in the first study to

examine children's responses to apologies in order to ensure that any lack of understanding that children displayed in the first study was not due to difficulties with the method.

Prior work has also not explored whether children cooperate more with remorseful transgressors. There is recent work showing that even 3-year-old children are discriminating cooperators and helpers. Children of this age reduce their helping toward harmful people and even toward people who intended to cause harm but were unable to do so (Vaish, Carpenter, & Tomasello, 2010). Moreover, 3- to 4-year-old children share more toys with a peer if that peer had previously shared toys with them (Levitt, Weber, Clark, & McDonnell, 1985), and when acting on behalf of a protagonist doll, they share more resources with dolls who have been generous to the protagonist doll or generous to another doll than with ungenerous dolls (Olson & Spelke, 2008). Given that a guilt display communicates that the transgressor is otherwise a reliable and cooperative group member, one may predict that children will cooperate more with a remorseful than an unremorseful transgressor.

Children in the present studies first watched videos of transgressors displaying guilt or no guilt about their transgressions (Study 1) or apologizing or not apologizing for their transgressions (Study 2) and then were asked a series of questions about the transgressors. The transgressors in the videos caused accidental rather than intentional harm. Although theories of guilt are unclear as to the role of intentionality (see, e.g., Baumeister, Stillwell, & Heatherington, 1994; Fessler & Haley, 2003), prior empirical work shows that adults expect an accidental transgressor to experience more guilt than an intentional transgressor (McGraw, 1987). Also, apologies are effective in increasing forgiveness only following accidental harm and actually seem to decrease forgiveness following intentional harm (Struthers, Eaton, Santelli, Uchiyama, & Shirvani, 2008). To ensure that the guilt displays in our study would be believable and effective, we therefore used accidental rather than intentional transgressions. We were reasonably sure that children would detect that the transgressions were accidental because the ability to distinguish between intentional and accidental actions is already present in infants by 9 months of age (Behne, Carpenter, Call, & Tomasello, 2005), and by around age 3 years, children can distinguish intentional from accidental transgressions (e.g., Nobes, Panagiotaki, & Pawson, 2009; Vaish et al., 2010).

Guilt seems to have no single, clear facial expression but rather is associated with a variety of remorseful facial and vocal expressions, confessions, acceptance of responsibility, statements that the harm was accidental, apologies, and expressions of the desire to repair, or actual attempts to repair (Berti et al., 2000; Fessler & Haley, 2003; Keltner, 1995; Keltner & Buswell, 1996; Zahn-Waxler & Kochanska, 1990). However, we chose to limit the guilt display in Study 1 to what might be the minimal components needed to clearly express guilt but without potential confounds. Thus, the guilt display in the present study consisted of a remorseful and concerned facial expression, confession, statement that the harm was accidental, and acceptance of responsibility. Apologies were excluded in Study 1 since we wanted to assess responses to guilt in the absence of apologies. Attempts to repair the damage also were excluded as these might bias children to see the remorseful transgressor as more prosocial but for reasons other than the guilt display. The guilt display in Study 1 was thus restricted to a few core features.

Thus, in Study 1, we assessed 4- and 5-year-old children's grasp of the appeasement functions of guilt displays by assessing their judgments of accidental transgressors who displayed guilt versus no guilt. On the basis of prior work (e.g., Smith et al., 2010; Wellman et al., 1979), we expected that at least by age 5 years, children would expect a victim to be madder at an unremorseful transgressor and to prefer a remorseful transgressor and that they would themselves prefer to interact with a remorseful transgressor. We also predicted that children would distribute more resources to the remorseful transgressor. Following Smith et al. (2010), we also asked children about the moral character of the transgressors (i.e., whether the remorseful or unremorseful transgressor is meaner), but we did not have specific predictions given Smith et al.'s findings that 4- to 5-year-olds failed to make appropriate inferences about the moral character of transgressors. Lastly, we also asked children to provide justifications for their judgments, as these can be useful in revealing the criteria on which children base their judgments (Gruneich, 1982; Turiel, 1998). In Study 2, we examined 4-year-old children's grasp of the appeasement functions of apologies using the same method as Study 1.

Study 1

Method

Participants. Participants were 4-year-old children ($N = 20$, 10 girls, 10 boys) who were between the ages of 48 months 25 days and 53 months 12 days ($M = 51$ months 7 days; $SD = 1$ month 12 days) and 5-year-old children ($N = 20$, 10 girls, 10 boys) who were between the ages of 60 months 1 day and 66 months 4 days ($M = 63$ months 1 day; $SD = 2$ months 8 days). Four additional children were tested but excluded due to experimenter error (two 5-year-olds), equipment failure (one 5-year-old), or unwillingness to participate (one 4-year-old). All children were native German speakers whose parents had given permission for them to participate in child development studies. Children were recruited from and tested in their day care centers in a medium-sized German city.

Design and materials. During the experiment, children sat at a table on which two laptop computers (PowerBook G4 computers [Apple Computer, Inc., Cupertino, CA] with 15-inch [38.1-cm], $1,440 \times 960$ -pixel screens) were placed next to one another, one to the left and one to the right of the child. All videos were played using the full-screen option in Quicktime Player. A camera recorded a frontal view of the children, and a microphone placed between the computers supplied sound to the camera. The procedure had two phases. In each phase, children saw one Guilt video and one No Guilt video, about which they received comprehension probe questions (as manipulation checks to make sure they grasped the content of the videos) and eight test questions. After the second phase (with a second set of Guilt and No Guilt videos), children received a distribution of resources task and one final test question about why they had distributed the resources in the way that they had. Thus, altogether, children watched four videos (two per phase) and answered 17 test questions (eight after each of the two phases and one after the distribution of resources task).

Video stimuli. Videos featuring three adult actresses (research assistants in the lab) served as stimuli. These videos featured a "transgressor" accidentally harming a "victim" and then

either displaying guilt or no guilt. One actress (Anya) always played the victim, and the other two actresses (Lisa and Susie) played the transgressors. Each video featured one target object: a doll, ball, clay bird, or picture.

All videos began with the three actresses seated around a table (see top panel of Figure 1). Anya, the future victim (sitting in the middle), excitedly told Lisa and Susie that she wanted to show them something, and then brought out and presented the target object for 45 s, as follows:

Doll. Anya said this was her favorite doll and then happily showed off the doll's hair, eyes, and so forth.

Ball. Anya said this was her new ball and then happily played with it by throwing it in the air, rolling it on the table, and so forth.

Bird. Anya brought out a clay bird and happily talked about how she had made it and how pretty it was. After 20 s, she proudly added a tail feather with some more clay. When she was done, she again stated how pretty the bird was.

Picture. Anya brought out a drawing of a butterfly and happily talked about how she had drawn it and how pretty it was. After 20 s, she proudly completed the drawing by adding the



Figure 1. The two camera angles used in both studies, at the start of the video stimuli (top panel) and during the transgressors' actions (bottom panel; here, the transgressor, on the right, is showing guilt). The individuals whose faces appear here gave signed consent for their likenesses to be published in this article.

antennae. When she was done, she again stated how pretty the butterfly was.

The first 15 s of each video showed all three actresses, after which the camera zoomed in on Anya so as to show the target object more clearly (the transgressors were now out of view). Toward the end of her 45-s presentation, the camera angle switched to a view of Anya and the transgressor assigned to that situation (the other transgressor remained out of view; see bottom panel of Figure 1). Anya now happily placed the target object on the table, and the transgressor then acted upon the object.

For the sake of simplicity, Lisa was always the transgressor in the doll and bird situations and Susie in the ball and picture situations. The transgressions thus proceeded as follows:

Doll. Lisa picked up the doll to admire it, but while she was playing with the doll's hair, its head came off and fell onto the table.

Ball. Susie played with the ball by throwing it up and catching it. The third time that she did this, she failed to catch the ball, and it fell to the ground. When she reached under the table to retrieve it (out of view), the sound of something tearing was audible, and when she brought the ball back up, it was torn and the filling was spilling out.

Bird. Lisa picked up the clay bird to admire it, but as she touched one of the wings, it became detached and the rest of the bird fell onto the table.

Picture. Susie admired the picture but while returning it to Anya, she accidentally tore it.

At the end of each incident, Anya said sadly, "Oh, my [target object]," and the transgressor then responded guiltily or not guiltily. In the Guilt condition, the transgressor looked remorseful and concerned (with furrowed brow and concerned eyes; cf. Vaish, Carpenter, & Tomasello, 2009; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992) and remorsefully said, "Oh, I've torn/broken your [target object]. I didn't want that to happen. It's my fault." (German: "Oh, ich habe dein [target object] kaputt gemacht. Das wollte ich nicht. Das ist meine Schuld.") While speaking, she alternated her gaze between Anya and the broken object. She then continued to look remorseful as she placed the object on the table. In the No Guilt condition, the transgressor looked neutral and said in a neutral tone of voice, "Yes, I've torn/broken your [target object]. Hmph [she shrugged], I don't care." (German: "Ja, ich habe dein [target object] kaputt gemacht. Hmph, das ist mir egal.") She also alternated her gaze between Anya and the broken object while speaking, and then neutrally placed the object on the table. In both the Guilt and the No Guilt condition, Anya now picked up the object and looked at it sadly while the transgressor continued looking remorseful or neutral. The video ended with a still frame of this scene, which remained on the screen for 6 s. The average duration of each video was 2 min. There were also shorter versions of all videos (starting just before Anya placed the object on the table) that ended with the same still frames.

Counterbalancing. For each target object, we created four videos in which the sides that the transgressors sat on and whether they showed guilt or no guilt were fully counterbalanced. For example, there were four videos of the doll situation: Lisa on the left and showing guilt or no guilt, and on the right and showing guilt or no guilt. There were thus 16 videos in all (four per target object), although each child only watched four of the 16 videos

(one per target object). During testing, the doll and ball videos (both featuring the victim's possessions) were always presented together, as were the bird and picture videos (both featuring objects that the victim had created).

Children were randomly assigned to one of 20 presentation orders which counterbalanced the sides on which transgressors sat, which transgressor acted first, which transgressor showed guilt, and whether children saw the bird and picture or the ball and doll video pair first. Other factors that were counterbalanced will be mentioned later. The computer on the left always showed the situations involving the transgressor on the left, and the computer on the right always showed the situations involving the transgressor on the right.

Procedure. All children were tested by the same female adult experimenter (E), who always sat to their left during the experiment (and who did not feature in the videos). E told children that she was going to show them videos of some people doing some things and that they should watch carefully and she then would ask them some questions. E opened the first video assigned to the child, introduced the three characters on the still opening frame, and played the video (e.g., a Guilt video of the ball situation on the left computer). At the end of the video, E paused the still frame and asked the first comprehension probe: "What did Lisa/Susie do to Anya's [target object]?" The child was expected to answer, "She broke it," or something similar. If the child answered with something less specific (e.g., "She dropped the ball"), E prompted the child further by asking, for example, "And what happened to the ball?"

The first probe was to ensure that the child understood that the transgressor had damaged the object. Once the child's response indicated that he or she understood this, E said, "That's right," and asked the second comprehension probe: "How does she [pointing to transgressor] feel now? Does she feel bad or does she not feel bad?" (Order of "feel bad" and "not feel bad" was counterbalanced across children). This second probe was to ensure that the child grasped how the transgressor felt, which was critical if the child was going to draw any inferences on this basis. If the child answered correctly ("Bad" or something similar in the Guilt case; "Not bad" or something similar in the No Guilt case), E said, "That's right. You've understood it correctly. Let's watch that last part again." E then played the shorter version of the video and paused it on the final still frame. If, however, the child answered the second probe incorrectly (e.g., "Bad" in the No Guilt case), E said, "Hmm, I'm not so sure about that. Let's watch that last part again, and I'll ask you the questions again afterwards." E then played the shorter version of the video, paused it on the still frame, and repeated both comprehension probes as before. If the child still answered the second probe incorrectly, E corrected her by saying, "No, she felt/did not feel bad about breaking/tearing the [target object]." (Thus, regardless of whether or not children correctly answered the comprehension probes, all children saw each video entirely once and partially once.)

E then opened the second video, which was in the other condition and on the other computer (e.g., a No Guilt video of the doll situation on the right computer). She reminded the child of the characters' names and then followed the same procedure as with the first video. Finally, after the child had seen both videos and answered the comprehension probes, E provided a reminder, for example: "So, Lisa/Susie [pointing to corresponding computer

screen] broke Anya's [target object] and she feels bad about it, and Susie/Lisa broke Anya's [target object] and she doesn't feel bad about it" (always starting with the first video children had seen in that phase). While providing this reminder (and throughout the procedure), E was careful to speak neutrally and not to nod or shake her head or in any other way provide evaluations of the transgressors. E then asked the following test questions:

1. Victim madder: "Whom is Anya madder at? Susie or Lisa?" (pointing to each in turn)
 - 1a. Victim madder-justification: "Why is she madder at her?"
2. Victim likes: "Whom does Anya like more? Susie or Lisa?" (pointing to each)
 - 2a. Victim likes-justification: "Why does she like her more?"
3. Child plays: "Whom would you prefer to play with? Susie or Lisa?" (pointing to each)
 - 3a. Child plays-justification: "Why would you like to play with her more?"
4. Meaner: "Who do you think is meaner? Susie or Lisa?" (pointing to each)
 - 4a. Meaner-justification: "Why do you think she is meaner?"

Questions 1, 2, 3, and 4 were forced-choice questions because our aim was to assess whether when presented with the choice, children would be able to use the information about the transgressors' guilt displays to respond in the hypothesized ways. In response to these forced-choice questions, children were expected to name and/or point to one transgressor. If a child responded "Both" or "Neither," E prompted her to choose one (this never occurred among the 5-year-olds and occurred only three times among the 4-year-olds: one child said "Both" once, and two children said "Neither" once each). If a child did not respond at all, E repeated the question once, but if the child still did not respond, E moved on to the next question. Questions 1a, 2a, 3a, and 4a were designed to elicit justifications for children's responses to the forced-choice questions. E thus let children respond freely to these questions and did not probe further. Note that E did not provide any feedback on the correctness of children's responses for any of the test questions. Following these eight test questions, E repeated the entire procedure and all of the questions with the second pair of videos (Phase 2).

For a given child, Phases 1 and 2 were matched in terms of which transgressor showed guilt and on which side each transgressor sat. However, the order in which the transgressors' names appeared in the test questions and the order of the first three pairs of questions were counterbalanced across children and across the two phases for a given child. The fourth pair (the "meaner" questions, similar to the moral character questions used by Smith et al., 2010) appeared either first or last (counterbalanced across children). This was because although it was feasible that children would not draw appropriate inferences about the transgressors' moral character (Smith et al., 2010), if they did, it would be important to know that they were not influenced by their responses to the preceding questions.

Finally, after the second phase, E said that she would see Lisa and Susie soon and could bring them something from the child. Then, in front of each computer, E placed a small container holding a photograph of the transgressor featured on the corresponding computer (the photographs featured the transgressors seated at the table and looking neutrally at the camera). E then gave the child three cloth flowers to distribute as he or she wanted.

If the child did not distribute all the flowers or asked E for guidance, E encouraged him or her to decide for himself or herself. When the child was done, E asked one final justification question, namely, why the child had given two (or three) flowers to Lisa/Susie (whoever received more flowers). Again, E let children respond freely and did not probe further.

Coding and reliability. A transcriber (blind to hypotheses) first transcribed children's verbal and/or pointing responses. From these transcriptions, the primary coder coded whether children responded correctly to the comprehension probes. Since E asked the second comprehension probe only after children answered the first probe satisfactorily, coding of responses to the first probe was only to make sure that E had followed this procedure and thus that all children understood that the target object was damaged. Coding of responses to the second probe assessed whether children grasped right away how the transgressor felt, whether they grasped it after watching the video again, or whether E had to correct them explicitly. For reliability, a second coder (blind to hypotheses) coded responses to the second comprehension probe for a random 25% of the sample. Reliability was perfect, $\kappa = 1$.

The primary coder also used the transcriptions to code children's responses to the forced-choice test questions (Questions 1, 2, 3, and 4). Responses were scored 1 if they were consistent with the hypotheses that children should (a) judge that the victim is madder at the unremorseful transgressor, (b) judge that the victim prefers the remorseful transgressor, (c) indicate that they themselves prefer to interact (play) with the remorseful transgressor, and (d) judge the unremorseful transgressor to have a worse moral character (i.e., to be meaner); responses not consistent with these hypotheses were scored 0. A second coder (blind to hypotheses) coded a random 25% of the sample. Reliability was perfect, $\kappa = 1$.

Children's distribution of the three flowers was coded from videotape and scored 0, 1, 2, or 3 to represent how many flowers children gave to the transgressor who displayed guilt. A second coder coded this for a random 25% of the sample. Agreement between coders was 100%.

Finally, children's justifications (i.e., their responses to Test Questions 1a, 2a, 3a, and 4a, and their justification for the distribution of resources task) were coded from the transcriptions and assigned scores of either 1 or 0 (see Table 1 for details of the coding scheme). A score of 1 was assigned to justifications that indicated relevant and sophisticated reasoning about the transgressors and their responses. These included justifications that referred to apologies or feelings of guilt, or involved moral evaluations. References to apologies were included in this category because, in fact, neither transgressor apologized. Inferring an apology from a guilt display or no apology from the absence of a guilt display thus required an impressive grasp of the emotions behind apologies. Note also that references to a transgressor's feelings of guilt (or the lack thereof) could be of two kinds: One kind—"Guilt (repeated)"—involved repeating phrases that had been used in the videos or by E (e.g., "Because she always said, 'Oh, I've broken your bird; I don't care'"), whereas the other kind—"Guilt (re-described)"—involved using phrases other than those used in the videos or by E (e.g., "Because she tore the picture and then did not feel sorry for it"). Although "Guilt (repeated)" was clearly relevant to our question of whether children are sensitive to displays of guilt, when children repeated the word or phrase used in the videos

Table 1
Coding Scheme for Justifications: Study 1

Score/category	Content
Score of 1	
Apology	Transgressor did (or did not) apologize; for example, "Because she said sorry."
Guilt (re-described)	Transgressor did (or did not) feel bad about what she had done (child uses words other than those used in the videos or by adult experimenter E; for example, "Because she tore the picture and then did not feel sorry for it.")
Moral character, evaluation, or norm	Transgressor is a good (or bad) person, transgressor's response to her act was good (or bad), or transgressor broke (or did not break) a moral norm; for example, "Because when one tears a picture, that's mean, and then on top of that to not care about it."
Guilt (repeated) ^a	Transgressor did (or did not) feel bad about what she had done (child uses words that had been used in the videos or by E); for example, "Because she always said, 'Oh, I've broken your bird; I don't care.'"
Score of 0	
Intention	Transgressor intended (or did not intend) to cause harm or to damage the object; for example, "She didn't do it on purpose."
Victim	Victim is sad or upset, or her belonging has been destroyed.
Own preference	Child expresses his or her own preference for the transgressor; for example, "Because I like her better."
Action	Transgressor damaged the [target object].
Object	[Target object] is damaged or can no longer be repaired.
Other, irrelevant, or not able to be coded	Response could not be put into any of the above categories (e.g., "Because that's how it is"), was irrelevant (e.g., "Because I always go to the zoo with my mother"), or could not be coded (e.g., because the child's speech could not be understood).

^a Analyses were conducted with this category scored as 1 and as 0.

or by E, there was no way of knowing whether they were engaging in higher level reasoning or not. That is, did they in fact understand the transgressors' reactions in a sophisticated way but nevertheless repeat the words they heard before, or did they not understand the transgressors' reactions in a sophisticated way and thus simply repeat the words they heard before? To account for both possibilities, we conducted two sets of analyses of children's justifications, one in which "Guilt (repeated)" was assigned a score of 1, and the other in which it was assigned a score of 0. Furthermore, note that "Guilt (re-described)" was assigned a score of 1 only if, in addition to re-describing feelings of guilt (or the lack thereof), the justification also linked the guilt to the transgression (see earlier example); simply re-describing the guilt feelings without this link received a score of 0.

A score of 0 was assigned to all other justifications, including justifications that indicated that children had understood what had happened but that were not diagnostic; that is, they did not set one of the transgressors apart from the other. For instance, justifications referring to a transgressor's intentions received a score of 0 since the transgressors' intentions were the same across scenarios; similarly, stating that Anya liked one of the transgressors less because that transgressor had broken her object received a score of 0 since an object was broken in all scenarios. If a child did not provide a justification on a particular question, no score was assigned for that question. A second coder (blind to hypotheses) coded justifications of a random 25% of children. Reliability was excellent, $\kappa = .82$.

Results

We first report results of the comprehension probes in order to provide information about how well children understood the content of the videos. We then report children's performance on test questions and the distribution of resources task. Preliminary analyses revealed that for both age groups, there were no significant effects of the side on which Lisa and Susie sat, whether Lisa or Susie acted first, whether Lisa or Susie showed guilt, or whether a Guilt or No Guilt video was presented first. There were also no significant effects of gender but one: Among 4-year-olds, in Phase 1, nine of 10 boys, but only three of nine girls, chose the remorseful transgressor in response to the test question "Whom would you prefer to play with?" ($p = .01$). However, as this effect did not emerge in Phase 2, and as no other gender effects emerged, gender was pooled for all subsequent analyses. All reported p values are two-tailed.

Comprehension probes.

Comprehension Probe 1. For all four videos, responses to the first comprehension probe (about what the transgressor had done) indicated that all children at both ages understood that the transgressors had damaged the target objects. Thus, all children grasped the basic premise of the videos.

Comprehension Probe 2.

5-year-olds. In Phase 1, when the transgressor showed guilt, 16 of 20 children in the 5-year-old group correctly identified the transgressor as feeling bad right away (binomial probability, using

a test proportion of .50, $p = .012$). The remaining four children responded correctly after rewatching the video. When the transgressor showed no guilt, 11 of 20 children in the 5-year-old group correctly identified the transgressor as not feeling bad right away (binomial probability, $p = .824$). Eight of the nine remaining children responded correctly after rewatching the video. Children's performance improved in Phase 2: When the transgressor displayed guilt, all 20 children correctly identified her right away as feeling bad, and when the transgressor showed no guilt, 17 of 20 children (binomial probability, $p = .003$) correctly identified her right away as not feeling bad, and the other three responded correctly after re-watching the video. Altogether, in the Guilt case, 16 of 20 children responded correctly right away in both phases, and in the No Guilt case, 11 of 20 did so.

4-year-olds. The younger children struggled to grasp how the transgressors felt, especially when the transgressor was unremorseful. In Phase 1, when the transgressor showed guilt, the majority of 4-year-olds (14 of 20) correctly identified her as feeling bad right away (binomial probability, $p = .115$). Of the remaining six children, five responded correctly after rewatching the video. However, when the transgressor showed no guilt, a significant majority (16 of 20) initially responded *incorrectly* (binomial probability, $p = .012$). Of those 16, eight responded correctly after rewatching the video, but eight still responded incorrectly and had to be corrected by E. Again, performance improved in Phase 2: When the transgressor showed guilt, 16 of 20 children correctly identified the transgressor right away as feeling bad (binomial probability, $p = .012$), and three of the remaining four responded correctly after rewatching the video. When the transgressor showed no guilt, nine of 20 children responded correctly right away (binomial probability, $p = .824$), and seven of the remaining 11 responded correctly after rewatching the video. Altogether, in the Guilt case, 13 of 20 children responded correctly right away in both phases, but in the No Guilt case, only two of 20 children did so.

Age comparisons. The 4- and 5-year-olds' initial responses to the second comprehension probe (pooled across Phases 1 and 2) were analyzed using chi-square tests. When the transgressor showed guilt, the number of children who responded correctly in both phases did not differ by age group ($p = .288$). However, when the transgressor showed no guilt, significantly more 5-year-olds than 4-year-olds responded correctly in both phases, $\chi^2(1, N = 40) = 9.23, p = .002$. The fact that when the transgressor showed no guilt, a large number of children at both ages nevertheless responded that the transgressor felt bad points to the importance of checking children's comprehension and, when necessary, correcting them about the transgressors' feelings before moving on to test their understanding of the appeasement functions of guilt.

Test questions.

Forced-choice questions. Preliminary analyses revealed no significant effects for either age group of whether or not children answered the comprehension probes correctly right away or of whether the "meaner" questions (Questions 4 and 4a) appeared first or last. These variables were thus not included in further analyses.

5-year-olds. The performance of the 5-year-olds was impressive: Children's responses to the eight forced-choice test questions (four per phase) indicate that they overwhelmingly drew all of

the appropriate, hypothesized inferences in both the first and the second phase (the proportion of children who responded in the predicted way on each question ranged from .80 to .95; binomial probabilities, all $ps < .013$). For further analyses, we pooled children's responses from both phases for each question. For example, responses to the "Victim madder" question in Phase 1 and in Phase 2 were pooled, and a score of 1 was assigned for the "Victim madder" question if both responses drew the appropriate, hypothesized inference; otherwise, a score of 0 was assigned. These more stringent analyses also indicated that the 5-year-olds drew all of the appropriate, hypothesized inferences (binomial probabilities, all $ps < .013$; see Figure 2). Furthermore, 19 of 20 children responded in the hypothesized way to more than half of the forced-choice questions (binomial probability, $p < .0005$).

4-year-olds. In stark contrast to the 5-year-olds, the younger children's responses to the eight forced-choice test questions indicate that they did not draw any of the hypothesized inferences in either phase (the proportion of children who responded in the predicted way on each question ranged from .35 to .65; binomial probabilities, all $ps > .262$). Pooling children's responses across phases for each question produced similar results (binomial probabilities, all $ps > .166$; see Figure 2). Altogether, only seven of 20 children responded in the hypothesized way to more than half of the forced-choice questions (binomial probability, $p = .263$).

Age comparisons. We conducted a chi-square analysis using pooled responses from both phases for each question. This analysis revealed that on all forced-choice test questions, significantly more 5- than 4-year-olds drew the hypothesized inference in both phases (all $ps < .005$; see Figure 2). Moreover, the proportion of the eight test questions answered in the hypothesized way was significantly higher among 5-year-olds ($M = .91, SD = .20$) than 4-year-olds ($M = .51, SD = .28$), $t(38) = 5.10, p < .0005$.

Justifications. Children's justifications were compared across the two age groups. Justifications were only included in analyses if children had answered the preceding forced-choice test question in the hypothesized way (but note that including all justifications did not change the pattern of any of the results). Furthermore, since assigning "Guilt (repeated)" a score of 1 versus 0 did not change the pattern of any of the results, we report analyses in which it was assigned a score of 1 so as to provide information about all children who provided relevant and appropriate responses (Table 2 provides the numbers obtained with both methods of coding). The majority of 5-year-olds (15 of 20) provided at least one higher level (score of 1) justification across all eight justification questions, indicating a sophisticated level of understanding and reasoning about the transgressors and their responses. However, very few 4-year-olds (only two of 18) provided any Level-1 justifications across the eight justification questions. The difference between the number of 5- versus 4-year-olds who provided at least one Level-1 justification was thus highly significant, $\chi^2(1, N = 38) = 15.64, p < .0005$.

Distribution of resources.

Distribution.

5-year-olds. Of the 20 children in the 5-year-old group, the first three did not receive the distribution of resources task. Of the remaining 17 children, 16 gave more flowers (two or three out of three) to the guilt-displaying transgressor (binomial probability, $p < .0005$).

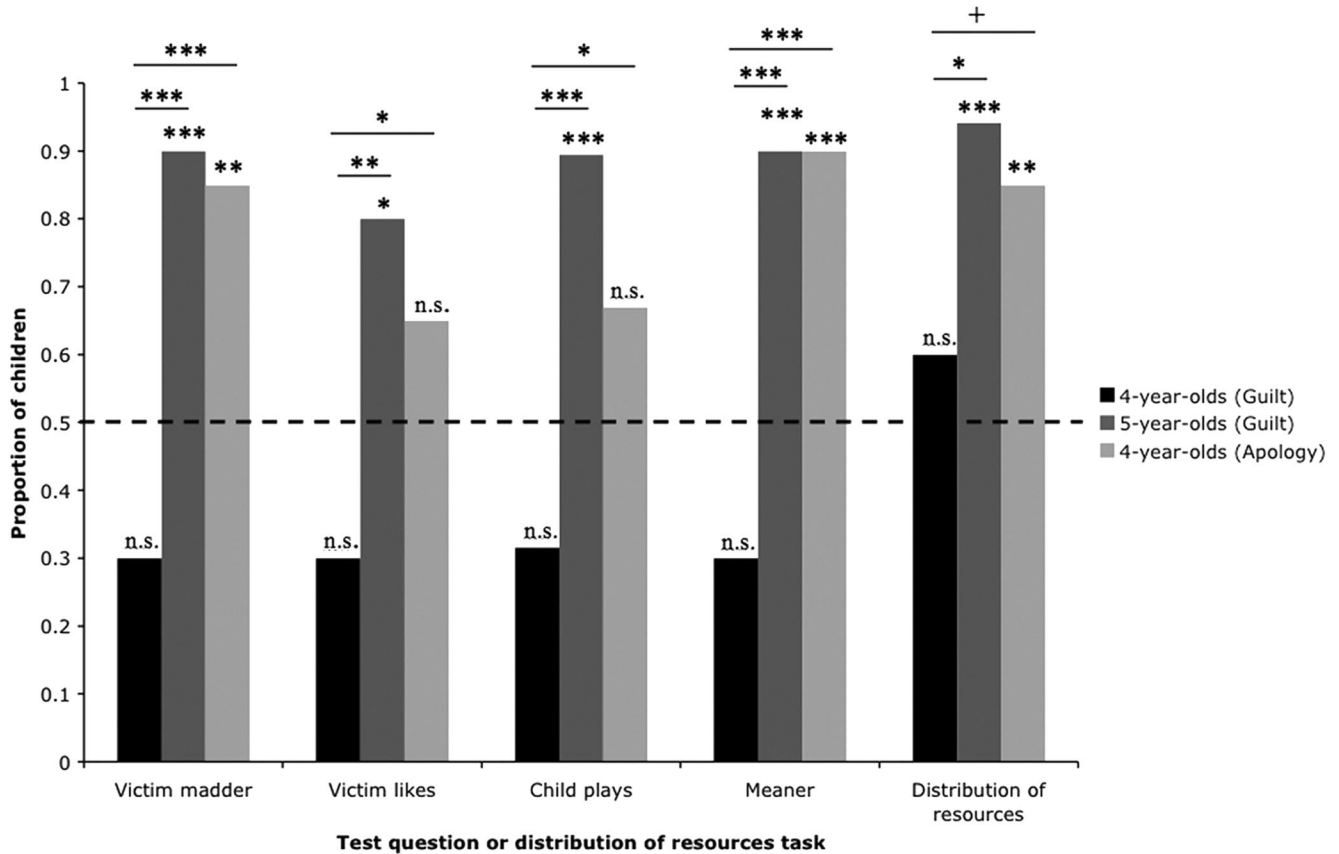


Figure 2. Proportion of children in Study 1 (Guilt) and Study 2 (Apology) who answered each test question and distributed resources in the predicted ways in both phases. The dashed line indicates chance level. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .0005$.

4-year-olds. A majority of 4-year-olds (12 of 20) gave more flowers to the guilt-displaying transgressor, but this proportion was not different from chance (binomial probability, $p = .503$).

Age comparisons. A chi-square analysis revealed that a greater proportion of 5- than 4-year-olds gave more flowers to the

remorseful person, Fisher's exact test (due to small N s in some cells), $p = .023$. The mean number of flowers given to the remorseful transgressor was also higher among 5-year-olds ($M = 2.06$; $SD = 0.43$) than 4-year-olds ($M = 1.55$, $SD = 0.61$), $t(35) = 2.90$, $p = .005$.

Table 2
Children's Justifications in Study 1 and Study 2

Variable	5-year-olds		4-year-olds			
	Study 1: Guilt Score		Study 1: Guilt Score		Study 2: Apology Score	
	1	0	1	0	1	0
"Guilt (repeated)" in Study 1 & "Apology" in Study 2 included ^a						
Test questions	15	5	2	16	15	5
No. of children in Phase 1; Phase 2	11; 15	9; 4	2; 1	13; 15	14; 13	5; 6
Distribution of resources	9	6	4	8	13	3
"Guilt (repeated)" in Study 1 & "Apology" in Study 2 excluded ^b						
Test questions	14	6	2	16	5	15
No. of children in Phase 1; Phase 2	10; 13	10; 6	2; 1	13; 15	2; 4	17; 15
Distribution of resources	9	6	4	8	8	8

Note. The values represent number of children who provided at least one higher level justification (score of 1) versus none (score of 0).
^a "Guilt (repeated)" responses in Study 1 & "Apology" responses in Study 2 were coded as scoring 1. ^b "Guilt (repeated)" responses in Study 1 & "Apology" responses in Study 2 were coded as scoring 0.

Justifications. Levels of children's justifications were compared across the two age groups. Children's justifications were included in analyses only if children had distributed the resources as predicted (i.e., given more flowers to the remorseful person), although the pattern of results remained the same when all justifications were included. Furthermore, assigning "Guilt (repeated)" a score of 1 versus a score of 0 did not change the pattern of results; we will thus report only analyses in which it was assigned a score of 1 (but see Table 2 for numbers obtained with both coding methods). Of the 16 children in the 5-year-old group who distributed the resources as predicted, one child was not asked the corresponding justification question due to experimenter error. Nine of the 15 remaining children provided Level-1 justifications. Among the 12 children in the 4-year-old group who responded as predicted on the distribution of resources task, four provided Level-1 justifications. Unlike the case with the justifications for the test questions, the two age groups did not differ in the number of children who provided Level-1 justifications on the distribution of resources task ($p = .168$).

Discussion

The results of Study 1 show that 5-year-olds robustly understand the appeasement functions of guilt displays: They know that the victim will be more upset with a transgressor who displayed no guilt than with one who did and, conversely, that the victim will like a transgressor who displayed guilt more than one who did not. They also robustly say that they themselves would prefer to affiliate (in the form of play) with a remorseful transgressor than with an unremorseful one. Moreover, contrary to Smith et al. (2010), who found that 4- to 5-year-olds were not able to make inferences about the moral characters of apologetic versus non-apologetic transgressors, our results showed that 5-year-olds were able to make such inferences. It is important to note that 4-year-olds did not draw any of these inferences. Finally, 5-year-olds, but not 4-year-olds, showed a strong preference to cooperate with (by distributing resources to) the remorseful than with the unremorseful transgressor. Our results thus suggest that between the ages of 4 and 5 years, children become sensitive to the appeasement functions of displaying guilt.

Note that the actors in the videos did not apologize, so children in our study could not have relied on apologies to draw their inferences. Thus, 5-year-olds seem to truly grasp the functions of displaying guilt rather than simply relying on whether an apology script was followed or not. Children's justifications also provide evidence for this idea, as several 5-year-olds stated that the guilt-displaying transgressor had apologized although in fact she had not offered an explicit apology. This suggests that at least some 5-year-olds grasp the true meaning behind apologizing, namely, to convey guilt and the appeasing information associated with it. The results of Study 1 suggest that 4-year-olds, on the other hand, seem not to grasp this.

However, it is also possible that the method used in Study 1 proved too challenging for 4-year-olds. After all, the method did rely heavily on children's verbal (comprehension and production) abilities, which may have posed a problem for the younger children. Additionally, perhaps the 4-year-olds found it difficult to process two complex social interactions and to keep both in mind long enough to answer questions about them. To test whether the

method proved too complex for 4-year-olds, we conducted a second study in which the method was kept the same as in Study 1 but one transgressor explicitly apologized (rather than displaying guilt without apologizing, as in Study 1). As mentioned previously, 4-year-olds seem to understand and draw at least some inferences from others' apologies: They judge stories in which a transgressor apologizes as better and more just than stories in which a transgressor does not apologize (Irwin & Moore, 1971; Wellman et al., 1979), and they attribute negative feelings to an apologetic transgressor and improved feelings to a victim who has received an apology (Smith et al., 2010). Our reasoning was thus that if 4-year-olds only had difficulty with the procedural demands in Study 1, they should perform similarly with this procedure regardless of whether they are presented with explicit apologies (Study 2) or guilt displays without apologies (Study 1). However, if their difficulty in Study 1 lay in understanding the emotions behind apologies and if their apparent understanding of apologies in fact rests rather superficially on the use of key words such as "sorry" or "apologize," then they should perform significantly better when presented with explicit apologies (Study 2) than when presented only with the emotions behind apologies (Study 1).

Study 2

Method

Participants. Participants were 4-year-old children ($N = 20$, 10 girls, 10 boys) who were between the ages of 49 months 10 days and 54 months 8 days ($M = 51$ months 29 days; $SD = 1$ month 16 days). One additional child was excluded due to an unwillingness to participate. Children were recruited and tested as in Study 1.

Design, materials, and video stimuli. The design and materials of Study 2 were identical to those in Study 1 except that children took part in Apology and No Apology conditions (rather than Guilt and No Guilt conditions). For the No Apology condition, we simply used the No Guilt videos of Study 1, but we created new videos for the Apology condition. These featured the same three actresses and target objects and were exactly like their Guilt counterparts in Study 1, the sole difference being that upon breaking the target object, the transgressor remorsefully said, "Oh, I've torn/broken your [target object]. I'm sorry. I apologize." (German: "Ich habe dein [target object] kaputt gemacht. Das tut mir Leid. Entschuldigung.") Counterbalancing and presentation of videos were exactly as in Study 1.

Procedure. The procedure of Study 2 was identical to that in Study 1 except that (a) the second comprehension probe was changed to "Did she [pointing to transgressor] apologize for what she did, or did she not apologize for what she did?" (order of "apologize" and "not apologize" counterbalanced across children) and (b) the reminder that E provided prior to asking the test questions was changed to, for example, "So Lisa/Susie [pointing to corresponding computer screen] broke Anya's [target object], and she apologized for it, and Susie/Lisa broke Anya's [target object], and she did not apologize for it." Note that as in Study 1, in response to the forced-choice test questions, children rarely responded "Both" (one child in response to one question) or "Neither" (two children, one in response to two questions and one in response to one).

Coding and reliability. Children's responses to the comprehension probes, forced-choice test questions, and distribution of resources task were coded as in Study 1. For reliability, a second coder (blind to hypotheses) coded children's responses for a random 25% of the sample. Reliability was perfect for all three measures: $\kappa = 1$ for the second comprehension probe as well as the forced-choice test questions, and 100% agreement for the distribution of the three flowers.

Children's justifications were again assigned scores of either 1 or 0. A score of 1 was assigned to justifications that referred to apologies or feelings of guilt or involved moral evaluations. However, note that in this study, justifications involving apologies could result from repeating phrases that had been used in the videos or by E (e.g., "Because she apologized" or "She said sorry") rather than from a sophisticated understanding or analysis of the transgressors' feelings or motivations for apologizing. That is, children who repeated apology-related phrases used in the videos or by E might in fact have understood the transgressors' reactions in a sophisticated way but nevertheless used the phrases that they heard before, or they may not have understood the transgressors' reactions in a sophisticated way and thus simply repeated the words that they heard before. In order to consider both possibilities, we conducted two sets of analyses of children's justifications, one in which "Apology" was assigned a score of 1, and the other in which it was assigned a score of 0. A score of 0 was assigned to all other justifications (see Table 3 for details of the coding scheme). A second coder (blind to hypotheses) coded justifications of a random 25% of children. Reliability was excellent, $\kappa = .94$.

Results

We will report results in the same order as for Study 1. Preliminary analyses revealed no significant effects of any of the counterbalanced variables or of gender. All reported p values are two-tailed.

Comprehension probes.

Comprehension Probe 1. For all four videos, responses to the first comprehension probe indicated that all children understood that the transgressors had damaged the target objects. Thus, all children grasped the basic premise of the videos.

Comprehension Probe 2. When the transgressor apologized, 18 of 20 children in Phase 1 and the same number in Phase 2 correctly identified this right away (binomial probabilities, $p < .0005$). In both phases, one of the remaining two children responded correctly after rewatching the video. When the transgressor did not apologize, 16 of 20 children in Phase 1 and 19 of 20 children in Phase 2 correctly identified this right away (binomial probabilities, $p < .0005$). In Phase 1, three of the four remaining children responded correctly after rewatching the video, and in Phase 2, the one remaining child responded correctly after rewatching the video. Altogether, in the Apology case, 17 of 20 children responded correctly right away in both phases (binomial probability, $p = .003$), and in the No Apology case, 16 of 20 did so (binomial probability, $p = .012$).

Age comparisons (across studies). We conducted chi-square tests to compare children's initial responses to the second comprehension probe (pooled across Phases 1 and 2) across studies and

Table 3
Coding Scheme for Justifications: Study 2

Score/category	Content
Score of 1	
Guilt	Transgressor did (or did not) feel bad about what she had done.
Moral character, evaluation, or norm	Transgressor is a good (or bad) person, transgressor's response to her act was good (or bad), or transgressor broke (or did not break) a moral norm; for example, "Because when one tears a picture, that's mean, and then on top of that to not care about it."
Apology ^a	Transgressor did (or did not) apologize.
Score of 0	
Intention	Transgressor intended (or did not intend) to cause harm or to damage the object; for example, "She didn't do it on purpose."
Victim	Victim is sad or upset, or her belonging has been destroyed.
Own preference	Child expresses his or her own preference for the transgressor; for example, "Because I like her better."
Action	Transgressor damaged the [target object].
Object	[Target object] is damaged or can no longer be repaired.
Other, irrelevant, or not able to be coded	Response could not be put into any of the above categories (e.g., "Because that's how it is"), was irrelevant (e.g., "Because I always go to the zoo with my mother"), or not be coded (e.g., because the child's speech could not be understood).

^a Analyses were conducted with this category scored as 1 and as 0.

ages. The proportion of 4-year-olds in Study 2 who responded correctly when the transgressor apologized did not differ from the proportion of 4-year-olds in Study 1 ($p = .144$) and the proportion of 5-year-olds in Study 1 ($p = 1.00$) who responded correctly when the transgressor showed guilt. Also, the proportion of 4-year-olds in Study 2 who correctly identified when the transgressor did not apologize did not differ from the proportion of 5-year-olds in Study 1 who correctly identified when the transgressor did not show guilt ($p = .091$), but the proportion of 4-year-olds in Study 2 who did so was significantly greater than the proportion of 4-year-olds in Study 1, $\chi^2(1, N = 40) = 19.80, p < .0005$. Thus, 4-year-olds in Study 2 clearly grasped whether or not the transgressors had apologized, and were substantially better at identifying when a transgressor had not apologized than were their counterparts in Study 1 at identifying when a transgressor did not feel bad.

Test questions.

Forced-choice questions. Preliminary analyses revealed no significant effects of whether or not children answered the comprehension probes correctly right away or of whether the “meaner” questions (Questions 4 and 4a) appeared first or last. These variables were thus not included in further analyses.

Children’s responses to the eight forced-choice test questions (four per phase) indicate that, as predicted, they drew nearly all of the appropriate, hypothesized inferences in both the first and the second phases (the proportion of children who responded in the predicted way on each question ranged from .72 to .95; binomial probabilities, all $ps < .013$ except the “Child plays” question in Phase 2, for which $p = .096$). For further analyses, we pooled children’s responses from both phases for each question (as in Study 1). These more stringent analyses indicated that a significant majority of children responded in the hypothesized way about whom the victim would be madder at (binomial probability, $p = .003$) and about which of the transgressors was “meaner” (binomial probability, $p < .0005$). A majority of children also responded in the hypothesized way about whom the victim would like more (13 of 20 children) and about whom children would prefer to play with (12 of 18 children), but these proportions were not significantly above chance (binomial probabilities, both $ps > .237$; see Figure 2). Nevertheless, a significant majority of children (18 of 20) responded in the hypothesized way to more than half of the eight forced-choice questions (binomial probability, $p < .0005$).

To compare children’s performance across studies and ages, we conducted chi-square analyses using pooled responses from both phases for each question in order to compare children’s performance across studies and ages. This analysis revealed that on all forced-choice test questions, there were no significant differences in the numbers of 4-year-olds in Study 2 versus 5-year-olds in Study 1 who responded in the hypothesized ways (all $ps > .123$). There was also no significant difference between these two groups in the proportion of the eight test questions answered in the hypothesized ways ($p = .449$). On the other hand, significantly more 4-year-olds in Study 2 than their counterparts in Study 1 drew the hypothesized inferences in both phases (all $ps < .034$; see Figure 2). Moreover, the proportion of the eight test questions answered in the hypothesized way was significantly higher among 4-year-olds in Study 2 ($M = .86; SD = .19$) than 4-year-olds in Study 1 ($M = .51, SD = .28$), $t(38) = 4.59, p < .0005$. Thus,

4-year-olds’ inferences about transgressors who did (not) apologize (Study 2) were more sophisticated and in keeping with predictions than 4-year-olds’ inferences about transgressors who did (not) display guilt (Study 1).

Justifications. As in Study 1, justifications were only included in analyses if children had answered the preceding forced-choice test question in the hypothesized way (but note that including all justifications did not change any of the results). Unlike in Study 2, however, assigning “Apology” a score of 1 versus a score of 0 did change the results. We will thus report analyses using both scoring methods (see also Table 2). When “Apology” was assigned a score of 0, only a few children (five of 20) were scored as providing at least one higher level (score of 1) justification across all eight justification questions. Chi-square analyses indicated that this proportion was significantly lower than the proportion of 5-year-olds in Study 1 who provided at least one higher level justification when “Guilt (repeated)” was scored 0 ($p = .004$). On the other hand, this proportion was not different from the proportion of 4-year-olds in Study 1 who provided at least one higher level justification when “Guilt (repeated)” was scored 0 ($p = .410$).

The results were very different when “Apology” was assigned a score of 1. In this case, a majority of 4-year-olds in Study 2 (15 of 20) were scored as providing at least one higher level (score of 1) justification across all eight justification questions. Chi-square analyses indicated that this proportion was not different from the proportion of 5-year-olds in Study 1 who provided at least one such justification when “Guilt (repeated)” was scored 1 ($p = 1.000$), but was significantly greater than the proportion of 4-year-olds in Study 1 who provided at least one such justification when “Guilt (repeated)” was scored 1 ($p < .0005$). Thus, with a stringent coding system (in which repeating phrases that were used during the presentations did not receive a higher score despite being relevant and appropriate), 4-year-olds in both studies provided similar levels of justifications, which were significantly lower than those provided by 5-year-olds in Study 1. However, with a more liberal coding system (in which repeating relevant phrases that were used during the presentations received a higher score), 4-year-olds were seen to provide sophisticated and appropriate justifications for their responses on the forced-choice test questions in the context of transgressors apologizing or not apologizing (Study 2), but were still not seen to do so in the context of transgressors displaying or not displaying guilt (Study 1).

Distribution of resources.

Distribution. A significant majority of the 4-year-olds in Study 2 (17 of 20) gave more flowers (two or three out of three) to the apologetic transgressor (binomial probability, $p = .003$).

We conducted chi-square analyses to compare children’s distribution of resources across studies and ages. These analyses revealed no significant difference in the proportion of 4-year-olds who gave more flowers to the apologetic transgressor (Study 2) versus the proportion of 5-year-olds who gave more flowers to the guilt-displaying transgressor (Study 1), Fisher’s exact test, $p = .609$. The number of flowers given to the apologetic transgressor by the 4-year-olds in Study 2 ($M = 1.9, SD = 0.45$) also did not differ from the number given to the guilt-displaying transgressor by the 5-year-olds in Study 1 ($M = 2.06; SD = 0.43$), $t(35) = 1.097, p = 0.280$. On the other hand, the proportion of 4-year-olds who gave more flowers to the apologetic transgressor (Study 2)

was nearly significantly higher than the proportion of 4-year-olds who gave more flowers to the guilt-displaying transgressor (Study 1), $\chi^2(1, N = 40) = 3.14, p = .077$. The number of flowers given to the apologetic transgressor by the 4-year-olds in Study 2 was also significantly greater than the number given to the guilt-displaying transgressor by the 4-year-olds in Study 1 ($M = 1.55, SD = 0.61$), $t(38) = 2.08, p = .045$.

Justifications. Children's justifications were only included in analyses if children had distributed the resources as predicted (i.e., given more flowers to the apologetic transgressor), although the results remained the same when all justifications were included. Since assigning "Apology" a score of 1 versus a score of 0 did impact the results of the justifications for the distribution of resources task, we will report analyses using both scoring methods (see also Table 2).

When "Apology" was assigned a score of 0, half of the children (eight of 16) were scored as providing a higher level (score of 1) justification. A chi-square analysis indicated that this proportion did not differ significantly from the proportion of 5-year-olds in Study 1 ($p = .576$) or from the proportion of 4-year-olds in Study 1 ($p = .378$) who provided a Level-1 justification when "Guilt (repeated)" was assigned a score of 0.

When "Apology" was assigned a score of 1, the majority of 4-year-olds in Study 2 (13 of 16) were scored as providing a Level-1 justification. This proportion again did not significantly differ from the proportion of 5-year-olds in Study 1 who provided a Level-1 justification when "Guilt (repeated)" was scored 1 ($p = .252$), but it was significantly greater than the proportion of 4-year-olds in Study 1 who provided a Level-1 justification when "Guilt (repeated)" was scored 1 ($p = .019$). Thus, with the stringent coding system (in which repeating relevant phrases that were used during the presentations did not receive a higher score), the 4-year-olds in Study 2 provided similar levels of justifications as the 4- and 5-year-olds in Study 1. However, when repeating relevant phrases from the presentations did receive a higher score, 4-year-olds were seen to use sophisticated and appropriate justifications for their distribution of resources in the context of transgressors apologizing or not apologizing (Study 2) more than in the context of transgressors displaying or not displaying guilt (Study 1).

Discussion

In Study 2, 4-year-olds displayed a sophisticated understanding of the appeasement functions that explicit apologies serve. They inferred that the victim would be less upset with an apologetic transgressor and would like an apologetic transgressor more than an unapologetic one. They stated that they themselves would prefer to play with an apologetic transgressor than with an unapologetic one and drew appropriate inferences about the moral character of unapologetic transgressors. These children also clearly preferred to cooperate with (by distributing more resources to) the apologetic transgressor. Finally, 4-year-olds in Study 2 explained their responses using higher levels of justifications (i.e., justifications that were relevant, appropriate, and sophisticated) more often than 4-year-olds in Study 1. Across all measures, therefore, the 4-year-olds in Study 2 performed very similarly to the 5-year-olds in Study 1.

These results clearly show that the problem for the 4-year-olds in Study 1 was not that the task was too challenging: When the task

was kept identical and only one (seemingly minor) change was made to one transgressor's display, the 4-year-olds' performance changed dramatically and became comparable to that of the 5-year-olds in Study 1. Moreover, they then demonstrated a good understanding of the functions that apologies serve, which replicates and extends prior work (e.g., Smith et al., 2010; Wellman et al., 1979). Thus, 4-year-olds do seem to have a robust grasp of the appeasement functions that explicit apologies serve but are perhaps not yet sensitive to the emotions that apologies stand for, namely, remorse and guilt; this sensitivity seems to emerge between the ages of 4 and 5 years.

General Discussion

In two studies, we assessed what preschoolers understand when a transgressor shows guilt. The 5-year-olds in Study 1 demonstrated a robust grasp of the appeasement functions of guilt displays, whereas the 4-year-olds did not. Moreover, the failure of the 4-year-olds in Study 1 was not due to difficulties with the procedure: In Study 2, when the procedure was kept identical but the content of one transgressor's display was changed from guilt alone to guilt with an apology, the 4-year-olds were able to draw the same sophisticated inferences as the 5-year-olds in Study 1. Our results thus suggest that between the ages of 4 and 5 years, children become sensitive to the appeasement functions of displaying guilt; by 4 years of age, children grasp the appeasement functions of explicit apologies but apparently not of the guilt and remorse that apologies represent.

The question that arises is why 4-year-olds in Study 1 did not demonstrate the same sensitivity to the appeasement functions of guilt displays as the 5-year-olds. One possibility is that although young children seem to experience guilt themselves (e.g., Zahn-Waxler & Kochanska, 1990), the understanding of such complex social emotions might emerge only later in development and is perhaps not fully in place by the age of 4 years (e.g., Harris et al., 1987). An alternative possibility arises from the fact that in Study 1, the predominant response among 4-year-olds was to attribute guilty feelings to a transgressor who displayed no guilt (as seen in their responses to the second comprehension probe). Perhaps children of this age are aware that they themselves would feel guilty in a similar situation, and they project this onto an unremorseful transgressor (Darby & Schlenker, 1982). Relatedly, children may have a script in which transgressors generally experience guilt (Arsenio, 1988; Harris, 1985; Widen & Russell, 2010), and perhaps the younger children were unable to inhibit this script enough to use the actual information presented to them. These ideas are compatible with research on the development of executive functioning, which shows domain-general difficulties including representational inflexibility and trouble switching between incompatible perspectives (e.g., Zelazo et al., 2003). For instance, 5-year-olds succeed at dimensional card sorting and executive control tasks, whereas younger children do not (e.g., Zelazo, 2006). The developmental change observed in Study 1 fits nicely with this developmental transition in children's executive functioning and executive control. Difficulties with executive functioning and control may thus have prevented the 4-year-olds in Study 1 from grasping that one of the transgressors did not experience guilt, which would naturally make it difficult to distinguish be-

tween or draw inferences about a remorseful versus an unremorseful transgressor.

However, in Study 2, most 4-year-olds had little difficulty identifying whether a transgressor had apologized or not, or in using this information to draw sophisticated inferences about the respective transgressors. These young children apparently know the rule that one should apologize after transgressing and also know the appeasement functions that apologies serve, but they may not yet appreciate that the underlying function of the rule is to express guilt and may not yet grasp the appeasement functions that displaying guilt as such serves (see also Darby & Schlenker, 1982). Moreover, if 4-year-olds have a script in which transgressors generally apologize, they were able to inhibit it enough to appropriately use the information provided in Study 2 (i.e., that one transgressor did not apologize). This is perhaps because a rule they have been taught from very early on by their caregivers and teachers takes precedence over the apology script they might be developing for themselves and because seeing a rule being broken is extremely salient during the preschool years (e.g., Rakoczy, Warneken, & Tomasello, 2008; Vaish, Missana, & Tomasello, 2011). In any event, our findings hint at the possibility that 4-year-old children's understanding of guilt displays relies perhaps rather superficially on key words rather than on an understanding of the emotions expressed by those words.

At this juncture, it is important to consider an alternative interpretation of our findings: that 4-year-olds might simply find it easier to respond to explicit verbal expressions of guilt and suffering (as in Study 2) than facial and vocal expressions that lack verbal specificity (as in Study 1). For instance, in Study 2, the remorseful transgressor used the phrase "Das tut mir Leid," which means "I'm sorry," but literally "Leid" translates to "suffering." Thus, one could argue that this emotion label helped 4-year-olds recognize the transgressor's suffering, whereas the nonverbal expressions of guilt in Study 1 were harder for children to process. Indeed, preschoolers demonstrate better understanding of the causes and consequences of various emotions when presented with emotion labels than with facial expressions alone (e.g., Russell & Widen, 2002; Widen & Russell, 2004). It is thus possible that 4-year-olds are unable to infer feelings of guilt from implicit cues alone and that they require more explicit verbal expressions to do so, but that once they have inferred this guilt (or the lack thereof), they do understand its appeasement functions as well as the 5-year-olds (as they demonstrated in Study 2).

However, we do not believe this to be a plausible alternative explanation for our results for two reasons. First, note that even in Study 1, when children had to decode the transgressors' guilt from facial and vocal expressions that lacked verbal specificity, the 4-year-olds were about as good as the 5-year-olds in correctly identifying right away during the comprehension probes that the guilt-displaying transgressor felt bad. (Indeed, their difficulty lay in being able to correctly identify the transgressor in the No Guilt case as *not* feeling bad.) This shows that the 4-year-olds did not find it any more difficult than the 5-year-olds to decode guiltlike feelings in the absence of emotion labels or explicit verbal expressions. Second, in Study 1, after the comprehension probes and before the test questions, children were always explicitly reminded how the transgressors felt (e.g., "She feels bad about it," or "She doesn't feel bad about it"). Thus, children's understanding of whether or not the transgressors felt guilty did not rely solely on

their own ability to decode the guilt display; rather, regardless of whether or not children responded correctly on the comprehension probes, they always were provided the correct information before being asked about the appeasement functions of displaying guilt. Together, these results and procedural safeguards allow us to rule out with some confidence the alternative interpretation that the 4-year-olds in Study 1 could not decode the transgressors' guilt; rather, we argue that 4-year-olds may not yet understand the appeasement functions served by displaying guilt but do understand the appeasement functions of explicit apologies.

One rather surprising finding in the present studies was that both 5-year-olds in Study 1 and 4-year-olds in Study 2 were able to draw appropriate inferences about the moral character of the transgressors (Test Question 4), given Smith and colleagues' (2010) finding that 4- to 5-year-olds were unable to do so on the basis of whether a transgressor apologized or not. There are many possible reasons for this discrepancy. First, we gave children a forced choice between two transgressors and asked them who was "meaner." This may have been easier for children to respond to than Smith and colleagues' question about whether their single transgressor was "nice or naughty most of the time." Second, since we did not include "most of the time" in our moral character question, it is possible that children were in fact responding more specifically about who was meaner in the situations that they had witnessed rather than more generally (which would be related to but not identical to evaluating overall moral character). Finally, our use of videos may have aided children's understanding of the victim and transgressors and their interactions more clearly than the drawings and stories used in prior work did (e.g., Darby & Schlenker, 1982; Smith et al., 2010; Wellman et al., 1979). This may have made it easier for children to make more appropriate evaluations. All in all, it seems that at least under some circumstances, preschoolers can judge moral character to some degree on the basis of whether a transgressor has apologized (by 4 years of age) and whether a transgressor has displayed guilt (by 5 years of age).

A caveat regarding the results of Study 1 concerns the potential confounding factor of intentionality. Specifically, the transgressor who displayed guilt stated that she "didn't want [the damage] to happen," whereas the transgressor who displayed no guilt did not say anything about her intentions. It is thus possible that 5-year-olds relied on this difference between the Guilt versus No Guilt situations rather than on the guilt displays as such. Note, though, that acknowledging that one did not intend to cause harm is in fact a fundamental part of displaying guilt (Fessler & Haley, 2003). Moreover, given that guilt seems not to have a single, clear facial expression (Berti et al., 2000; Zahn-Waxler & Kochanska, 1990) and given that we excluded apologies and attempts to repair from the guilt displays, it was imperative for us to retain at least some verbal cues that are critical for conveying guilt. An examination of children's justifications also shows that children likely did not rely primarily on intentionality information. Specifically, only a small number of children (six 5-year-olds and one 4-year-old) referred to the transgressors' intentions in their justifications, and five of those six 5-year-olds and the one 4-year-old also produced higher level justifications (e.g., apology), suggesting that these children drew their inferences not only from intentionality information but also from other aspects of the transgressors' reactions. Finally, in Study 2, the apologetic transgressor did not say that her actions were

unintentional, yet 4-year-olds in that study were able to distinguish between and respond appropriately about the two transgressors. Altogether, then, it seems unlikely that children in Study 1 relied solely on intentionality information to draw their inferences.

Our findings contribute importantly to the literature on children's emotional understanding more generally. Prior work had shown that by middle childhood, children demonstrate some understanding of complex, social emotions such as guilt (Brody & Harrison, 1987; Harris et al., 1987; Wiggers & van Lieshout, 1985). However, to our knowledge, there had been no previous systematic assessment of whether and when children grasp the appeasement functions of guilt displays as such, in the absence of apologies that children are taught from very early on to use and to expect from others. Our studies show for the first time that certainly by 5 years of age, children are attuned to the important appeasement functions that guilt displays serve (Darby & Schlenker, 1989; Keltner et al., 1997; Leary et al., 1996). This indicates that beyond being able to recognize the facial expressions and grasp the causes and consequences of certain complex emotions (e.g., Berti et al., 2000; Harris et al., 1987; Widen & Russell, 2010), preschool-aged children are also developing an understanding of the sophisticated social functions that these emotions serve (e.g., Goffman, 1956; Keltner et al., 1997), an understanding that is likely critical for children's entry into and smooth functioning in large and complex social groups.

References

- Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development, 59*, 1611–1622. doi:10.2307/1130675
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*, 243–267. doi:10.1037/0033-2909.115.2.243
- Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41*, 328–337. doi:10.1037/0012-1649.41.2.328
- Berti, A. E., Garattoni, C., & Venturini, B. (2000). The understanding of sadness, guilt, and shame in 5-, 7-, and 9-year-old children. *Genetic, Social, and General Psychology Monographs, 126*, 293–318.
- Brody, L. R., & Harrison, R. H. (1987). Developmental changes in children's abilities to match and label emotionally laden situations. *Motivation and Emotion, 11*, 347–365. doi:10.1007/BF00992849
- Castelfranchi, C., & Poggi, I. (1990). Blushing as discourse: Was Darwin wrong? In W. R. Crozier (Ed.), *Shyness and embarrassment: Perspectives from social psychology* (pp. 230–252). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511571183.009
- Darby, B. W., & Schlenker, B. R. (1982). Children's reactions to apologies. *Journal of Personality and Social Psychology, 43*, 742–753. doi:10.1037/0022-3514.43.4.742
- Darby, B. W., & Schlenker, B. R. (1989). Children's reactions to transgressions: Effects of the actor's apology, reputation, and remorse. *British Journal of Social Psychology, 28*, 353–364.
- Fessler, D. M. T., & Haley, K. J. (2003). The strategy of affect: Emotions in human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 7–36). Cambridge, MA: MIT Press.
- Goffman, E. (1956). Embarrassment and social organization. *American Journal of Sociology, 62*, 264–271. doi:10.1086/222003
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Garden City, NY: Anchor.
- Grueneich, R. (1982). Issues in the developmental study of how children use intention and consequence information to make moral evaluations. *Child Development, 53*, 29–43. doi:10.2307/1129636
- Harris, P. L. (1985). What children know about the situations that provoke emotions. In M. Lewis & C. Saarni (Eds.), *The socialization of emotion* (pp. 161–186). New York, NY: Plenum.
- Harris, P. L., Olthof, T., Terwogt, M. M., & Hardman, C. E. (1987). Children's knowledge of the situations that provoke emotion. *International Journal of Behavioral Development, 10*, 319–343.
- Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg (Ed.), *The development of prosocial behavior* (pp. 281–338). New York, NY: Academic Press.
- Irwin, D. M., & Moore, S. G. (1971). The young child's understanding of justice. *Developmental Psychology, 5*, 406–410. doi:10.1037/h0031614
- Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*, 441–454. doi:10.1037/0022-3514.68.3.441
- Keltner, D., & Anderson, C. (2000). Saving face for Darwin: The functions and uses of embarrassment. *Current Directions in Psychological Science, 9*, 187–192. doi:10.1111/1467-8721.00091
- Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition & Emotion, 10*, 155–172. doi:10.1080/026999396380312
- Keltner, D., Young, R. C., & Buswell, B. N. (1997). Appeasement in human emotion, social practice, and personality. *Aggressive Behavior, 23*, 359–374. doi:10.1002/(SICI)1098-2337(1997)23:5<359::AID-AB5>3.0.CO;2-D
- Ketelaar, T., & Au, W. T. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition & Emotion, 17*, 429–453. doi:10.1080/02699930143000662
- Kochanska, G., DeVet, K., Goldman, M., Murray, K., & Putnam, S. P. (1994). Maternal reports of conscience development and temperament in young children. *Child Development, 65*, 852–868. doi:10.2307/1131423
- Kochanska, G., Gross, J. N., Lin, M.-H., & Nichols, K. E. (2002). Guilt in young children: Development, determinants, and relations with a broader system of standards. *Child Development, 73*, 461–482. doi:10.1111/1467-8624.00418
- Leary, M. R., Landel, J. L., & Patton, K. M. (1996). The motivated expression of embarrassment following a self-presentational predicament. *Journal of Personality, 64*, 619–636. doi:10.1111/j.1467-6494.1996.tb00524.x
- Levin, J., & Arluke, A. (1982). Embarrassment and helping behavior. *Psychological Reports, 51*, 999–1002.
- Levitt, M. J., Weber, R. A., Clark, M. C., & McDonnell, P. (1985). Reciprocity of exchange in toddler sharing behavior. *Developmental Psychology, 21*, 122–123. doi:10.1037/0012-1649.21.1.122
- McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology, 53*, 247–256. doi:10.1037/0022-3514.53.2.247
- Miller, R. S. (1987). Empathic embarrassment: Situational and personal determinants of reactions to the embarrassment of another. *Journal of Personality and Social Psychology, 53*, 1061–1069. doi:10.1037/0022-3514.53.6.1061
- Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*, 3219–3237. doi:10.2307/1131775
- Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby effect. *Emotion, 9*, 118–122. doi:10.1037/a0014540
- Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child Development, 58*, 889–909. doi:10.2307/1130530

- Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology, 104*, 382–397. doi:10.1016/j.jecp.2009.08.001
- Ohbuchi, K., Kameda, M., & Agarie, N. (1989). Apology as aggression control: Its role in mediating appraisal of and response to harm. *Journal of Personality and Social Psychology, 56*, 219–227. doi:10.1037/0022-3514.56.2.219
- Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition, 108*, 222–231. doi:10.1016/j.cognition.2007.12.003
- O'Malley, M. N., & Greenberg, J. (1983). Sex differences in restoring justice: The down payment effect. *Journal of Research in Personality, 17*, 174–185. doi:10.1016/0092-6566(83)90029-6
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*, 875–881. doi:10.1037/0012-1649.44.3.875
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12–21. doi:10.1037/0012-1649.33.1.12
- Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development, 11*, 30–52. doi:10.1111/1467-9507.00185
- Saarni, C., Campos, J. J., Camras, L., & Witherington, D. (2006). Emotional development: Action, communication, and understanding. In W. Damon and R. M. Lerner (Series Eds.) and N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 226–299). New York, NY: Wiley.
- Semin, G. R., & Manstead, A. S. R. (1981). The beholder beheld: A study of social emotionality. *European Journal of Social Psychology, 11*, 253–265. doi:10.1002/ejsp.2420110302
- Smith, C. E., Chen, D., & Harris, P. L. (2010). When the happy victimizer says sorry: Children's understanding of apology and emotion. *British Journal of Developmental Psychology, 28*, 727–746. doi:10.1348/026151009X475343
- Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M., & Shirvani, N. (2008). The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story. *Journal of Experimental Social Psychology, 44*, 983–992. doi:10.1016/j.jesp.2008.02.006
- Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 863–932). New York, NY: Wiley.
- Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective-taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45*, 534–543. doi:10.1037/a0014322
- Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development, 81*, 1661–1669. doi:10.1111/j.1467-8624.2010.01500.x
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383–403. doi:10.1037/0033-2909.134.3.383
- Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*, 124–130. doi:10.1348/026151010X532888
- Vaish, A., & Woodward, A. (2010). Infants use attention but not emotion to predict others' actions. *Infant Behavior and Development, 33*, 79–87. doi:10.1016/j.infbeh.2009.11.003
- Wellman, H. M., Larkey, C., & Somerville, S. C. (1979). The early development of moral criteria. *Child Development, 50*, 869–873. doi:10.2307/1128956
- Widen, S. C., & Russell, J. A. (2004). The relative power of an emotion's facial expression, label, and behavioral consequence to evoke preschoolers' knowledge of its cause. *Cognitive Development, 19*, 111–125. doi:10.1016/j.cogdev.2003.11.004
- Widen, S. C., & Russell, J. A. (2010). Children's scripts for social emotions: Causes and consequences are more central than are facial expressions. *British Journal of Developmental Psychology, 28*, 565–581. doi:10.1348/026151009X457550d
- Wiggers, M., & van Lieshout, C. F. (1985). Development of recognition of emotions: Children's reliance on situational and facial expressive cues. *Developmental Psychology, 21*, 338–349. doi:10.1037/0012-1649.21.2.338
- Zahn-Waxler, C., & Kochanska, G. (1990). The origins of guilt. In R. A. Thompson (Ed.), *The Nebraska Symposium on Motivation 1988: Socio-emotional development* (Vol. 36, pp. 183–258). Lincoln: University of Nebraska Press.
- Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology, 28*, 126–136. doi:10.1037/0012-1649.28.1.126
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301. doi:10.1038/nprot.2006.46
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., . . . Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*, i–151.

Received October 6, 2010

Revision received March 22, 2011

Accepted March 29, 2011 ■