CHILD DEVELOPMENT

Child Development, March/April 2017, Volume 88, Number 2, Pages 597–611

Following My Head and My Heart: Integrating Preschoolers' Empathy, Theory of Mind, and Moral Judgments

Courtney L. Ball, Judith G. Smetana, and Melissa L. Sturge-Apple
University of Rochester

Associations among hypothetical, prototypic moral, and conventional judgments; theory of mind (ToM); empathy; and personal distress were examined in 108 socioeconomically diverse preschoolers (M_{age} = 42.94 months, SD = 1.42). Repeated measures analysis of covariance with empathy, false beliefs, and their interaction as covariates indicated that empathy was significantly associated with judgments of greater moral but not conventional transgression severity, particularly for psychological harm, and with deserved punishment for unfairness. False beliefs were associated with (combined) moral criterion judgments of rule and authority independence and inalterability. Empathy also was positively associated with criterion judgments but only for children low in ToM. Personal distress was unrelated to judgments. Results demonstrate the importance of both affective and cognitive processes in preschoolers' moral judgments.

Research indicates that through their social experiences, young children develop rudimentary concepts of justice, fairness, and rights (Smetana, Jambon, & Ball, 2014). Both affective and cognitive processes are thought to play a role in the construction of these early moral concepts. Despite the identified need for an integrative developmental approach to the study of moral cognitions and emotions, the mutual contribution of these processes to early moral judgments has not yet been systematically examined (Malti & Latzko, 2010; Turiel, 2015). In particular, preschoolers' empathy (Hoffman, 2000) and theory of mind (ToM) understanding (Decety & Jackson, 2004) advance their knowledge of others' psychological and emotional experiences and thus may be constituent elements of early moral judgments. Therefore, the current study addresses this gap in the literature by investigating associations among preschoolers' empathy, false belief understanding, and moral judgments.

Early Moral Development

Infants are predisposed to be sensitive and responsive to others' emotional expressions (Trevarthen & Aitken, 2001). This propensity helps young children acquire knowledge of their social and internal worlds, such as understanding why others transgress or the consequences of different social acts. Through these experiences, preschoolers are thought to construct qualitatively different categories, or domains, of social knowledge, including moral and social-conventional ones (see Smetana et al., 2014). Moral concepts entail prescriptive judgments of right and wrong concerning issues of welfare (harm), justice (comparable treatment and distribution), and rights. In contrast, social-conventional concepts focus on arbitrary, agreed-upon norms, and regularities (like etiquette) that structure social interactions in different social systems. Moral transgressions typically, though not invariably, are deemed more serious and deserving of punishment than conventional violations (Tisak & Turiel, 1988). Similarly, because the consequences of immoral acts are intrinsically harmful or cause injustice for others, even young children judge that moral prescriptions (but not conventional norms) are *obligatory*, *generalizable*, and *unalterable* (e.g., wrong regardless of whether a rule exists or an authority figure mandates otherwise). These dimensions are seen as formal *criteria*, drawn from philosophical definitions of morality, for distinguishing moral from nonmoral issues. Accordingly, *criterion judgments* assess whether young children recognize distinctively moral features of actions.

This research is supported by the National Institute of Nursing Research (R21 NR010857-01) awarded to Melissa L. Sturge-Apple. Special thanks to Michael A. Skibo, Jennifer H. Suor, and Project CONNECT undergraduate research assistants for their help with data collection and entry; to participating mothers and children for their cooperation; and to Jessica R. Robinson for her coding contributions.

Correspondence concerning this article should be addressed to Courtney L. Ball, Department of Clinical & Social Sciences in Psychology, University of Rochester, Meliora Hall, RC 270266, Rochester, NY 14627. Electronic mail may be sent to cball@psych.rochester.edu.

© 2016 The Authors
Child Development © 2016 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2017/8802-0024
DOI: 10.1111/cdev.12605

Children as young as 3 years of age apply these criteria in making moral and conventional judgments (Smetana & Braeges, 1990), although some evidence suggests that this varies according to harm type (Helwig, Hildebrandt, & Turiel, 1995; Helwig, Zelazo, & Wilson, 2001). Children's thinking about psychological harm and unfairness typically develops later than an understanding of physical harm (Smetana, 1981; Wainryb & Brehl, 2006). This is potentially because psychological harm and unfairness must be inferred, whereas physical harm is directly observable. To recognize that psychological harm has occurred, one must consider the victim's thoughts and feelings as well as the actor's intent. Fairness violations (e.g., stealing or not sharing) may concretely illustrate inequality, but interpreting the act as harmful also requires the understanding that the victim desires the deprived resource. This capacity is limited by competing self-interest until around age 4 (Birch & Billman, 1986; Dunn & Munn, 1987). Same-age preschoolers with more advanced affective and cognitive capacities that facilitate these inferences therefore may have more mature judgments about relatively abstract moral harms (Dunn, Cutting, & Demetriou, 2000; Wainryb & Brehl, 2006).

Empathy and Moral Development

In mature moral reasoning, children know *and*, importantly, care about the emotional and psychological experiences of others (Hoffman, 2000). Young preschoolers demonstrate an emerging awareness of the subjectivity of others' emotional states and increasingly experience emotional concern for "victims" (Davidov, Zahn-Waxler, Roth-Hanania, & Knafo, 2013). Empathy for the actual or inferred distress of others is one of the earliest manifestations of very young children's moral capacities (Hoffman, 2000) and may be a central organizing feature of moral concepts (Turiel, 2015). The definition and scope of empathy varies across conceptualizations (see Decety & Meyer, 2008), but here we consider empathic concern, or sympathy, as an affective response to the perceived emotional state or condition of another that is similar to what the other person is feeling and evokes concern for the other (Eisenberg, Spinrad, & Morris, 2014). Empathy-related responding, or a vicarious response to another's distress, does not always elicit other-oriented concern. It can, instead, result in personal distress, a self-oriented and aversive response that diverts attention from the victim's suffering and devotes cognitive and affective processes to self-comfort (Eisenberg et al., 1988). Thus, empathy, but not personal distress, has been found to facilitate prosocial moral development, or acting to benefit others without apparent personal gain (Eisenberg et al., 2014).

Accordingly, empathic concern for the victims of moral transgressions underscores the emotional salience and moral significance of the social exchange (Arsenio, 1988) and may reflect children's perceptions of the act's severity (Malti & Latzko, 2010). A few studies have demonstrated positive links between preschoolers' prosocial moral reasoning (Lane, Wellman, Olson, LaBounty, & Kerr, 2010; Miller, Eisenberg, Fabes, & Shell, 1996) and moral emotions, or affective responses generally acknowledged to influence a person's understanding of the prescriptive nature of moral norms (Nussbaum, 2001). However, these findings may not extend to moral reasoning about harm, which uniquely centers on obligation (Weller & Lagattuta, 2013). Prosocial acts are not obligatory in all contexts and therefore entail a component of personal choice that is absent from the universal obligation not to harm others (Nucci & Turiel, 2009). Given the dearth of existing research, associations between empathy and early moral judgments of harm must be investigated.

ToM and Moral Development

Much research in recent years indicates that young children's psychological and moral knowledge are interrelated (see Lagattuta & Weller, 2014). ToM is the ability to attribute mental states—such as belief, desires, emotions, or intentions—to others. By 3–4 years of age, children begin to understand that individuals can have false beliefs, or incorrect assumptions about reality. This emerging ability is commonly identified as a developmental milestone of ToM (Wellman & Liu, 2004). A rudimentary understanding of false beliefs may enable children to make more informed or flexible moral evaluations by facilitating accurate inferences about the intentions or experiences of others (Buttelmann, Carpenter, & Tomasello, 2009; Smetana, Jambon, Conry-Murray, & Sturge-Apple, 2012). Indeed, children with more advanced false belief understanding focus more on the psychological states of others in their moral justifications (Dunn et al., 2000). Inversely, young children's yet incomplete mastery of ToM also constrains their moral judgments in important ways. For example, children lacking false belief understanding judge accidental transgressions as harshly as intentional acts (Killen, Mulvey,

Richardson, Jampol, & Woodward, 2011). Similarly, 3-year-olds can accurately predict consequences in prototypic, causal situations (e.g., hitting causes pain), but not in situations where responses are atypical (e.g., hitting causes pleasure; Zelazo, Helwig, & Lau, 1996).

Although together these studies suggest higher ToM development may facilitate early moral judgments (or constrain them when ToM is low), not all moral evaluations require mental state understanding. Young children are aware of moral and social rules and make prototypic moral evaluations well before they evince comparable mental state understanding (Dunn et al., 2000; Smetana & Braeges, 1990). Thus, ToM is influential in early moral reasoning, but the relation is complex, and the available findings are mixed.

Thinking, Feeling, and Judging

Although research demonstrates links between preschoolers' understanding of other minds (ToM) and their interpretations of moral events, little research has examined how ToM interacts with empathy in early moral understanding. This is surprising, as empathy and ToM are often seen as interdependent developmental processes (Cutting & Dunn, 1999). Empathy may facilitate young children's understanding of others' minds by directing them to psychological consequences, whereas ToM may advance their discernment of others' emotional responses and subsequent empathy (Harris, Johnson, Hutton, Andrews, & Cooke, 1989). In support of this view, more empathic 4- and 5-year-olds make more advanced inferences about others' thoughts, emotions, and intentions (Findlay, Girardi, & Coplan, 2006). Moreover, many theorists agree that empathy requires a basic level of self–other differentiation and correspondingly, an understanding that the origin of distress is the other, not the self (Eisenberg et al., 2014; Hoffman, 2000). Accordingly, as 14- to 18-month-olds more clearly differentiate self from other, they also demonstrate more other-directed empathic concern and soon after can empathize with victims even when they do not express negative emotion (Vaish, Carpenter, & Tomasello, 2009).

Inversely, limitations in preschoolers' ability to take others' perspectives may inhibit empathic concern (Harris et al., 1989). By age 3, children's empathic responses are greater when they perceive that the distressed individual has legitimate cause (e.g., was seriously harmed rather than mildly inconvenienced; Hepach, Vaish, & Tomasello, 2013). Children who lack the requisite psychological knowledge to perceive and understand the negative consequences of more abstract harms might not respond empathically, despite being sensitive to the victims' distress, because they do not view the distress as legitimate. Indeed, preschoolers who better understand others' emotions and cognitions demonstrate more mature prosocial moral reasoning and heightened tendencies to sympathize with others (Dunn et al., 2000; Lane et al., 2010). Thus, the contribution of empathy to preschoolers' moral judgments may be facilitated or constrained by ToM understanding and vice versa. Therefore, both processes must be investigated concurrently with moral judgments to capture these complex associations.

The Current Study

This study examined associations among empathy, false belief understanding, and moral judgments in a socioeconomically diverse sample of preschoolers. We focused on these associations in 3½-year-old children, based on research demonstrating that there is significant variability and growth in moral judgments (Smetana, Rote, et al., 2012), ToM understanding (Flavell, Mumme, Green, & Flavell, 1992), and empathic concern between the ages of 3 and 4 years (Davidov et al., 2013; Rochat et al., 2009). Thus, this is an optimal age for examining these associations. Consistent with previous research (e.g., Smetana et al., 2014), we asked preschoolers to judge act severity, deserved punishment, and criterion judgments (e.g., rule and authority independence and rule inalterability) about prototypic moral (physical harm, psychological harm, and unfairness) and conventional transgressions. Transgressors' beliefs and intentions and the transgressions' consequences were left ambiguous to capture variability theoretically associated with empathy or ToM. As such, the interview required child participants to interpret the depicted situation and imagine the likely consequences in order to formulate judgments based on those inferred consequences.

The central hypothesis of the present study was that individual differences in empathy (assessed using children's behavioral responses and maternal reports) and false belief understanding (as a standard "litmus test" for ToM in our target age group; Wellman & Liu, 2004) would be associated with moral but not conventional judgments. Specifically, we hypothesized that higher levels of empathy would be associated with judgments that moral transgressions are more serious and deserving of punishment. Past research has yielded inconsistent

findings on the association between ToM understanding and ratings of transgression severity and deserved punishment (e.g., Baird & Astington, 2004; Smetana, Jambon et al., 2012), so we examined but had no a priori hypotheses about these associations.

Criterion judgments require children to think about the consequences of immoral acts under different conditions (e.g., with and without the presence of authority). Higher scores on criterion judgments index moral maturity because they indicate that individuals recognize that moral prescriptions are obligatory regardless of context. Therefore, we hypothesized that more advanced ToM would be associated with judgments that moral transgressions are more authority and rule independent and unalterable. Although even young children are generally concerned for others' welfare, they nonetheless struggle to apply concerns for others' perspectives in their moral reasoning, possibly due to rudimentary ToM skills (Killen et al., 2011). Therefore, we predicted that empathy also would be positively associated with more mature moral criterion judgments but only for children higher in ToM.

We further expected that these associations would differ according to the type of moral harm under consideration. Past research has shown that preschoolers almost unanimously view physical harm as wrong but vary in their judgments of psychological harm and unfairness (Helwig et al., 1995, 2001). Therefore, we expected that transgressions involving physical harm would be judged as more serious, deserving of punishment, and independent of authority and rules than either psychological harm or unfairness. We also hypothesized that being relatively high on either empathy or ToM would advance children's evaluations of these relatively more abstract harms. Specifically, for psychological harm and unfairness but not physical harm, we expected that higher levels of empathy would be associated with judgments that moral transgressions are more serious and deserving of punishment and that greater ToM would be associated with more mature criterion judgments. Consistent with previous research (Lane et al., 2010), we therefore predicted that preschoolers high on *both* empathy and ToM would have the most mature moral judgments overall.

Empathy but not personal distress directs observers of immoral acts to the victim and source of their shared distress. Thus, as an additional test of specificity, we hypothesized that personal distress would not be associated with moral judgments. Moreover, there is some evidence that preschoolers' verbal skills influence their false belief understanding (Milligan, Astington, & Dack, 2007) and moral judgments (Smetana & Braeges, 1990). Therefore, we controlled for estimated verbal IQ to ensure that results were not an artifact of differences in task comprehension. Finally, as sex differences favoring girls often have been found in empathy (Eisenberg et al., 2014) and ToM (Charman, Ruffman, & Clements, 2002), the potential moderating effect of child sex on associations was examined. Few sex differences in preschoolers' moral and conventional judgments have been found, however, and thus no a priori hypotheses were tested.

In summary, the primary hypotheses were that (a) empathy will be positively associated with prototypic moral judgments of transgression severity and deserved punishment, whereas ToM and empathy (for preschoolers higher in ToM) will be associated with moral criterion judgments. We also expected that (b) children will rate physical harm as more serious, deserving of punishment, and higher on different moral criteria than psychological harm or unfairness, and correspondingly, that associations will differ according to the form of moral harm, with empathy and ToM associated with judgments of psychological (but not physical) harm and unfairness. In addition, as tests of the specificity of our hypotheses, we expected that (c) empathy, ToM, and their interactions will be associated with judgments regarding moral but not conventional transgressions, and (d) personal distress will not be related to moral judgments.

Method

Participants/Sample

The sample consisted of 108 mother–child dyads recruited through local women, infants, and children assistance offices and via flyers posted in the community (e.g., doctors' offices, day cares, and libraries) in a medium-sized city in the northeastern United States. Participants were 3½ years old (M = 42.94 months, SD = 1.42, 52 boys) and racially diverse (64% Caucasian, 14% African American, 15% biracial/multiracial, 7% other). Ethnicity approximated the demographics of the city. The sample was also socioeconomically diverse, as determined by mothers' highest level of education (17% high school level or less, 17% some college or trade school, 43% completed college, 23% advanced degrees), reported government assistance (32% receiving assistance), household annual income (M = \$59,950, SD = \$45,225), and family structure

(58.8% married). Approximately 45% of families were residents of low-income neighborhoods concentrated in the city, whereas the other 55% resided primarily in middle-income suburban neighborhoods.

Measures

Child Empathy Induction

A hurt-knee simulation was adapted from previous research to assess children's empathic responses (Miller et al., 1996). After spending approximately an hour with an experimenter, the child was told that it was time to go to another room and was asked to wait by the door. The experimenter rose to follow and simulated bumping his or her knee on the tabletop, making a loud bang and feigning distress (adopting a pained expression and exclaiming, "Owww—I hit my knee, and it really hurts!"). Throughout the 1-min simulation, the experimenter evinced vocal and facial expressions of pain while rubbing the knee and avoiding eye contact with the child. The first 30 s of the episode were characterized by moderate distress vocalizations; in the following 30-s recovery period, the experimenter gradually indicated that she or he was feeling better. To ensure standardization across the five experimenters who performed the simulation, two coders reliably rated ($\alpha = .91$) the quality of enacted distress on a 4-point scale from 1 (*unconvincing*) to 4 (*extremely convincing*). Enactments were mostly convincing ($M = 3.81$), and entering these scores as a covariate in the analyses did not alter the effects; thus, they are not discussed further.

Children's affective responses to the simulation were coded using criteria adapted from Eisenberg et al.'s (1988) Time Sampling of Affect coding scheme. Empathic concern and personal distress responses were coded for presence and intensity every 10 s on a 5-point rating scale from 0 (*no expression of emotion*) to 4 (*pronounced expression, entire period*). Empathic concern and personal distress were both assessed by facial indicators as well as body language and behavioral cues (see Appendix S1 for a complete description of the coding). To determine interrater reliability, two trained coders rated 20% of the distress simulations. Reliability (intraclass correlation coefficients) was .91 for empathic concern and .84 for personal distress.

Mother Reports of Child Empathy

Mothers' ratings of their child's empathy were assessed on five items from Eisenberg et al. (1988). Four items (e.g., "My child gets upset when she/he sees another child being hurt," "My child usually feels sorry for other children who are being teased") were rated on a 4-point Likert scale ranging from 1 (*not at all true of my child*) to 4 (*very true of my child*). One item ("How often does your child feel sympathetic?") was rated on a 5-point Likert scale ranging from 1 (*very rarely or not at all*) to 5 (*extremely often*). Scores for each item were standardized and combined; the scale demonstrated acceptable reliability ($\alpha = .77$).

Empathic concern codes and mother report of empathy were significantly but moderately correlated ($r = .32$) and were averaged to create a composite empathy score.

Theory of Mind

We administered two standard ToM tasks assessing false belief understanding (see Wellman & Liu, 2004 for a detailed description of both tasks). The first, an "unexpected contents" false belief task, was adapted from Gopnik and Astington's (1988) "Smarties task." The second was the "change of location" ("Maxi and the chocolate") false beliefs task (Wimmer & Perner, 1983). Children had to answer several control questions correctly to ensure that they understood both tasks. Nearly all control questions (approximately 95%) were correctly answered on the first attempt, and all children answered correctly before the target questions were asked. The "Smarties" task was scored as 0 (*incorrect response*) or 1 (*correct response*). The "Maxi" task had two target questions; children had to answer both correctly to be coded as passing the task (0 = *1 or 2 incorrect responses*, 1 = *2 correct responses*). The change of location and unexpected contents tasks were passed by 36% and 17% of participants, *SD*s = 0.44, 0.31, respectively. The scores for the two tasks were summed to create a composite ToM score ranging from 0 (*failed both tasks)* to 2 (*passed both tasks)*.

Social Events Interview

The story stimuli for this study consisted of eight 8 × 11 in. colored illustrations depicting six prototypical moral transgressions: physical harm (hitting, shoving), psychological harm (name calling, teasing), and unfairness (stealing a snack, grabbing a toy) and two prototypical conventional transgressions (eating spaghetti with fingers, calling mom by first name). Four sets of pictorial stimuli corresponded to the race (White or Black) and sex of the

child participant. Biracial children were shown pictures most closely matched to their own skin tone. In each story, the relationship between the transgressor and victim was described as friendly, and the acts were identified as intentional, but beyond this, the pictures were presented with minimal contextual information (see Appendix S2 for full interview text). Children were trained to use a frown scale, a 3-point Likert scale with faces depicted in graduated size and severity of expression from neutral to a deep frown. The scale ranged from 1 (*just OK*), depicted by a small, neutral face, to 3 (*very bad*), depicted by the largest face and deepest frown.

Children were presented with each story and asked in a fixed order a standardized set of questions adapted from previous research (Smetana, 1985). Children first were asked, "Is it OK or not OK for (transgressor's name) to X?" If the child responded that it was not OK, they then were asked, "Is it a little bit bad or very bad?" Responses were coded on a 3-point scale of *severity* ranging from 1 (*just OK*) to 3 (*very bad*). Criterion judgments were assessed using the 3-point frown scale. For *authority independence*, children were asked, "What if your (parent figure) said it would be alright to X? How wrong or alright would it be to X then?" We next asked, "What if there was no rule about X? How wrong or alright would it be to X then?" assessing rule *independence*, and "Who makes the rules (tells you what is right and wrong) at your house? What if (person named) decided to change the rule to say it was alright to X? How wrong or alright would it be to change the rule?" assessing *inalterability*. Higher scores indicated greater authority independence, rule independence, and inalterability. Finally, to assess *deserved punishment*, children were asked, "Should (transgressor) get in trouble for X, like get put in time out?" If the child responded "Yes," they were asked, "A little bit or a lot?" Responses were scored on a 3-point scale ranging from 1 (*no punishment)* to 3 (*a lot of punishment*). Mean scores were obtained separately for items pertaining to physical harm, psychological harm, unfairness, and conventions for each of the five judgments.

Chi-square tests of the transgression pairs (e.g., hitting and shoving stories) indicated that judgments did not differ across the two variants for each transgression type. Accordingly, judgments were averaged across the two stories for each transgression type to create three composite *moral harm* scores, one each for physical harm, psychological harm, and unfairness, and one composite *conventional violation* score for each of the five judgments (severity, authority independence, rule independence, inalterability, and deserved punishment). We did not expect meaningful differences among the authority independence, rule independence, and inalterability judgments, and preliminary analyses did not yield significant differences across the three criteria, either for each harm type or in the pattern of associations. Furthermore, a principal components analysis of the five judgments supported a three-factor solution, with authority independence, rule independence, and inalterability loading on one component: factor loadings = .80, .91, .78, respectively, and severity and punishment each loading highly on separate factors. Therefore, the scores were averaged to create one composite *criterion judgment* variable for each of the three types of moral harm and one composite conventional violation.

Verbal IQ

Children's verbal intelligence was assessed using the short-form of the Wechsler Preschool and Primary Scale Intelligence Test, 3rd ed. (WPPSI–III). Two subtests (information and picture naming) validated for use with 2- to 4-year-olds (Sattler, 2008) were used to estimate a verbal IQ composite score for each child. Items for both subtests were coded as pass/fail and scored in accordance with the WPPSI–III manual, normed based on child age to have a mean of 100 and a standard deviation of 15 (M = 105.18, SD = 15.99 here).

Procedure

Data were collected in 2011–2012. Participating mother–child dyads were assessed during an approximately 2-hr visit to a university laboratory. Additional data (not examined here) were collected during the session, but all social-cognitive tasks occurred at the beginning of the visit and were completed within the 1st hour. After obtaining consent and permission, mothers reported on their demographic background and their child's empathy on a laboratory computer. Concurrently in a separate room, a trained interviewer administered to the child (in order) the WPPSI (~15 min), Social Events Interview (~25 min), and ToM tasks (~15 min).

The child and the interviewer sat at a child-sized table alone in the laboratory, but the sessions were filmed by two hidden cameras controlled by a laboratory assistant in the adjacent room. The Social

Events Interview stories were divided into two sets, both including one of each type of social transgression (physical harm, psychological harm, unfairness, and convention). The two sets of stimuli, as well as the order of the stories within each set, were administered in counterbalanced order across participants. Between the first and second set of interview stories, the child participated in an engaging game to reduce fatigue (searching for various attractive animals pictured in a wall mural). Next, children completed the two ToM tasks. Child responses to all tasks were coded by a trained assistant observing in the filming room and later confirmed by another assistant reviewing the recording. At the end of the session, the hurt-knee paradigm was enacted. Children's responses to the simulation were filmed with one camera focused on the child's face and the second camera capturing body positioning and movements or gestures.

Results

The means, standard deviations, and correlations among the main study variables are presented in Table 1. As Table 1 shows, the study variables were moderately correlated, as expected. Sex was correlated with empathy, with girls higher in empathy than boys. Although severity and deserved punishment judgments have been highly correlated in past research, the associations were only moderate here. Associations between preschoolers' empathy and ToM have varied across studies, and here these variables were uncorrelated.

All continuous independent variables were mean centered, as recommended (Aiken & West, 1991). All analyses were run first with verbal IQ included as a covariate to control for differences in language ability and child sex included as a between-subjects factor to test for evidence of sex moderation. Only two effects out of a possible 36 involving verbal IQ or child sex were significant: Higher verbal IQ was associated with moral severity judgments, $\beta = .41$, $p = .001$, but only for boys. Furthermore, excluding child sex and verbal IQ from models did not alter the overall pattern of results for the variables of interest. Thus, for parsimony, verbal IQ and child sex were omitted from presented analyses.

Study hypotheses were examined in multiple steps. First, we conducted detailed analyses testing our main hypotheses that individual differences in empathy and ToM would be associated with judgments about different types of moral harm. Then, we tested the specificity hypotheses in separate analyses, examining effects first for moral versus conventional transgressions and next for personal distress rather than empathic concern.

Associations Among Empathy, ToM, and Different Moral Judgments

The first set of analyses examined associations among empathic concern, ToM, and moral judgments regarding different forms of harm. Severity, deserved punishment, and the composite criterion judgments were each analyzed separately. Hypotheses were tested using repeated measures analysis of covariance (ANCOVA), with moral harm type (physical harm, psychological harm, unfairness) as the within-subject factor and the mean-centered composite empathy and ToM scores as well as their interaction included as covariates. ANCOVA is comparable to multiple hierarchical regression analyses when the independent variables of interest (i.e., empathy and ToM) are continuous (see Rutherford, 2001). Utilizing a repeated measures design increased statistical power by accounting for individual differences that might otherwise obscure effects within levels of a factor (e.g., physical vs. psychological harm). Thus, repeated measures ANCOVA provided a powerful and appropriate analytic method for testing our hypotheses (Wainryb, Shaw, Laupa, & Smith, 2001).

Moral Severity and Deserved Punishment Judgments

Our hypothesis that higher levels of empathy would be associated with judgments of greater moral transgression severity and deserved punishment was partially confirmed. A significant main effect of empathy, $F(1, 103) = 4.25$, $p = .04$, $\eta_p^2 = .04$, indicated that higher empathy was associated with judgments that moral transgressions were more serious, $\beta = .13$. However, a similar main effect for deserved punishment was not found, $p = .37$. ToM was not significant either as a main effect or in interaction with empathy for either severity or deserved punishment judgments, all ps $> .20$.

Similarly, our hypothesis that physical harm transgressions would be judged as more serious and deserving of punishment than psychological harm or unfairness was confirmed only partially. There was a significant main effect of moral harm type for severity judgments, $F(2, 206) = 4.41$, $p = .01$, $\eta_p^2 = .04$, but not for deserved punishment judgments, $p = .15$. Bonferroni-corrected pairwise comparisons indicated that, as predicted, children

Table 1
Means, Standard Deviations, and Correlations Among Study Variables

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Child sex	—	—	—															
2. Verbal IQ	105.18	15.99	.12	—														
3. Empathy	1.74	1.33	.22*	.12	—													
4. Personal distress	0.69	0.96	−.09	−.08	−.13	—												
5. ToM	0.36	0.55	−.08	.21	−.10	.18	—											
6. Sev—physical	2.51	0.63	−.06	.39**	.00	−.17	.02	—										
7. Sev—psychological	2.42	0.71	−.09	.27*	.27**	.04	.07	.44**	—									
8. Sev—unfairness	2.35	0.70	.06	.24*	.16	−.18	.02	.56**	.54**	—								
9. Sev—conventional	2.12	0.77	.11	.11	.14	−.13	−.06	.39**	.38**	.33**	—							
10. CJ—physical	2.34	0.59	.01	.03	.08	−.05	.11	.22*	.28**	.26**	−.01	—						
11. CJ—psychological	2.31	0.59	−.11	.20	−.04	−.07	.00	.19*	.27**	.22*	.10	.47**	—					
12. CJ—unfairness	2.24	0.64	.09	.13	.19*	−.11	.16	.33**	.34**	.54**	.24*	.51**	.43**	—				
13. CJ—conventional	2.12	0.55	.09	−.05	.08	−.11	−.15	.09	.18	.12	.20*	.39**	.52**	.25**	—			
14. Pun—physical	2.53	0.66	−.15	.16	.01	.07	.22*	.26**	.34**	.24*	.18	.27**	.09	.12	.10	—		
15. Pun—psychological	2.42	0.67	.01	.18	.20*	.12	.09	.34**	.52**	.39**	.33**	.15	.19*	.21*	.10	.59**	—	
16. Pun—unfairness	2.43	0.71	.06	.02	.13	.00	.13	.18	.32**	.38**	.30**	.25**	.16	.25**	.18	.47**	.56**	—
17. Pun—conventional	2.28	0.73	−.02	.03	.06	.13	.07	.17	.24*	.17	.52**	.03	.11	.05	.28**	.45**	.45**	.44**

Note. Means for empathy, personal distress, and ToM are uncentered. Empathy and personal distress were scored on 5-point scales with 4 = *higher affect*; ToM was scored on a 3-point scale with 2 = *passed both tasks*; and judgments were scored on 3-point scales with 3 = *very bad or a lot of punishment*. Sev = severity; Pun = deserved punishment; CJ = criterion judgments; ToM = theory of mind. *$p < .05$. **$p < .01$.

judged physical harm as significantly more serious than unfairness, $p = .01$. However, contrary to expectations, severity judgments for psychological harm did not differ significantly from either physical harm or unfairness, ps = .68, .31, respectively.

We further hypothesized that effects of harm type would be moderated by empathy. This hypothesis was supported, as shown by the significant Moral Harm × Empathy interactions for judgments of severity, $F(2, 206) = 4.99$, $p = .01$, $\eta_p^2 = .05$, and deserved punishment, $F(2, 204) = 2.94$, $p = .05$, $\eta_p^2 = .03$. As hypothesized, higher empathy was significantly associated with greater severity ratings for psychological harm, $t(103) = 3.01$, $\beta = .28$, $p < .001$, $\eta_p^2 = .08$; marginally associated with greater severity for unfairness, $t(103) = 1.90$, $\beta = .16$, $p = .06$, $\eta_p^2 = .03$; and not significantly associated with severity for physical harm, $p = .86$ (see Figure 1). In contrast, greater empathy was significantly associated with judging unfairness—but not physical or psychological harm, ps > .52—as more deserving of punishment, $t(103) = 1.96$, $\beta = .16$, $p = .05$, $\eta_p^2 = .04$ (see Figure 2).

Moral Criterion Judgments

Our hypothesis that higher ToM would be associated with judgments that moral transgressions are inalterably wrong and independent of authority and rules was supported. As expected, the main effect of ToM was significant for moral criterion judgments, $F(1, 103) = 4.70$, $p = .03$, $\eta_p^2 = .04$. More advanced ToM was associated with more mature criterion judgments (e.g., greater rule and authority independence and inalterability), $\beta = .20$. Consistent with the hypothesis that for children high in ToM, higher empathy would be associated with more mature criterion judgments, the main effect of empathy was not significant, $p = .22$, but the Empathy × ToM interaction was, $F(1, 103) = 4.20$, $p = .04$, $\eta_p^2 = .04$. To interpret the interaction, we computed high- and low-centered ToM variables at +1 and −1 *SD* from the composite ToM scores. Contrary to our predictions, however, simple slope analyses (see Figure 3) indicated that for preschoolers lower (but not higher, $p = .51$) in ToM, higher empathy was associated with more mature moral criterion judgments, $F(1, 102) = 7.02$, $\beta = .31$, $p = .01$, $\eta_p^2 = .06$.

Our hypotheses regarding moral harm type were not supported in the omnibus analysis of moral criterion judgments. The main effect of moral harm type was not significant, $p = .26$, and associations did not differ by harm type, all ps > .47. Although the Moral Harm × ToM interaction was not significant, we examined parameter estimates to assess our a priori hypothesis that ToM would be associated with criterion judgments of psychological harm and unfairness but not physical harm.

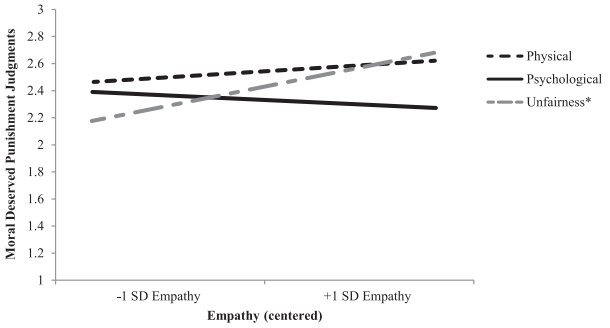

Figure 2. Empathy × Moral Harm Type interaction for deserved punishment judgments.
Note. *$p < .05$.

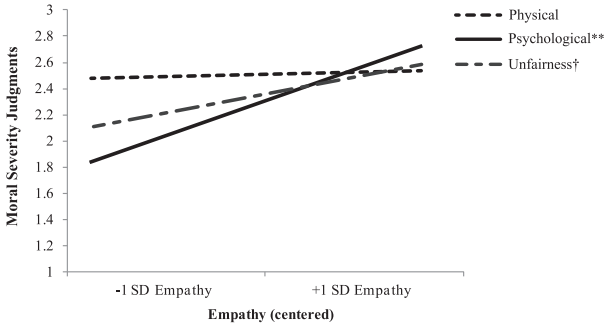

Figure 1. Empathy × Moral Harm Type interaction for severity judgments.
Note. †$p < .10$. **$p < .01$.

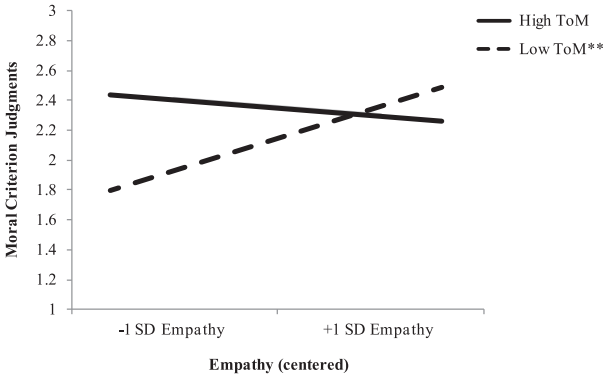

Figure 3. Empathy × Theory of Mind interaction for moral criterion judgments.
Note. ToM = theory of mind. **$p < .01$.

Unexpectedly, higher ToM was significantly associated with higher criterion judgments of physical harm, $t(103) = 2.29$, $\beta = .25$, $p = .02$, $\eta_p^2 = .05$, marginally associated with unfairness, $t(103) = 1.94$, $\beta = .20$, $p = .08$, $\eta_p^2 = .03$, and not significantly associated with psychological harm, $p = .19$.

However, the ToM × Empathy interaction was significant for criterion judgments about psychological harm, $t(103) = 2.01$, $\beta = -.28$, $p = .04$, $\eta_p^2 = .04$, but not physical harm or unfairness, ps = .17, .10, respectively. To interpret the interaction, we computed high- and low-centered empathy variables at +1 and −1 *SD*. Simple slope analyses indicated that for preschoolers lower (but not higher, $p = .31$) in empathy, higher ToM was associated with more mature criterion judgments about psychological harm, $F(3, 105) = 4.40$, $\beta = .29$, $p = .01$, $\eta_p^2 = .05$.

Moral Versus Conventional Judgments

To determine whether the obtained associations were specific to young children's evaluations of moral issues or applied to social judgments more broadly, we next examined children's judgments of moral and conventional transgressions. As type of moral harm was not of interest here, we calculated three composite moral severity, deserved punishment, and criterion judgment variables by averaging those judgments across the three forms of moral harm. We again employed repeated measures ANCOVA with empathy, ToM, and their interaction as covariates, but this time with transgression domain (moral, conventional) as the within-subject factor. Consistent with past research, there were significant domain main effects for severity, deserved punishment, and criterion judgments, Fs(1, 108) = 24.87, 6.83, 14.22, ps < .001, $\eta_p^2 = .20, .06, .12$, respectively. Moral as compared to conventional violations were considered more serious ($M = 2.44, 2.10$), more deserving of punishment ($M = 2.47, 2.29$), and more wrong in the absence of rules and authorities and inalterable (with these three criterion judgments combined, $M = 2.31, 2.12$).

For severity judgments, there was also a main effect of child empathy, $F(1, 104) = 4.12$, $p = .04$, $\eta_p^2 = .04$. Children higher in empathy judged moral and conventional transgressions as more serious, $\beta = .13$. Although the Domain × Empathy interaction did not reach significance, $p = .34$, parameter estimates supported our a priori hypothesis that child empathy would be positively associated with moral but not conventional, $p = .13$, severity judgments, $t(104) = 1.94$, $p = .05$, $\eta_p^2 = .04$.

For judgments of deserved punishment, the main effect of transgression domain was qualified by a significant interaction with ToM, $F(1, 102) = 4.78$, $p = .03$, $\eta_p^2 = .05$. Preschoolers higher in ToM judged moral transgressions as more deserving of punishment than conventional transgressions, $F(1, 102) = 11.32$, $p < .001$, $\eta_p^2 = .10$, but preschoolers with lower ToM did not differ, $p = .83$. Contrary to expectations, empathy, as a main effect or interaction, was not associated with deserved punishment judgments, all ps > .25. For criterion judgments, no further main effects or interactions were found.

Empathic Concern Versus Personal Distress

To determine if effects were specific to empathy, the two sets of analyses were rerun with personal distress rather than empathic concern as a covariate. As predicted, none of the main effects or interactions with personal distress were significantly associated with judgments.

Finally, we examined whether similar findings were obtained if only behavioral observations *or* mother reports of empathy were used. The analyses were rerun, substituting in turn the two different empathy assessments for the composite empathy measure. Neither analysis yielded significant effects. Thus, only the composite empathy variable was sufficiently sensitive to capture the associations reported here.

Discussion

This study examined associations among empathy, ToM, and moral judgments in 3½-year-olds. This age was of particular interest as empathic responding (Hoffman, 2000), ToM (Lane et al., 2010), and moral evaluations (Smetana, Rote, et al., 2012) are all developing, resulting in a great deal of variation in children's abilities. Consistent with recent conceptualizations of morality (Malti & Latzko, 2010; Smetana et al., 2014), we found that empathy and ToM were associated with moral judgments in distinct but related ways that varied according to the type of moral harm and judgment. That is, associations differed when preschoolers were evaluating severity, deserved punishment, or different moral criteria such as whether the act's wrongness was independent of rules and authority.

As expected, we found that children judged acts causing physical harm as more serious than unfairness. Although previous research has shown that even infants (Geraci & Surian, 2011) and toddlers

are sensitive to unequal resource distribution, prefer fair distributions to unfair ones, and will share their resources to amend inequalities (Brownell, Iesue, Nichols, & Svetlova, 2013), our results are consistent with other studies showing that concrete harm is particularly salient during early childhood (e.g., Helwig et al., 1995, 2001). We found that preschoolers do view taking another's resources as wrong but not as wrong as inflicting physical harm on another.

Although severity judgments for psychological harm unexpectedly did not differ from either physical harm or unfairness, we found that associations between empathy or ToM and judgments were differentiated according to the form of moral harm. Consistent with our focal hypotheses, preschoolers higher in empathy judged psychological harm and (marginally) unfairness as more serious than their less empathic peers. Understanding that psychological harm entails negative consequences for the victim might be easier for young children who are more sensitive to the harmful psychological or emotional outcomes of immoral acts. Also as expected, empathy was not associated with severity judgments regarding physical harm, most likely because the consequences are immediately evident, even to preschoolers who are less sensitive to others' distress. In such cases, empathy may not be necessary to identify and judge the harm to be serious, even in the absence of prohibitions.

On the basis of prior research, we expected severity and deserved punishment judgments to be highly correlated (Smetana, Jambon, et al., 2012) and their associations with empathy to be similar. However, in our study, correlations between these two ratings were low to moderate, and their associations with empathy differed. Consistent with our hypotheses, higher empathy was associated with ratings of greater deserved punishment for fairness violations but not for physical harm or, unexpectedly, for psychological harm. During the preschool years, spontaneous sharing of valued resources remains challenging for children and occurs infrequently (Grusec, 1991), whereas struggles over toys with peers or siblings become increasingly common (Eckerman & Peterman, 2001). Thus, punishment may be particularly important in helping children attend to and enforce issues of fairness. Correspondingly, as more empathic preschoolers tended to judge fairness violations as more serious than their less empathic peers, they also assigned more punishment.

We predicted and found that greater ToM, assessed by false belief understanding, was associated with more mature moral criterion judgments (e.g., moral transgressions treated as more independent of rules and authority and moral rules as more unalterable). Because we did not specify the consequences of transgressions, these judgments required children to interpret the situation and imagine the likely consequences of these violations. Preschoolers who develop ToM earlier than their peers may be better at predicting what others will think and feel in alternate, hypothetical scenarios (e.g., "What if her mom said that it was okay for her to hit? Would it be okay then?"). This capacity may facilitate their understanding that the prescriptive nature of moral concepts is not contingent on authority mandates or rules and that prototypical moral transgressions are invariably wrong. It should be noted that, as would be expected among children at the ages studied here (Wellman & Liu, 2004), only a small percentage of children in our study passed both of the ToM tasks used here. Although our participants were not highly skilled in understanding other minds, their scores reflected a great deal of variability, which was useful in testing study hypotheses, particularly regarding interactions between empathy and ToM.

Contrary to our expectations, parameter estimates indicated that ToM was significantly associated with criterion judgments about physical rather than psychological harm and unfairness. Although ToM was not associated with criterion judgments about psychological harm, individual differences in empathy may help explain this unanticipated finding. Associations between preschoolers' empathy and ToM have varied across studies (see Malti, Gasser, & Gutzwiller-Helfenfinger, 2010), and in our sample, these variables were uncorrelated. Our findings suggest that empathy is particularly important for evaluations of psychological harm, underscored by the significant interaction between ToM and empathy predicting criterion judgments about psychological harm but not other types of moral harm. Higher ToM was associated with more mature criterion judgments about psychological harm only for less empathic preschoolers. Therefore, children high on empathy may demonstrate more advanced judgments about psychological harm generally; however, preschoolers lower on empathy but relatively high in ToM may apply their advanced mental knowledge to psychological harm criterion judgments. The association between preschoolers' ToM and moral criterion judgments did not differ according to the form of moral harm in the ANCOVAs, however, so this finding should be interpreted with caution.

Nonetheless, consistent with this interpretation, we found that among preschoolers lower (but not higher) in ToM, higher empathy was associated with more mature moral criterion judgments. We hypothesized that the association between empathy and moral criterion judgments would vary according to ToM understanding, but we expected the relation to be strongest in children higher in ToM. Instead, the results suggest that children with less advanced ToM may rely more on affective information such as the perceived presence or absence of distress when evaluating acts. Children can judge immoral acts by observing their negative consequences for others and experiencing these situations as emotionally salient and thus might not require ToM for these judgments (Arsenio & Ford, 1985). For preschoolers higher in ToM, empathy was not associated with moral criterion judgments. If children understand the negative impact of psychological harm or unequal treatment on the victim, they may not need to experience empathy in order to recognize that the act is wrong. Together, these results highlight the importance of considering affective and cognitive processes in tandem in order to capture the unique and reciprocal associations of each with young children's developing moral judgments.

We hypothesized that empathy and ToM may facilitate early moral judgments by enabling young children to appreciate the harmful consequences of moral transgressions on others. Accordingly, we did not expect or find the same associations for social conventions, which do not inherently involve harm. Consistent with past findings (Smetana et al., 2014), moral violations (of different types) were considered more serious, more deserving of punishment, more wrong in the absence of rules and authorities, and less alterable than conventional violations. As predicted, empathy was associated with judgments of severity for moral but not conventional events. Past research demonstrates that elementary school children rate conventional infractions as emotionally neutral, whereas they associate negative emotions with immoral acts (Arsenio & Ford, 1985). Because mothers are unlikely to become significantly distressed if their children spontaneously call them by their first names or violate dinner etiquette, greater empathy may afford children little advantage in understanding why doing so may be wrong.

Unexpectedly, we also found that preschoolers with relatively low ToM did not distinguish between moral and conventional issues in judgments of deserved punishment. Although unanticipated, this result is consistent with the idea that greater ToM is associated with understanding increasingly abstract forms of harm. As both moral and conventional violations were illustrated without evident consequences and are both legitimately punishable, children lower in ToM might struggle to make nuanced distinctions and therefore believe that punishment should be indiscriminate. The associations reported in the focal moral harm analysis among empathy, ToM, and judgments of deserved punishment or moral criteria were not replicated in the social transgression analysis. This highlights the importance of precision in definitions and assessments of moral judgments, as the inclusion of different forms of social violations (i.e., conventions) obscured key associations.

Similarly, if empathy is associated with more sophisticated moral judgments because it directs attention to the victim and the harm caused, then mere aversion to distressing acts is not sufficient to promote moral understanding. Personal distress is consistently associated with directing attention and behaviors away from the victim and the upsetting event (Eisenberg et al., 2014). Accordingly, as expected and unlike empathy, personal distress was not associated with moral judgments. This suggests that self-oriented aversive responses to others' distress do not highlight the distinctive features of the social encounter nor facilitate advanced moral understanding. These results, or lack thereof, further underscore the importance of specificity and how a lack of precision in the theoretical conceptualization and corresponding methodological operationalization and assessment of the process(es) of interest could obfuscate empirical conclusions.

Limitations and Future Directions

Although the results of the current study are novel, the study has several limitations. First, the narrow age range potentially limits the generalizability of the findings. As the relations among these processes may vary in different developmental periods, further research should explore these associations both across a broader age range and longitudinally to test our hypotheses developmentally. Similarly, because we targeted preschoolers, whose verbal abilities are limited, we did not assess justifications. A few studies have obtained justifications from children as young as 4 years of age (e.g., Dunn et al., 2000), but these justifications are often undifferentiated ("it's bad" or "it's not nice"). Furthermore, this would have been difficult here, given the length and demands of the interview. Future research with school-aged children should directly

assess reasoning to better capture potential differences in moral judgment development.

Although we carefully controlled the depiction of victims' emotions in our story illustrations, psychological harm is more difficult to illustrate than other harm types. Indeed, this premise motivated our hypotheses regarding moral harm types. Although the pattern of results suggests potential minor differences in interview stimuli did not influence judgments, this possibility cannot be entirely discounted. Nonetheless, it is worth noting that although disparities in language skills and comprehensions have been found to account for some variability in ToM and moral judgments at this early age (e.g., Dunn et al., 2000), the associations reported here were obtained regardless of children's verbal intelligence. Furthermore, our utilization of multimethod assessments of empathy, including both observed empathic responses in the laboratory and mothers' reports of dispositional empathy, was a strength of this research. We found that neither maternal report nor affect coding of empathic concern alone was sufficiently sensitive to capture hypothesized associations, underscoring the importance of using multimethod measures of empathy in future research. Subsequent work also fruitfully could include physiological measures such as skin conductance or heart rate variability (e.g., Eisenberg et al., 1996) or neuroscience assessments like event-related potential or functional magnetic resonance imaging (e.g., Decety, Michalska, & Kinzler, 2012).

Finally, other moral emotions may be important for the development of early moral concepts. For example, Hoffman (2000) has suggested and Kochanska, Gross, Lin, and Nichols (2002) have demonstrated that guilt following indiscretions advances children's moral understanding. Future research should explore the role of moral emotions such as guilt or shame in developing moral judgments about harm. Despite these limitations, the findings of this study demonstrate the importance of considering both affective and cognitive processes in preschoolers' evaluations of moral issues. Because young children's social experiences are complex (Malti et al., 2010), sophisticated moral judgments may be facilitated by—but are not reducible to—understanding the thoughts and emotions of others.

References

Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, *59*, 1611–1622. doi:10.2307/1130675

Arsenio, W. F., & Ford, M. E. (1985). The role of affective information in social-cognitive development: Children's differentiation of moral and conventional events. *Merrill-Palmer Quarterly*, *31*, 1–17.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37–49. doi:10.1002/cd.96

Birch, L. L., & Billman, J. (1986). Preschool children's food sharing with friends and acquaintances. *Child Development*, *57*, 387–395. doi:10.2307/1130594

Brownell, C. A., Iesue, S. S., Nichols, S. R., & Svetlova, M. (2013). Mine or yours? Development of sharing in toddlers in relation to ownership understanding. *Child Development*, *84*, 906–920. doi:10.1111/cdev.12009

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*, 337–342. doi:10.1016/j.cognition.2009.05.006

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development*, *11*, 1–10. doi:10.1111/1467-9507.00183

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*, 853–865. doi:10.1111/1467-8624.00061

Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., & Knafo, A. (2013). Concern for others in the first year of life: Theory, evidence, and avenues for research. *Child Development Perspectives*, *7*, 126–131. doi:10.1111/cdep.12028

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*, 71–100. doi:10.1177/1534582304267187

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology*, *20*, 1053–1080. doi:10.1017/S0954579408000503

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex*, *22*, 209–220. doi:10.1093/cercor/bhr111

Dunn, J., Cutting, A. L., & Demetriou, H. (2000). Moral sensibility, understanding others, and children's friendship interactions in the preschool period. *British Journal of Developmental Psychology*, *18*, 159–177. doi:10.1348/026151000165625

Dunn, J., & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology*, *23*, 791–798. doi:10.1037/0012-1649.23.6.791

Eckerman, C. O., & Peterman, K. (2001). Peers and infant social/communicative development. In J. Bremner & A. Fogel (Eds.), *Handbook of infant development* (pp. 326–350). Malden, MA: Blackwell.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Developmental Psychology*, *32*, 195–209. doi:10.1037/0012-1649.32.2.195

Eisenberg, N., Schaller, M., Fabes, R. A., Bustamante, D., Mathy, R. M., Shell, R., & Rhodes, K. (1988). Differentiation of personal distress and sympathy in children and adults. *Developmental Psychology*, *24*, 766–775. doi:10.1037/0012-1649.24.6.766

Eisenberg, N., Spinrad, T., & Morris, A. (2014). Empathy-related responding in children. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 184–207). New York, NY: Taylor & Francis. doi:10.4324/9780203581957.ch9

Findlay, L. C., Girardi, A., & Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly*, *21*, 347–359. doi:10.1016/j.ecresq.2006.07.009

Flavell, J. H., Mumme, D. L., Green, F. L., & Flavell, E. R. (1992). Young children's understanding of different types of beliefs. *Child Development*, *63*, 960–977. doi:10.2307/1131247

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*, 1012–1020. doi:10.1111/j.1467-7687.2011.01048.x

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26–37. doi:10.2307/1130386

Grusec, J. E. (1991). Socializing concern for others in the home. *Developmental Psychology*, *27*, 338–342. doi:10.1037/0012-1649.27.2.338

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*, 379–400. doi:10.1080/02699938908412713

Helwig, C. C., Hildebrandt, C., & Turiel, E. (1995). Children's judgments about psychological harm in social context. *Child Development*, *66*, 1680–1693. doi:10.1111/j.1467-8624.1995.tb00958.x

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's judgments of psychological harm in normal and noncanonical situations. *Child Development*, *72*, 66–81. doi:10.1111/1467-8624.00266

Hepach, R., Vaish, A., & Tomasello, M. (2013). Young children sympathize less in response to unjustified emotional distress. *Developmental Psychology*, *49*, 1132–1138. doi:10.1037/a0029501

Hoffman, M. (2000). *Empathy and moral development. Implications for justice and caring*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511805851.003

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119*, 197–215. doi:10.1016/j.cognition.2011.01.006

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

Lagattuta, K., & Weller, D. (2014). Interrelations between theory of mind and morality. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 385–407). New York, NY: Taylor & Francis. doi:10.4324/9780203581957.ch18

Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., & Kerr, D. C. R. (2010). Theory of mind and emotion understanding predict moral development in early childhood. *British Journal of Developmental Psychology*, *28*, 871–889. doi:10.1348/026151009x483056

Malti, T., Gasser, L., & Gutzwiller-Helfenfinger, E. (2010). Children's interpretive understanding, moral judgments, and emotion attributions: Relations to social behaviour. *British Journal of Developmental Psychology*, *28*, 275–292. doi:10.1348/026151009X403838

Malti, T., & Latzko, B. (2010). Children's moral emotions and moral cognition: Towards an integrative perspective. *New Directions for Child and Adolescent Development*, *2010*, 1–10. doi:10.1002/cd.272

Miller, P. A., Eisenberg, N., Fabes, R. A., & Shell, R. (1996). Relations of moral reasoning and vicarious emotion to young children's prosocial behavior toward peers and adults. *Developmental Psychology*, *32*, 210–219. doi:10.1037/0012-1649.32.2.210

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*, 622–646. doi:10.1111/j.1467-8624.2007.01018.x

Nucci, L., & Turiel, E. (2009). Capturing the complexity of moral development and education. *Mind, Brain, and Education*, *3*, 151–159. doi:10.1111/j.1751-228X.2009.01065.x

Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge, UK: Cambridge University Press.

Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology*, *40*, 416–442. doi:10.1177/0022022109332844

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM approach*. London: Sage.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. San Diego: JM Sattler.

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, *52*, 1333–1336. doi:10.2307/1129527

Smetana, J. G. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology*, *21*, 18–29. doi:10.1037/0012-1649.21.1.18

Smetana, J. G., & Braeges, J. L. (1990). The development of toddler's moral and conventional judgments. *Merrill-Palmer Quarterly*, *36*, 329–346.

Smetana, J., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 23–45). New York, NY: Taylor & Francis. doi:10.4324/9780203581957.ch2

Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal associations between young children's developing moral judgments and theory of mind. *Developmental Psychology*, *48*, 1144–1155. doi:10.1037/a0025891

Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental changes and individual differences in young children's moral judgments. *Child Development*, *83*, 683–696. doi:10.1111/j.1467-8624.2011.01714.x

Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology*, *24*, 352–357. doi:10.1037/0012-1649.24.3.352

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, *42*, 3–48. doi:10.1111/1469-7610.00701

Turiel, E. (2015). Morality and prosocial judgments. In D. Schroeder & W. Graziano (Eds.), *The oxford handbook of prosocial behavior* (pp. 137–152). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195399813.013.022

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

Wainryb, C., & Brehl, B. A. (2006). I thought she knew that would hurt my feelings: Developing psychological knowledge and moral thinking. *Advances in Child Development and Behaviour*, *34*, 131–173. doi:10.1016/S0065-2407(06)80006-6

Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology*, *37*, 373–386. doi:10.1037/0012-1649.37.3.373

Weller, D., & Lagattuta, K. (2013). Helping the in-group feels better: Children's judgments and emotion attributions in response to prosocial dilemmas. *Child Development*, *84*, 253–268. doi:10.1111/j.1467-8624.2012.01837.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. doi:10.1111/j.1467-8624.2004.00691.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. doi:10.1016/0010-0277(83)90004-5

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, *67*, 2478–2492. doi:10.2307/11316

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Affect Coding Manual

Appendix S2. Social Events Interview