Cognition 119 (2011) 197–215

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The accidental transgressor: Morally-relevant theory of mind

Melanie Killen [a,*], Kelly Lynn Mulvey [a], Cameron Richardson [a], Noah Jampol [a], Amanda Woodward [b]

[a] *Department of Human Development, University of Maryland, College Park, MD, USA*
[b] *Department of Psychology, University of Chicago, USA*

ARTICLE INFO

Article history:
Received 18 May 2010
Revised 5 January 2011
Accepted 20 January 2011
Available online 4 March 2011

Keywords:
Theory of mind
Moral judgments
Developmental social cognition
Social knowledge

ABSTRACT

To test young children's false belief theory of mind in a morally relevant context, two experiments were conducted. In Experiment 1, children ($N = 162$) at 3.5, 5.5, and 7.5 years of age were administered three tasks: prototypic moral transgression task, false belief theory of mind task (ToM), and an "accidental transgressor" task, which measured a morally-relevant false belief theory of mind (MoToM). Children who did not pass false belief ToM were more likely to attribute negative intentions to an accidental transgressor than children who passed false belief ToM, and to use moral reasons when blaming the accidental transgressor. In Experiment 2, children ($N = 46$) who did not pass false belief ToM viewed it as more acceptable to punish the accidental transgressor than did participants who passed false belief ToM. Findings are discussed in light of research on the emergence of moral judgment and theory of mind.

1. Introduction

Understanding the intentions of another person reflects a core aspect of moral judgment (Killen & Smetana, 2008; Turiel, 2006; Zelazo, Helwig, & Lau, 1996) and theory of mind (Perner, Frith, Leslie, & Leekam, 1989; Wellman, 1990; Wellman & Liu, 2004; Woodward, Sommerville, & Guajardo, 2001). For several decades, researchers in the field of moral development have demonstrated how young children, as early as 3 and 4 years of age, evaluate moral transgressions on the basis of the negative intrinsic consequences to others rather than on external consequences such as teacher mandates or punishment (Smetana, 2006; Turiel, 2006). In addition, several decades of research on children's theory of mind has documented the emergence of increasingly robust reasoning about others' mental states during the same time period (Carpendale & Lewis, 2006; Perner et al., 1989; Wellman, Cross, & Watson, 2001).

1.1. Moral judgment and theory of mind

Recently, there has been interest in whether theory of mind competence is related to understanding another's intentions regarding *morally relevant* actions (Astington, 2004; Chandler, Sokol, & Wainryb, 2000; Knobe, 2005; Lagattuta, 2005; Leslie, Knobe, & Cohen, 2006; Wellman & Miller, 2008; Zelazo et al., 1996). The foci of the studies differ but converge on the overall expectation that theory of mind and moral judgment are interrelated. What is apparent is that the way that theory of mind is assessed is fairly consistent across studies, with measures including false belief competence in childhood (most often) and measures assessing reasoning about the desires of others. The moral judgment tasks, however, reflect a wide range of measures, from punishment acceptability for transgressions, to ratings of severity of a transgression as well as the

[*] Corresponding author. Address: Department of Human Development, 3304 Benjamin Building, University of Maryland, College Park, MD 20742-1131, USA. Tel.: +1 301 405 3176 (office); fax: +1 301 405 2891.
E-mail address: mkillen@umd.edu (M. Killen).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.01.006

desire to acquire something that has been prohibited by adults. Moreover, the designs often reflect the administration of two discrepant tasks, one for moral judgment and one for theory of mind.

A small handful of studies have measured children's evaluations of tasks that involve both theory of mind competence as well as moral judgment evaluations. Early research that demonstrated children's differentiation of accidental and intentional acts, contrary to Piaget's theory (1932) about the lack of differentiation, revealed, for example, that 3–7 year olds judged that recipients identified a bad outcome (harm) to be the result of an intentional rather than accidental action, and that children would assign more blame when an act was intentional rather than accidental (Yuill & Perner, 1988). Chandler and colleagues (Carpendale & Chandler, 1996; Chandler, Sokol, & Hallett, 2001) also demonstrated that children 5–7 years old rated intentional acts as "more bad" than accidental acts. Further, Leslie, Knobe, and colleagues (2006) found that contrary to the traditional expectation that theory of mind is necessary for moral judgment, children's moral decisions regarding attributions of blame influence their interpretations of other's intentions (theory of mind). Thus, these studies provided support for the theory that young children think about the motives of others, assign blame when acts are intentional, and, at times, interpret intentionality from a moral perspective.

Thus, while existing research has demonstrated that young children distinguish between intentions and outcomes, the connections documented, to date, are rather global, and more detailed investigations are necessary to specify how these connections are made in early childhood. To bolster this point, recent neuroscience research has examined the neural underpinnings between mental state attribution and moral evaluations in accidental transgression scenarios and has revealed a well-specified region of the brain that is integral to intent attribution during act evaluation (Young & Saxe, 2009). Moreover, accumulated research with adults has revealed a strong relationship between theory of mind competence and moral judgment (Knobe, 2005; Pettit & Knobe, 2009), but, as mentioned, questions about origins and development remain to be investigated.

Most developmental studies, to date, have used only one assessment for a single scenario, with some studies focused on children's affective responses towards the target and other studies on children's assignment of blame or punishment. What is lacking is a multi-measure approach in which prototypic moral judgment and false belief tasks are administered along with an embedded morally-relevant false belief theory of mind task in a single study. This type of design would directly address developmental questions about these early social cognitive competencies.

1.2. Moral judgment research

In her review of the literature, Smetana (2006) identified a robust measure for assessing moral judgment, in which children are asked to provide social reasons for what makes an act wrong from a moral viewpoint, such as focusing on physical or psychological harm, the unfairness of the act, or the lack of equal and just treatment of others in contrast to a conventional viewpoint, such as focusing on authority mandates, punishment, or rule violations. This measure, which has served as a prototypic moral judgment assessment given its extensive empirical validation (Helwig, Tisak, & Turiel, 1990; Killen, 2007; Nucci, 2001; Smetana, 2006; Turiel, 1998, 2008), derives from social domain theory (Turiel, 2006), and has been replicated in many countries, with cross-cultural generalizability, as well as with urban and rural samples of high and low socioeconomic status (for a review see Wainryb, 2006). In this methodology, assessments are made of children's evaluations of moral transgressions, such as an act of harm, in which few other competing considerations are involved (making it "prototypic") and children are asked to judge the act as well as to provide justifications for their judgments.

1.3. False belief theory of mind competency

A "prototypic" verbal measure of theory of mind that has been used extensively in past research is the false belief task (Wimmer & Perner, 1983). This task assesses children's ability to use a person's belief state to predict his or subsequent actions when those beliefs differ from reality and from the child's own knowledge. Typically, children younger than 4–5 years of age fail this task, predicting the person's actions based on reality rather than the person's false belief (Wellman et al., 2001). This task was originally assumed to measure the onset of ability to represent others' false beliefs (Wellman, 1990). Recent findings with infants (Onishi & Baillargeon, 2005) and young children (Friedman & Leslie, 2005), however, have cast doubt on this strong conclusion about onset. Even so, it is clear that this task remains relevant to the measurement of the child's ability to recruit false belief information to reason about explicit scenarios involving intentionality (Wellman & Liu, 2004).

Prototypic false belief "theory of mind" tasks measure one's access to knowledge about the physical world, such as whether another person who did not witness a location change of an object will know where to look for it ("I know it's been moved, but X does not know it's been moved and therefore will look in the place that he/she last saw it"); by design, the task itself has limited social content. The social aspect of the competence is the realization of how other people's minds work (in contrast to how other non-social objects work) but the non-social aspect of the false belief task is the lack of a specified social relationship between the "mover" and the "owner" of the object.

Further, in the case of the location change false belief task, marbles are moved from one box to another box when another child is out of the room (the participant is asked where the returning child will look for the marbles) and no social information is provided regarding who owns the marbles, the intentions of the "mover" of the marbles, or the relationship between the "mover" and the "observer" (e.g., friends, strangers). Yet the relationship between individuals in social situations has been shown to be significantly related to young children's evaluations of acts. For example, when preschoolers are told that "child X called child Y a name" children evaluate the act as "a game" when it's between two friends and as "acting mean" when it is between two children who are not friends (Slomkowski & Killen, 1992). While the false belief task has revealed much about children's judgments of intentionality, the task itself does not require one to consider social relationships and social information, which are reflected in actual situations in which judgments of intentionality are made by children in their daily interactions.

1.4. Morally-relevant theory of mind

What happens when the false belief task involves social and morally relevant considerations? For example, if the object being displaced in the task is a highly desirable one but is destroyed during the displacement by a "mover" then the relationship between the mover and the "owner" of the object might bear on the moral judgment involved in evaluating the situation. These considerations have to do with what we refer to as "morally-relevant theory of mind" because the understanding of another's knowledge state is relevant for attributions of intentions and desires regarding inter-individual treatment.

If we compare the examples of the standard "false contents" false belief task (what will a child who has been out of the room while the contents of a box have been changed to a non-traditional object think is in the box?) with a morally relevant "false contents" task (what will a child who is a classroom helper think is in a bag that she threw out?) arguments can be formulated for two main predictions. On the one hand, the difference between a passive observer's role (what will the child think is in the box?) and a potential "transgressor's" role (what did the child who threw away another's special object that was in a bag think was in the bag?) could make a false belief theory of mind task more difficult for children because they may be distracted by the salience of a "victim" who has lost a desired object. Thus, when asked what the potential transgressor thought was in the bag children may more readily, and mistakenly, suggest that the transgressor, like the participants, knew that there was a special object in the bag when throwing it away, rather than recognizing that the transgressor thought that there was trash in the bag, which would be the correct attribution. More generally, the moral context of the task may increase the complexity of the situation and thus make the task more demanding for children. On the other hand, any change in salience in the scenario may create an easier task for children due to their own increased attention to the details of the scenario, leading them to recognize that the "transgressor" does not know all of the details of the story and that he or she has not intended harm to the victim.

In fact, transforming a false belief task into one with moral valence involves a host of new considerations as well as assessments. In a morally-relevant false belief task, there are two new roles, a victim (the one who owns the desired object and left the room during the displacement) and a potential transgressor, both roles having social and moral value. False belief competence can be measured with respect to the potential transgressor (or actor) in terms of false contents (does the actor know what is in the bag?) as well as with respect to the owner of the contents, the victim, in terms of location change, after it is displaced (where will the owner look for the object when he/she returns?). Measuring both false contents and location change provides a window into the dynamics of theory of mind knowledge in a morally relevant context. In the prototypic false belief task, socially meaningful motives are rarely ascribed to the person who transfers the object; yet it is in ongoing peer interactions that theory of mind competence is both applied and developed (Carpendale & Lewis, 2006).

Thus, assessments that probe the participants' attributions of the knowledge states in the situation (did the transgressor think it was all right to throw out the bag, and why?), evaluations of the potential transgressor's actions (do you think it was all right to throw out the bag, and why?), and assessments that measure false belief theory of mind (what did the transgressor think was in the bag? and where will the recipient look for the desired object?) will provide a multi-measure design for investigating the interrelationship between moral judgments and false belief theory of mind competence. In addition, measures regarding how participants' view the emotions of the victim about the loss of the desired object (how will X feel about losing the desired object?) and about the potential transgressor (how will X feel about the child who threw away the desired object?) provide additional information regarding the participants' social and moral interpretations of the situation.

The information obtained from embedding a false belief theory of mind task in a context with morally relevant meaning contributes to understanding the application of different types of knowledge to a problem, as well as the interrelation of these two forms of cognition. Does children's theory of mind competence look different in a situation with moral meaning? Do children's moral judgments bear on their interpretations of others' intentions? Most situations confronted by children in their daily lives involve the simultaneous activation of these forms of knowledge. Children's theory of mind competence and moral understanding are called upon when making decisions on the playground and among peers in social situations. Not surprisingly, then, is the fact that one of the most frequent source of interpersonal conflicts among children has to do with the misattribution of intentions (Crick & Dodge, 1994; Zelazo et al., 1996). Understanding the interrelation of these forms of knowledge in a realistic context was both addressed and measured in the current two experiments.

1.5. Current study

The goal of this study, then, was to directly assess children's moral judgments and false belief theory of mind competence distinctly (using prototypic assessments) as well as to measure these forms of cognition within one task, developed for this study, which involved embedding a false belief assessment within a morally relevant peer scenario and administering questions that focused on the morally relevant dimensions of the task. Our research question was whether theory of mind knowledge bears on moral judgments, and whether evaluating theory of mind competence in a morally relevant context makes it more difficult to evaluate other's knowledge states than evaluating this competence in a standard task.

We administered a prototypic *moral transgression task* (pushing someone off a swing), two prototypic *theory of mind tasks* (ToM) (location change and false contents tasks to measure false belief for both roles in the scenario), and a *morally-relevant theory of mind task* (MoToM) (see Appendix for the descriptions of the three tasks). The MoToM task involved a story in which a child leaves a concealed desired object (cupcake in a paper bag) on a table, which is thrown away by a classroom helper while the child is out of the room. Assessments focused on the interrelation of the competencies across the three tasks, including attributions of intentions of the actor, evaluations of the act, and justifications for judgments. We administered the same measures, as appropriate, across all three tasks to provide the opportunity for direct comparisons of judgments. This design introduced new assessments in the prototypic moral transgression task which included probing children's perceptions of others' intentions (did the transgressor think it was okay to push X off the swing?). Three age groups were included in this project, from 3½ to 7½ years of age. We extended the age group past 5 years and up to 7½ years due to pilot results that revealed change in responses past the typical age range at which children "attain" false belief competence (Wellman & Liu, 2004). Specifically, pilot findings indicated that the oldest children were not at ceiling for correct responses in the morally relevant scenario. Another new aspect of the design was the inclusion of reasoning measures (asking children for their open-ended response for their judgments, "Why?") for the false belief theory of mind assessment. Prior theory of mind research has not probed children's reasons for their answers. Measuring false belief theory of mind in a morally relevant context, however, necessitated adding reasoning measures to fully capture the presence or absence of moral judgments (Smetana, 1995).

1.6. Hypotheses

It was expected both that participants who lacked explicit false belief competence in the morally relevant task would attribute negative intentions to the accidental transgressor and attribute negative feelings on the part of the victim towards the accidental transgressor. While it was expected that, with age, children would recognize that the accidental transgressor did not have negative intentions, it was also expected that this pattern would hold for theory of mind competence, independent of age. That is, when a participant can correctly identify when an actor has a false belief then that participant should also be aware of a lack of negative intent on the part of the accidental transgressor, regardless of their age. Regarding the types of justifications that participants would use, it was expected that participants who did not pass the morally-relevant false belief task would use moral justifications regarding the outcome of the loss of a desired object (e.g., "he will be sad to lose his cupcake"; "she threw away his cupcake "), and that participants who passed morally-relevant false belief tasks would use justifications reflecting a recognition of the lack of negative intentions (e.g., "she didn't mean to throw the cupcake away"; "he didn't know that it was in the bag").

Concerning the prototypical false belief task and the moral transgression task separately, it was expected that age-related changes would emerge for passing the false belief task as shown in previous literature. In the morally-relevant false belief scenario, whether children's false belief knowledge would differ between the potential transgressor and the potential victim was not known (given the lack of prior research). It was expected that moral reasons such as "harm to another" would be given for responses to the moral transgression, which was not expected to reveal age-related differences, based on prior findings (Smetana, 2006). It was an open question whether age-related differences would emerge regarding attributions of negative intentions of the transgressor given the absence of prior findings with this measure for a moral transgression task item.

Two experiments were conducted. Experiment 1 was the main study which tested children's false belief theory of mind, moral judgment, and morally-relevant theory of mind with a large sample ($N = 162$). Experiment 2 was a follow up ($N = 46$) designed to address the dissociation between intent judgments and act acceptability judgments for the oldest children. Specifically, a punishment assessment was included to address whether participants who viewed the act as wrong also viewed the transgressor as deserving of punishment. It was expected that the punishment acceptability item ("Do you think [the transgressor] should get in trouble for throwing the bag away?") would more directly measure evaluation of transgressor action and intent, whereas the moral judgment item ("When [the transgressor] threw out the bag, do you think [s]he was doing something that was alright or not alright?") would measure the negative valence of the outcome.

2. Experiment 1

2.1. Method

2.1.1. Participants

Children ($N = 162$) from the suburbs of a large Mid-Atlantic city participated. The sample consisted of three age groups: 62 (30 female) 3–4 year olds ($M = 49.2$ months, $SD = 6.6$, $range = 35.9–59.8$); 62 (36 female) 5–6 year olds ($M = 71.5$ months, $SD = 7.9$, $range = 60.0–83.9$); and 38 (24 female) 7–8 year olds ($M = 92.6$ months, $SD = 7.0$, $range = 84.0–106.4$). The participants came from preschools and elementary schools serving a middle- to low-income population. Parental consent was obtained for all participants.

2.1.2. Design and assessments

A within-participants design was used; all participants received all tasks in a fixed order. There were three assessment tasks, including a short warm-up task, to familiarize participants with the Likert scale. The three main assessment tasks (see Appendix) were: (1) accidental transgression, referred to as "MoToM" for "morality and theory of mind" (theory of mind false belief measurements were embedded in a morally relevant context); (2) moral transgression (prototypic); and (3) theory of mind (false belief, prototypic). For the three tasks, there were two versions for gender (names used in the stories matched the gender of the participant). There were three types of measurement items, including judgment (yes/no; all right/not all right), Likert (four-point scale) and justification (responses to "Why?").

2.1.3. Warm-up task

The purpose of the warm-up task was to familiarize participants with the four-point Likert response format (1 = not a lot; 4 = a lot). Specifically, participants were asked: "Can you show me how much you like pizza?" (the Experimenter pointed to the scale), and "Can you show me how much you like playing outside?" The scale was deemed to be a valid assessment tool, as assessed by the participants' consistency between their verbal response to the questions posed and the point on the scale which they indicated best represented their verbal response.

2.1.4. Accidental transgression (MoToM)

The MoToM accidental transgression task involved hearing a short vignette involving one child, who was a classroom helper and while cleaning up the room threw away a paper bag on a table that unbeknownst to the classroom helper had another child's special cupcake inside (the owner of the cupcake was outside). The measurements focused on the intentions of the accidental "transgressor," and the judgments and attributions of the cupcake owner (the "victim"). The exact story was the following (gender names matched the gender of the participant):

> "This is Tommy/Tammy (pointing to Tommy/Tammy) and this is Josh/Jane (pointing to Josh/Jane). Tommy has brought in a cupcake from home and is keeping it in this paper bag. Tommy puts the paper bag on the table then goes outside to play. Josh is helping the teacher clean up the classroom and sees the paper bag. Josh throws the paper bag in the trash."

Participants were asked to respond to eight items (for Likert scales: 1 = not all right and 4 = all right). The first 5 items referred to the accidental transgressor: (1) *theory of mind (false contents) of the accidental transgressor* ("What did Josh, the boy who threw out the paper bag, think was in the bag?"); (2) *accidental transgressor: intentions of the actor* ("When Josh threw out the bag, did he think he was doing something that was all right or not all right?"; *Likert*); (3) *justification for intentions of the accidental transgressor* ("Why?"); (4) *accidental transgressor: evaluation of the act* ("When Josh threw out the bag, do you think he was doing something that was all right or not all right?"; *Likert*); and (5) *justifications for evaluation of the act* ("Why?"). The next three items referred to the actions of the victim: (6) *theory of mind of (location change) of the victim* ("Now Tommy wants to eat the cupcake that he brought in from home . . .Where will Tommy look for his cupcake?"); (7) *attributions of the emotional state of the victim* ("How will Tommy feel about losing his cupcake?"); and (8) *attributions of the victim emotion towards the accidental transgressor* ("How will Tommy feel about Josh?"; *Likert*).

2.1.5. Moral transgression

The second task, referred to as the moral transgression task, presented participants with a standard prototypic moral transgression, frequently used in the literature (see Smetana, 2006); pushing someone off a swing. The exact vignette was the following:

> "This is David/Diane (pointing to David/Diane). This is Martin/Mary (pointing to Martin/Mary). Diane is playing on the swings outside. Mary comes over and pushes her off the swing so that she can get on it. Diane falls down on the ground and hurts her knee."

Again, as in the first task, participants were asked to respond to standard assessments referring to both the transgressor and victim. Specifically, participants responded to six items. Four items referred to the transgressor: (1) *prototypic transgressor: intentions of the actor* ("When Mary pushed Diane, did Mary think she was doing something that was all right or not all right?"; *Likert*); (2) *justifications for the intentions of the actor* ("Why?"); (3) *prototypic transgressor: evaluation of the act* ("When Mary pushed, do you think she was doing something that was all right or not all right?"; *Likert*); and (4) *justification for evaluation of the act* ("Why?"). Two additional items referred to the victim: (5) *attributions of the emotional state of the victim* ("How will Diane feel about getting pushed?"); and (6) *attributions of the victim emotions towards the transgressor* ("How will Diane feel about Mary?"; *Likert*).

2.1.6. Theory of mind

The final two tasks presented participants with two prototypic false belief ToM vignettes, false contents and location change (Wellman & Liu, 2004).

False contents. The exact wording of the *false contents* task was the following:

> "See this box (pointing to a crayon box). This is a crayon box. Now here is Marta, she is cleaning up the classroom and puts some crackers in the empty crayon box."

There were two target questions for the *false contents* task were: (1) *contents false belief* ("When the other children come back in from playing outside, what will they think is in the crayon box?"); and (2) *own belief* ("What is really in the crayon box?"). The final question was administered to test for participants' memory of the story: (3) *memory* ("Did the children who were playing outside see Marta put the crackers in the box?").

Location change. The exact wording of the *location change* task was the following:

> "Laura is using the markers before recess over at the art table. Laura goes outside to play and the teacher, Ms. Smith, puts the markers in the cabinet."

There were two target questions for the location change task: (1) *location false belief* ("When Laura comes back inside from recess, where will she look for the markers?"); and (2) *own belief* ("Where are the markers really located?"). As above, the third item was administered to test for participants' memory of the story: (3) *memory* ("Did Laura see where Ms. Smith put the markers?").

For analyses conducted with the false belief tasks, a five-point scale was created based on responses to the two target questions (the false belief question and the own belief question) in both the false contents as well as the location change tasks (0 = none passed, 4 = all passed). Participants who passed all four questions were designated as having false belief competence, while participants who passed three or less were designated as having less than full false belief theory of mind competence. (While the "own belief" questions are often used as inclusion criteria, we used these questions to create the false belief scale to provide more than two data points. To check that these items did not reflect a lack of general cognitive understanding, however, as opposed to specific false belief knowledge, analyses were conducted using the traditional assessments for false belief by excluding participants who did not pass the "own belief" check questions ($n = 25$) and there were no significant differences between the two samples. Thus the full sample was included for the analyses using the complete scale.)

Further, participants who failed the memory check were excluded from analyses; 15% of the participants were excluded due to the memory check failures (all of these children were in the youngest age group; 3–4 year olds). Analyses conducted to compare false belief competence in the prototypic and MoToM tasks used the contents false belief and the location false belief items from the respective false belief ToM tasks.

2.1.7. Procedure

Trained research assistants individually interviewed participants in a quiet room, with sessions lasting approximately 25–30 min. Participants were presented with separate picture cards to aid comprehension; the picture cards were images of the objects in the story rather than one picture reflecting the entire scenario. Specifically, for the MoToM task, participants were shown the following six pictures: two child picture cards, a trash can, a bag, a cupcake, and a table. For the moral transgression task, participants were shown the following picture cards: two child picture cards, and a swing set. For the ToM, false contents task: a child protagonist, crackers, and a crayon box. Finally, for the ToM, location change task: a child protagonist, a teacher, a table, a cabinet, and markers. Interviewers referred to the picture cards during the administration of the tasks. Pictures of the children were designed to be neutral in expression.

2.1.8. Coding and reliability

Participants' justifications were coded by using coding categories used in the literature (Ardila-Rey & Killen, 2001; Killen & Smetana, 1999) as well as based on the results of extensive pilot study. The coding system comprised six categories, including: (1) *harm* (e.g., "She will get hurt if she pushes her down"); (2) *negligence* (e.g., "He should have looked in the bag before he threw it away"); (3) *lack of negative intent* (e.g., "she didn't know the cupcake was in the bag"); (4) *social-conventional* (e.g., "it's against the rules to push"); (5) *psychological* (e.g., "he was being selfish"); and, (6) *undifferentiated* (e.g., I don't know"). Because both the social-conventional as well as the psychological categories were used infrequently (<10%), analyses were conducted with three coding categories: (1) *harm*, (2) *negligence*, and (3) *lack of negative intent*.

The coding was conducted by two coders blind to the hypotheses of the study. On the basis of 25% of the interviews ($N = 150$ data points), inter-rater reliability was high, with Cohen's $\kappa = .85$. Less than 5% of the participants used two codes. Justifications which received a double-code were coded with a weight of .50 for each code, while justifications which received a single code were coded with a weight of 1.0.

2.2. Results

2.2.1. Plan for analysis

Analysis of variance (ANOVAs) was used to test hypotheses pertaining to the assessments. ANOVA-based statistical tests to analyze proportions were used due to our repeated measures designs (which are not easily analyzed using other approaches such as log-linear). This is because ANOVAs are robust to the problem of empty cells, whereas other data analytic procedures (e.g., log-linear models) necessitate cumbersome data manipulation to address the empty cells issue (see Posada & Wainryb, 2008, p. for a fuller explanation and justification of this data analytic approach). Further, a recent review of analytic procedures for these types of data (covering 20 years in APA psychology journals) indicated that linear models with repeated procedures (particularly ANOVA) are appropriate compared to log-linear analysis for this type of within-subjects design (see Wainryb, Shaw, Laupa, & Smith, 2001, footnote 4). Initial analyses revealed no significant effects for gender, thus gender was not included in further analyses. Follow-up analyses included Univariate ANOVAs for between-subjects effects and Bonferroni *t*-tests for within-subjects interaction effects. In cases where sphericity was not met, corrections were made using the Huynh–Feldt method.

The report of the analyses will be in the order of the prototypic moral transgression task, followed by the prototypic false belief ToM task, and then the morally-relevant theory of mind (MoToM) task. This order is used to establish children's baseline abilities before reporting the results for the complex MoToM task.

2.2.2. Prototypic moral transgressions

A Univariate ANOVA was conducted on evaluations of the prototypic moral transgression for age (3.5, 5.5, 7.5 years). Replicating previous findings, all participants viewed the act as wrong (with ratings of 1 = very bad and 2 = bad), $M = 1.26$, $SD = .55$. An age difference was also found. Older participants viewed the transgression as more wrong than did the younger two age groups, $F(2, 154) = 7.35$, $p < .01$, $\eta^2 = .08$ ($M_{youngest} = 1.52$, $SD = .08$; $M_{middle} = 1.34$, $SD = .08$; $M_{oldest} = 1.00$, $SD = .11$). Follow-up tests revealed that the oldest age group viewed it as more wrong than the younger two age groups (no differences between the younger two groups).

2.2.3. Prototypic theory of mind (ToM)

Analyses of the traditional false belief theory of mind tasks confirmed that the false belief tasks replicated previous research findings. A Univariate ANOVA was conducted on the proportion of participants who passed the four-item false belief ToM task for age (3.5, 5.5, 7.5 years). As expected, there were significant changes for age, with less than a quarter of the youngest children passing the test, and the vast majority of the older children fully passing the two tasks, $F(2, 141) = 40.63$, $p < .001$, $\eta^2 = .36$ ($M_{youngest} = .24$, $SD = .43$; $M_{middle} = .73$, $SD = .46$; $M_{oldest} = .97$, $SD = .16$). In addition, separate Univariate ANOVAs were conducted for each task for age, demonstrating significant age-related increases for passing the tests; for *contents false belief*, $F(2, 157) = 32.02$, $p < .001$, $\eta^2 = .29$ ($M_{youngest} = .44$, $SD = .50$; $M_{middle} = .87$, $SD = .34$; $M_{oldest} = 1.00$, $SD = .00$), and for *location false belief*, $F(2, 157) = 45.97$, $p < .001$, $\eta^2 = .36$, ($M_{youngest} = .25$, $SD = .43$; $M_{middle} = .74$, $SD = .44$; $M_{oldest} = .97$, $SD = .16$).

2.2.4. Morally-relevant theory of mind (MoToM)

We investigated whether the same age-related changes found for standard false belief theory of mind assessments would be revealed in a morally-relevant theory of mind task. To test the hypothesis that false belief theory of mind knowledge increased with age in a morally relevant scenario (MoToM), 2 separate Univariate ANOVAs were conducted on the *false contents* theory of mind assessment (cupcake, trash) and on the *location change* theory of mind assessment (trash, table) for age (3.5, 5.5, 7.5 years). As expected, there were significant increases with age for passing the *false contents* theory of mind in a morally relevant scenario, $F(2, 154) = 34.81$, $p < .001$, $\eta^2 = .31$, ($M = .29$, $SD = .45$; $M = .67$, $SD = .48$; $M = 1.00$, $SD =.00$), and for *location change* in a morally relevant scenario, $F(2, 159) = 37.35$, $p < .001$, $\eta^2 = .32$, ($M = .31$, $SD = .46$; $M = .76$, $SD = .43$; $M = .97$, $SD = .16$). Follow-up tests revealed all groups were significantly different, *ps* < .01.

2.2.5. Relations between false belief ToM and MoToM

Following confirmation that both the traditional and morally relevant ToM tasks showed the expected age-related trends, we examined relations between pass rates on both forms of false belief ToM assessments. To investigate whether participants' success on the false belief ToM tasks was stable across the false belief ToM contexts (prototypic and embedded 'MoToM'), two chi-square tests were computed by organizing the data according to the four possible patterns of pass/fail across both the location change (i.e., *location false belief* assessment (art table, cabinet) and the MoToM *location change* assessment (table, trash)) and false contents (i.e., *contents false belief* assessment (crackers, crayons) and the MoToM *false contents* assessment (cupcake, trash)) tasks. More specifically, when looking at performance on the two false contents tasks (i.e., prototypic and embedded) participants could either: (1) pass both tasks; (2) fail both tasks; (3) pass the prototypic task and fail the embedded task, or (4) pass the embedded task and fail the prototypic task (note that this same pattern of results covered all possible outcomes for the location change tasks as well). For the *false contents* tasks, more participants were stable (i.e., more participants either failed or passed both tasks than failed only one task) than unstable in their *false contents* knowledge for the two assessments, $\chi^2(3, N = 155) = 39.16$, $p < .001$. Similarly, for the *location change* tasks, more participants were stable than unstable for the *location change* knowledge for the two assessments, $\chi^2(3, N = 160) = 88.73$, $p < .001$. Thus, these results indicated that participants' false belief theory of mind knowledge was stable across contexts.

While stability across contexts was found, we also assessed the predicted direction of the stability. In order to test the hypotheses that false belief ToM assessments will be more challenging for children in a morally relevant context (MoToM) than in the prototypic false belief ToM context, chi-square tests were computed for children who passed only one of the two false contents tasks. As expected, more participants failed the MoToM *false contents* task and succeeded on the prototypic *false contents* task, $\chi^2(1, N = 37) = 11.92$, $p < .01$.

Disconfirming our hypothesis, for the false belief ToM *location change* assessments, there were no differences in the number of participants failing only the MoToM *location change* assessment, which involved understanding the false belief of the victim, and those failing only the false belief ToM *location change* assessment. Thus, MoToM *false contents*, which involved understanding the false belief of the transgressor, was more difficult than false belief ToM *false contents*, whereas *location change* was equally difficult in both contexts.

2.2.6. Attributions of intentions of actor and evaluation of act: Accidental transgressor

Without an understanding of false belief, children should have difficulty with intent considerations. In order to assess participants' false belief competence in the MoToM context, the MoToM *false contents* assessment was used, as this is the most relevant false belief ToM assessment to the scenario in question. It was hypothesized that in the MoToM context, with age, participants who passed false belief theory of mind would judge the accidental transgressor's intentions as positive, and would evaluate the action as more all right than would participants who did not pass the embedded false belief theory of mind assessment. To test this hypothesis, a 3 (age: 3.5, 5.5, 7.5) × 2 (MoToM *false contents*: pass, fail) × 2 (accidental transgression: intentions of actor, evaluation of act) ANOVA was conducted with repeated measures on the last factor. A main effect was found for Accidental Transgression, $F(1, 150) = 35.63$, $p < .001$, $\eta^2 = .19$, revealing that participants attributed positive intentions to the accidental transgressor ($M = 2.26$, $SD = .08$), but evaluated the act itself more negatively ($M = 1.63$, $SD = .08$).

An interaction effect was found for age by accidental transgression, $F(2, 150) = 4.05$, $p < .05$, $\eta^2 = .05$, revealing, as expected, that, with age, participants evaluated the accidental transgressor's intentions as well as the accidental transgression itself as more all right. As shown in Fig. 1, the youngest age group judged the intentions of the actor ($M = 1.62$, $SD = .13$) and evaluated the act ($M = 1.45$, $SD = .13$) negatively (no difference). While participants did, with age, evaluate the transgression as more all right,

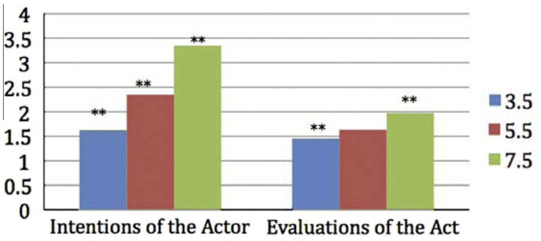

Fig. 1. Attributions of the intentions of the actor and evaluations of the act of the accidental transgressor. *Note*: All age groups differed significantly ($p < .01$) for the intentions of the actor. The youngest age group differed significantly from the oldest age group ($p < .01$) for the evaluations of the act.

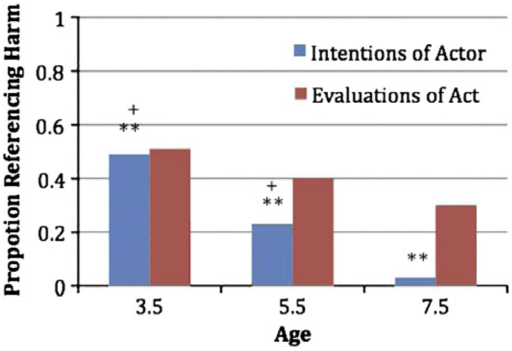

Fig. 2. Justification Referencing harm in rating the intentions of the actor and the evaluation of the act in MoToM. *Note*: For the intentions of the actor, all groups differed significantly at $p < .05$. Age groups did not differ for the evaluations of the act.

the two older age groups did not rate the intentions and evaluate the act as equally all right. The middle age group judged the accidental transgressor's intentions ($M = 2.35$, $SD = .13$) as more all right than their evaluation of the act ($M = 1.63$, $SD = .13$). Similarly, the oldest age group judged the intentions of the actor ($M = 3.35$, $SD = .16$) as more all right then how they evaluated the act ($M = 1.97$, $SD = .16$).

Follow-up tests revealed that 5.5 year olds and 7.5 year olds were significantly different in their judgment of the intentions of the actor and their evaluation of the act, *ps* < .001. Further, follow-up tests revealed that all age group comparisons were significantly different, *ps* < .01. Thus, with age, children were able to recognize the accidental nature of the transgression, but evaluated the act itself as negative.

An interaction effect was also found for MoToM *false contents* (pass, fail) by accidental transgression (intentions, evaluations), $F(1, 150) = 10.63$, $p < .01$, $\eta^2 = .06$. While children who did not pass false belief ToM judged the intentions of the actor and evaluated the act as equally all right ($M = 1.56$, $SD = .13$; $M = 1.50$, $SD = .12$, respectively), children who passed false belief ToM judged the intentions of the actor as more all right ($M = 2.72$, $SD = .11$) than they evaluated the act ($M = 1.71$, $SD = .10$). Follow-up tests revealed that children who passed false belief ToM judged the intentions of the actor and evaluated the act differently, $p < .001$. Specifically, follow-up tests confirmed that participants who passed false belief ToM judged the intentions of the actor as significantly more all right than did participants who did not pass false belief ToM, $p < .001$. This supported our hypothesis that participants who passed false belief ToM would attribute more positive intentions to the accidental transgressor than would participants without false belief ToM and revealed, similar to the age-related finding, that even children who passed false belief ToM evaluated the act to be wrong, despite their knowledge of the accidental nature of the transgression.

To confirm that false belief ToM was related to participants' judgments of the accidental transgressor's intentions and their evaluations of the act above and beyond age, a 2 (MoToM *false contents*: pass, fail) × 2 (accidental transgression: intentions, evaluation) ANOVA was conducted, with age as a covariate and with repeated measures on the last factor. This revealed a significant false belief ToM by accidental transgressor interaction effect, $F(1, 152) = 7.99$, $p < .01$, $\eta^2 = .05$. Participants who passed false belief ToM evaluated the transgressor's intentions as significantly more all right ($M = 2.66$, $SD = .14$) than how they themselves evaluated the act ($M = 1.66$, $SD = .11$). Participants without false belief ToM did not differ (follow-up test was not significant) in their judgment about the transgressor's intentions ($M = 1.63$, $SD = .13$) and the act itself ($M = 1.92$, $SD = .14$).

2.2.7. Justifications for intentions and evaluation: Accidental transgressor

To test the hypothesis that, in a MoToM context, the justifications for their judgment of the intentions of the accidental transgressor differed by age, a 3 (ages: 3.5, 5.5, 7.5) × 3 (justifications: harm, negligence, no negative intent) ANOVA, with repeated measures on the last factor revealed a significant main effect for justification, $F(2, 318) = 62.84$, $p < .001$, $\eta^2 = .28$, indicating that harm justifications were used most often ($M = .54$, $SD = .03$), followed by no negative intent ($M = .25$, $SD = .03$), and negligence ($M = .03$, $SD = .01$). As expected, a justification by age interaction effect, $F(4, 318) = 26.21$, $p < .001$, $\eta^2 = .24$, indicated that children used different justifications with age for their attributions of intentions of the actor. Follow-up tests indicated that the use of harm justifications decreased with age, with half of the youngest participants using harm reasons ($M = .49$, $SD = .05$), less than half of the middle group ($M = .23$, $SD = .05$) and virtually none of the oldest group using harm reasons ($M = .03$, $SD = .06$). Follow-up tests revealed that all age groups differed significantly ($p < .05$). Almost no participants used negligence (*Ms* = .03, .06, .00).

There was an increase in the use of "no negative intent" with age ($M = .20$ $SD = .05$; $M = .52$, $SD = .05$; $M = .91$, $SD = .07$, for 3.5, 5.5, and 7.5 year olds, respectively). Follow-up tests revealed that all age groups differed significantly, $p < .001$. As predicted, with age, harm justifications decreased, indicating that participants did not

Table 1
Justifications for the evaluations of the act by age.

Participant age (in years)	Harm		Negligence		No negative intent	
	MoToM	Prototypic moral	MoToM	Prototypic moral	MoToM	Prototypic moral
3	.51 (.06)	.44 (.06)	.02 (.04)	.01 (.01)	.13 (.05)	.02 (.02)
5	.40 (.06)	.57 (.06)	.14 (.04)	.01 (.01)	.21 (.05)	.03 (.02)
7	.30 (.08)	.88 (.07)	.18 (.05)	.00 (.01)	.41 (.06)	.00 (.02)

Note: Harm = participants' justifications that referred to harm to the victim; negligence = participants' justifications that referred to the lack of effort to avoid transgression; no negative intent = participants' justifications that referred to the lack of negative intentions on the part of the transgressor. Numbers reflect the proportion of participants justifying their judgment with the respective codes. Standard deviations are in parentheses.

evaluate the accidental transgressor's act as a moral violation; instead, with age, participants justified their answers by referring to "no negative intent." The frequent use of harm justifications by the youngest group may reflect the youngest children focusing on the outcome, and not the intention in this question, whereas the two older groups were more adept at distinguishing outcome from intent when reasoning about the intent of the transgressor. Thus, as shown in Fig. 2, references to harm decreased with age for the intentions of the actor.

A 3 (ages: 3.5, 5.5, 7.5) × 3 (justifications: harm, negligence, no negative intent) ANOVA, with repeated measures on the last factor was conducted to test whether the justifications used by participants for their own evaluations of the accidental transgressor's act varied by age. A significant main effect for justification was found, $F(2, 318) = 15.54$, $p < .001$, $\eta^2 = .08$, with most participants using harm reasons ($M = .40$, $SD = .04$), about a quarter of participants using "no negative intent" ($M = .25$, $SD = .03$) and a minority of participants cited negligence ($M = .11$, $SD = .02$) for their evaluation of the accidental transgressor's act. A justification by age interaction effect, $F(4, 318) = 3.96$, $p < .01$, $\eta^2 = .04$, indicated that children used different justifications with age (see Table 1). Use of harm justifications did not differ with age (although there was a trend) ($M = .51$, $SD = .06$; $M = .40$, $SD = .06$; $M = .30$, $SD = .08$, for 3.5, 5.5, 7.5 year olds, respectively).

As expected, references to "no negative intent" increased with age ($M = .13$, $SD = .05$; $M = .21$, $SD = .05$; $M = .41$, $SD = .06$, for ages 3.5, 5.5, 7.5). Follow-up tests revealed that the youngest age group and the middle age group were significantly different than the oldest age group, $p < .01$. Thus, as shown in Fig. 3, for both intentions of the actor and evaluations of the act, references to "no negative intent" increased with age.

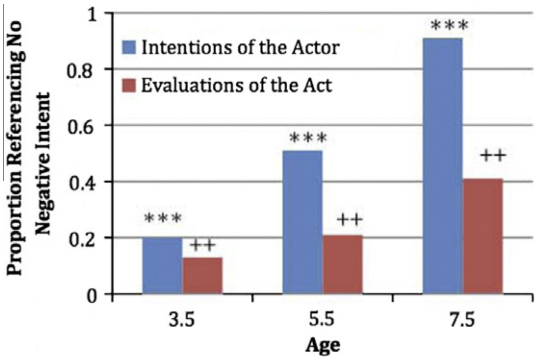

Fig. 3. Justifications referencing no negative intent in rating the intensions of the actor and the evaluation of the act in MoToM. *Note*: For the intentions of the actor, all age groups were significantly different, $p < .001$. For the evaluations of the act, the youngest age group and the middle age group were significantly different than the oldest age group, $p < .01$.

Whereas almost no participants referenced negligence when providing justifications for the transgressor's intentions, references to negligence when justifying participants' own evaluation of the transgressor's act increased with age ($M = .02$, $SD = .04$; $M = .14$, $SD = .04$; $M = .18$, $SD = .05$, for ages 3.5, 5.5, 7.5, respectively). The youngest age group was significantly different than the middle age group, $p < .05$, and the oldest age group, $p < .01$. In sum, participants who evaluated the act from their own viewpoint used harm justifications most often, reflecting their concern with the harm caused by the transgression, even if it was accidental. Additionally, the recognition that the transgressor did not have negative intentions increased with age, as predicted. Finally, and similar to the findings for "no negative intent" justifications, references to the transgressor's negligence increased with age, indicating that the oldest children recognized the lack of negative intent while simultaneously condemning the actor.

2.2.8. Victim emotions regarding the morally-relevant theory of mind task

In the MoToM scenario, virtually all participants judged that the victim would feel bad about losing the cupcake (with MoToM *false contents*, $M = 1.00$, $SD = .00$; without MoToM *false contents*, $M = .94$, $SD = .25$). While participants with false belief theory of mind viewed the intentions of the actor more positively than children without false belief theory of mind, they expected the victim to feel bad about the transgressor ($M = 1.53$, $SD = .64$), just as did children who did not pass false belief ToM ($M = 1.34$, $SD = .65$). This could be due to a belief that the victim's emotions are driven by outcome, regardless of whether the victim understands the transgressor's intent. Alternatively, it could be due to a belief that the victim misunderstood the transgressor's intentions, or that the victim focused on the transgressor's negligence.

2.2.9. Attribution of intentions of the actor and evaluation of the act: Prototypic moral transgression

The same judgments measured in the ToM and MoToM tasks were assessed for the prototypic moral transgression task in order to determine what impact understanding the nature of a transgression had on assessing traditional moral transgressions. In this instance, the traditional false belief theory of mind scale was used as a measure of false belief ToM competence, with participants who passed all four false belief ToM questions considered to have false belief ToM and participants who passed less than four questions labeled as not demonstrating false belief ToM competence.

A 3 (ages: 3.5, 5.5, 7.5) × 2 (false belief ToM scale: pass, fail) × 2 (prototypic transgression: intentions of actor, evaluation of act) ANOVA, with repeated measures on the last factor, was conducted to examine attribution of intentions of the actor and evaluation of the prototypic moral transgression. A main effect for prototypic transgression was found, $F(1, 151) = 4.86$, $p < .05$, $\eta^2 = .03$, revealing that participants viewed the intentions of the actor as less wrong ($M = 1.64$, $SD = .18$) than their own evaluation of the act itself ($M = 1.25$, $SD = .10$). Interaction effects were found for false belief ToM (pass, fail) by prototypic transgression (intentions of the actor, evaluations of the act), $F(1, 150) = 4.56$, $p < .05$, $\eta^2 = .03$, as well as for age by prototypic transgression, $F(2, 150) = 4.16$, $p < .05$, $\eta^2 = .05$.

The false belief ToM interaction effect revealed that participants who did not pass false belief ToM (failed) did not differ for the attributions of intentions of the actor and the moral evaluation of the act ($M = 1.26$, SD=.07; $M = 1.24$, $SD = .19$ for intentions of the actor and evaluation of the act, respectively). For participants who did have false belief ToM (passed), and similar to the MoToM findings, the intentions of the actor were evaluated as more positive ($M = 2.03$, $SD = .12$) than the evaluation of the act ($M = 1.26$, $SD = .33$). Participants' positive rating for the transgressor's intentions could have been due to identifying with a peer in a play situation and refraining from attributing negative intentions to him or her, or focusing on positive intentions, having just evaluated the accidental scenario, which included positive intentions. Follow-up tests revealed that children who passed false belief ToM judged the intentions and the act differently, $p < .001$. More specifically, participants who passed false belief ToM judged the intentions of the actor as significantly more all right than did participants who did not pass false belief ToM.

The age interaction effect revealed that the youngest participants judged the transgressor's intentions as less all right ($M = 1.36$, $SD = .16$) than the middle ($M = 1.92$, $SD = .14$) or oldest age group ($M = 1.65$, $SD = .48$). Follow-up tests revealed that the youngest age group attributed more negative intentions to the transgressor than did the middle age group, $p < .05$. In contrast to the age-related decrease in attribution of negative intention, older children judged the act to be less all right ($M = 1.00$, $SD = .27$) than did either of the two younger age groups (age 3.5: $M = 1.40$, $SD = .09$; age 5.5: $M = 1.35$, $SD = .08$), $ps < .001$.

To confirm that false belief ToM abilities were related to evaluations of the prototypic transgression above and beyond age, a 2 (ToM: pass, fail) × 2 (prototypic transgression: intentions of the actor, evaluation of the act) repeated measures ANOVA, with the repeated measures on the last factor, with age in months as a covariate, was conducted. An interaction effect was found, indicating differences in judgment between intentions and evaluation of the act, by false belief competence $F(1, 138) = 9.27$, $p < .01$, $\eta^2 = .06$. In line with the finding above, participants who passed false belief ToM judged the transgressor's intentions as significantly more all right ($M = 2.13$, $SD = .11$) than participants evaluated the act to be all right ($M = 1.28$, $SD = .09$), $p < .001$, whereas participants who did not pass false belief ToM judged that the transgressor's intentions ($M = 1.48$, $SD = .16$) and the act itself were similarly not all right ($M = 1.23$, $SD = .07$). Thus, though the prototypic moral transgression was judged as wrong by all participants, false belief ToM skill seems to enable participants to consider possible reasons why a transgressor may have committed a transgression and to look for positive intentions, even in actions where the intention appears quite negative. Even in situations where there is a prototypic, clear-cut transgression, false belief theory of mind skill played a role in evaluating intentions.

2.2.10. Justifications for prototypic moral judgments

A 3 (ages: 3.5, 5.5, 7.5) × 3 (justifications: harm, negligence, no negative intent) ANOVA, with repeated measures on the last factor was conducted to confirm that in a prototypic moral transgression context participants would differ in their justifications for prototypic transgressor's intentions. A main effect for justifications was found, $F(2, 314) = 104.84$, $p < .001$, $\eta^2 = .40$, indicating that harm justifications were used most often ($M = .53$, $SD = .04$), with no use of negligence ($M = .00$), and almost no use of "no negative intent" ($M = .09$, $SD = .02$). Harm reasons were used most often by all age groups ($M = .63$, $SD = .06$; $M = .42$, $SD = .06$; $M = .55$, $SD = .08$, for ages 3.5, 5.5, 7.5, respectively), and accounted for over 85% of the reasoning we assessed (.53/.62 reflects 85% of the total harm reasons used because .38 reflected other categories not assessed).

A 3 (ages: 3.5, 5.5, 7.5) × 3 (justifications: harm, negligence, no negative intent) ANOVA, with repeated measures on the last factor was conducted to test if the justifications used by participants for their evaluations of the transgressor's act varied by age. A main effect for justifications was found, $F(2, 316) = 248.81$, $p < .001$, $\eta^2 = .61$, with the majority of participants using harm reasons ($M = .63$, $SD = .04$) and very few participants referencing negligence ($M = .01$, $SD = .00$) or "no negative intent" ($M = .02$, $SD = .01$). As shown in Table 1, while in the MoToM condition references to harm decreased with age for the evaluation of the act, in the prototypic condition, references to harm increased with age (see Table 1). This is likely because children acquired false belief theory of mind with age and were thus better able to recognize the positive intentions of the transgressor, with age, in the MoToM condition. As they recognized the positive intentions in the MoToM condition, their references to harm decreased.

2.2.11. Victim emotions: Prototypic moral transgression

For the prototypic moral transgression victim emotions were evaluated as a function of false belief theory of mind. Virtually all participants felt that the victim would feel bad about being pushed off the swing (who passed false belief

ToM, $M = 1.00$, $SD = .01$; who did not pass false belief ToM, $M = .96$, $SD = .04$) and would feel bad about the transgressor (who did not pass false belief ToM, $M = 1.34$, $SD = .08$; who passed false belief ToM, $M = 1.31$, $SD = .06$).

2.3. Discussion

There were several novel findings in Experiment 1 regarding children's theory of mind knowledge, specifically false belief competence, in a morally relevant context. The task in this study provided a measure of children's attribution of mental states of a potential transgressor as well as of an owner of a desired object that was "accidentally" destroyed. With age, children's attributions of positive intentions of the actor increased. Thus, while young children at 3.5 years of age attributed negative intentions to an "accidental transgressor," it was not until 7.5 years, 2.5 years beyond the canonical "5 year old" false belief knowledge marker, that children attributed positive intentions to the actor, that is, that the actor did not mean to throw away a desired object belonging to another child in the classroom. This finding was confirmed when the analyses focused on presence or absence of false belief ToM competence; children who passed false belief ToM attributed more positive intentions to the actor than did children who did not pass false belief ToM (controlling for age). While this finding revealed the underlying competency related to attributions of positive intentions, the age-related findings provided an indication of when the transformation was occurring given that we included three age groups, 3.5, 5.5, and 7.5 years.

While all children evaluated the act as wrong, younger children evaluated the act as more wrong than did older children. Thus, younger children interpreted an "accidental transgression" as a "prototypic moral transgression" in which the negative intentions and outcomes were clear and unambiguous. Interestingly, regardless of false belief ToM ability, the majority of all children evaluated the accidental transgression as wrong. These findings demonstrated that even children who passed false belief ToM, who understood the actor's false belief in the MoToM problem, still rated his/her actions as wrong. What we do not know is whether they evaluated the act as wrong because of the negative outcome for the recipient (the owner of the cupcake), whether it was a problem of coordinating the intent and outcome information, or whether it was a result of the perception that the transgressor acted negligently (Nobes, Panagiotaki, & Pawson, 2009). One way to further understand this pattern would be to test children's judgments of punishment acceptability of the actor. Do children who evaluate the outcome as negative give more priority to outcomes than to intentions? That is, was it that these participants knew that the actor did not have negative intentions but they nonetheless evaluated the act as wrong due to the salience of the negative outcome, or was it that these participants were having trouble integrating intent into their evaluation? This latter interpretation would indicate that the participants knew that the actor did not have negative intentions but they could not bring that information to bear on their judgments about outcomes (Zelazo et al., 1996). This central issue motivated the design of Experiment 2 which was to test children's judgments about punishment acceptability, that is, should the actor be punished? Zelazo et al. (1996) showed that punishment acceptability is influenced by intention judgments. More specifically, with age, participants were more punitive when an actor intended to and succeeded in harming a victim than when the "transgressor's" intentions were positive. Additionally, new research has also shown that when children are given information about outcome, intentions and negligence, that they are influenced strongly by intentions in making punishment judgments (Nobes et al., 2009). We designed this study to measure punishment acceptability in a familiar peer interaction context. Following the findings for Experiment 2, we will discuss the results for both Experiments in light of the literature and our hypotheses.

3. Experiment 2

Follow-up assessments were administered in a second experiment in order to clarify participants' negative evaluations of the accidental transgression in Experiment 1. It was expected that even though they evaluated the act as unacceptable, participants with false belief competence would not proscribe punishment for the accidental transgressor, and that this relation would hold as a result of viewing the transgressor as having unintentionally caused harm. Further, those participants without false belief competence would proscribe punishment.

3.1. Method

3.1.1. Participants

As with Experiment 1, children ($N = 46$) residing in the suburbs of a large Mid-Atlantic city were recruited to participate, and as with Experiment 1, these participants were recruited from preschools and elementary schools serving a middle- to low-income population (these children did not participate in Experiment 1). This sample consisted of three age groups: 11 (6 female) 3–4 year olds ($M = 52.6$ months, $SD = 5.8$, *range* = 43.3–59.2); 24 (15 female) 5–6 year olds ($M = 73.3$ months, $SD = 8.2$, *range* = 60.5–83.5); and 11 (4 female) 7–8 year olds ($M = 92.8$ months, $SD = 10.6$, *range* = 84.1–109.7). Parental consent was obtained for all participants.

3.1.2. Design and assessments

As in the first experiment, a within-participants design was used; all participants received all tasks described in Experiment 1 in a fixed order. Specifically, each child was administered both the warm-up task as well as the four tasks: (1) accidental transgression (MoToM); (2) moral transgression; (3) theory of mind (false contents); and (4) theory of mind (location change). There were three types of measurement items, including judgment (yes/no; all right/not all right), Likert (four-point scale) and justification (responses to "Why?"). In addition to these main assessments, additional items were administered, which are described below.

3.1.3. Follow-up assessments for accidental and moral transgression tasks

Past work (Zelazo et al., 1996) suggests that participants' evaluations of a transgressor's actions provide a different measure of intent than ratings of the acceptability of punishment for those actions. As a result, in addition to asking participants to evaluate the accidental transgressor's actions, participants were asked to respond to the following punishment acceptability question: "Do you think Josh should get in trouble for throwing the bag away?"; Likert (0 = no punishment and 2 = a lot of punishment) (gender names matched the gender of the participant).

3.1.4. Procedure

The procedure was identical to that employed in Experiment 1.

3.1.5. Coding and reliability

Coding and the assessment of the reliability of the coding system were identical to that employed in Experiment 1.

3.2. Results for Experiment 2

Experiment 2 replicated the findings of Experiment 1, with additional assessments administered to test for punishment acceptability for the potential transgressor in the MoToM task.

3.2.1. The role of punishment

Punishment acceptability. To confirm the hypothesis that children who responded correctly about the transgressor's false belief would evaluate the transgressor in a MoToM scenario less harshly than children who did not respond correctly about the transgressor's false belief, a Univariate ANOVA was conducted on punishment acceptability, with age in months as a covariate. As expected, participants who responded correctly about the transgressor's false belief viewed it as less acceptable to punish the transgressor ($M = .73$, $SD = .16$) than did participants who did not respond correctly about the transgressor's false belief ($M = 1.61$, $SD = .25$), $F(1, 31) = 7.862$, $p < .01$, $\eta^2 = .21$.

To test the hypothesis that punishment acceptability in the MoToM context decreases with age, a Univariate ANOVA was conducted on the punishment acceptability judgment by age (3 ages: 3.5, 5.5, 7.5). Confirming our hypothesis, punishment was viewed as very acceptable by the youngest children ($M_{youngest} = 1.75$, $SD = .29$), and much less acceptable by the two older age groups ($M_{middle} = .80$, $SD = .21$, $M_{oldest} = .64$, $SD = .25$). Follow-up tests indicate that the youngest age group endorsed significantly more punishment acceptability than did the middle group ($p < .05$) and the oldest groups ($p < .01$).

Children with false belief theory of mind should be able to recognize the accidental nature of the transgression in the MoToM condition, and thus should be more likely to differentiate between punishment acceptability in a MoToM scenario and in a moral transgression scenario than would children who did not show false belief theory of mind competence. This is because children view punishment as more acceptable when a transgression has negative rather than neutral (or positive) intentions (Nobes et al., 2009). In order to test the hypothesis that children with false belief theory of mind would differentiate between punishment acceptability in a MoToM scenario and in a moral transgression scenario more than would children who did not pass false belief, a 2 (prototypic ToM: pass, fail) × 2 (story: MoToM, moral transgression) ANOVA was conducted with repeated measures on the last factor, and age in months as a covariate. The significant interaction effect, $F(1, 27) = 4.75$, $p < .05$, $\eta^2 = .15$, indicated that participants who passed false belief judged that it was more acceptable to punish a transgressor in the moral transgression scenario ($M = 1.91$, $SD = .10$) than in the MoToM scenario ($M = .69$, $SD = .20$), whereas participants who did not pass false belief did not differentiate between the scenarios, MoToM ($M = 1.45$, $SD = .29$) and moral transgression ($M = 1.80$, $SD = .14$).

In order to assess if intentions of the actor and evaluation of the act were related to rankings of punishment acceptability in a MoToM scenario, separate Univariate ANOVAs were conducted on intentions of the actor (all right, not all right) and evaluation of the act (all right, not all right) by punishment acceptability, with age as a covariate. Intentions of the actor were found to be significantly related to punishment acceptability $F(1, 34) = 5.91$, $p < .05$, $\eta^2 = .16$, with participants who judged the transgressor's intentions as not all right advocating for significantly more punishment ($M = 1.38$, $SD = .21$) than did participants who judged the transgressor's intentions as all right ($M = .65$, $SD = .18$). In contrast to attributions of intent, participants advocated for similar degrees of punishment regardless of how participants evaluated the act (all right: $M = .80$, $SD = .38$, not all right: $M = 1.00$, $SD = .15$).

For the prototypic moral transgression scenario, separate Univariate ANOVAs were conducted on intentions of the actor (all right, not all right) and evaluation of the act (all right, not all right) by punishment acceptability, with age as a covariate, in order to assess if intentions of the actor and evaluations of the act related to rankings of punishment acceptability. Unlike in the MoToM scenario, no significant differences were found for intentions of the actor (all right: $M = 1.75$, $SD = .45$, not all right: $M = 1.95$, $SD = .23$) or evaluation of the act (all right: $M = 1.87$, $SD = .35$, not all right: $M = 2.00$, $SD = .00$). Thus, while some participants judged that the transgressor in a prototypic moral transgression scenario might have had positive intentions they still believed that this transgressor should be punished for the transgression.

Justifications for punishment acceptability. In order to test the hypothesis that children without false belief theory of mind would justify punishment acceptability for a MoToM transgression using harm justifications, a Univariate ANOVA was conducted and revealed a significant effect, $F(1, 43) = 8.75$, $p < .01$, $\eta^2 = .16$. As expected, the vast majority of participants who did not pass false belief in a MoToM scenario justified punishment acceptability using harm reasons ($M = .73$, $SD = .12$) while only a minority of participants who passed false belief in a MoToM scenario used harm reasons ($M = .30$, $SD = .09$).

In order to test the hypothesis that participants who passed false belief in a MoToM scenario would reject punishment acceptability for a MoToM transgression by citing that the transgressor had no negative intentions, a Univariate ANOVA was conducted for using "no negative intent" justifications for explaining their view about punishment acceptability. As expected, participants who passed false belief in a MoToM scenario explained their view about punishment acceptability by citing that the transgressor had no negative intent whereas participants who did not pass false belief in a MoToM scenario never used this explanation ($M = .37$, $SD = .10$; $M = .00$), $F(1, 43) = 8.30$, $p < .01$, $\eta^2 = .16$.

3.2.2. Justifications across punishment acceptability and evaluation of the act: Accidental transgressor

In order to test the hypothesis that harm justifications for punishment acceptability and evaluation of the act did not differ, two separate repeated measures ANOVAs were conducted on the proportional use of harm justifications for the evaluation of the act and punishment acceptability, once for participants with morally-relevant false belief ToM, and once for participants without morally-relevant false belief ToM. Results revealed no differences in the use of harm justifications across the two assessments as a function of passing or failing morally embedded false belief competence.

It was anticipated, however, that there would exist differences in the use of "no negative intent" justifications across these two assessments for participants who had demonstrated false belief competence. In order to test the hypothesis that "no negative intent" justifications for punishment acceptability and evaluation of the act differed by false belief competence, separate ANOVAs were conducted on the proportional use of no negative intent justifications for the evaluation of the act and punishment acceptability for participants with morally-relevant false belief ToM and for participants without morally-relevant false belief ToM. Results revealed differences in justifications across assessment for participants with false belief competence $F(1, 29) = 4.619$, $p < .05$, $\eta^2 = .13$. Participants with false belief competence more frequently cited no negative intent when justifying punishment acceptability judgments than when justifying their evaluation of the act (justification, act evaluation: $M = .15$, $SD = .35$; justification, punishment acceptability $M = .37$, $SD = .49$). In contrast, participants without false belief competence did not make use of no negative intent in their justifications for their judgments ($M = .00$; $M = .00$).

3.3. Discussion

The findings for Experiment 2 confirmed our expectations that children who passed false belief theory of mind in a MoToM scenario would view it as less acceptable to punish the accidental transgressor than did participants who did not pass false belief in a MoToM scenario. Thus, the findings in Experiment 1, in which children who passed false belief in a MoToM scenario viewed the accidental transgressor as doing something "bad" may have reflected a focus on the victim's loss of the desired object (cupcake) and not moral culpability. Further, in Experiment 2, participants who passed false belief in a MoToM scenario judged that it was more acceptable to punish a transgressor in the moral transgression scenario than in the MoToM scenario, whereas participants who did not pass false belief in a MoToM scenario did not differentiate between the scenarios, revealing that the transgressors in these situations were not viewed as the same by children who passed false belief in a MoToM scenario. Participants' reasons for their judgments provided further support, with participants who passed false belief in a MoToM scenario citing no negative intent for the MoToM condition.

3.4. General discussion

How is theory of mind related to moral judgments? In the current study, we designed two experiments to investigate the developmental origins of these propositions by assessing false belief theory of mind competence embedded in a morally relevant social context. Specifically, we focused on false belief knowledge as one example of theory of mind competence, and the wrongfulness of property damage and physical harm as examples of morally relevant outcomes.

The crux of the study was to assess the extent to which false belief theory of mind knowledge was directly related to children's moral evaluations of wrongdoing. Previous studies have related false belief theory of mind competencies to moral judgment responses by comparing responses to independent tasks, reflecting each type of knowledge, or by using tasks which focused on differentiating intentions and consequences without independent assessments of moral judgment and false belief. Not only did our findings reveal new knowledge about how theory of mind (false belief) competence is related to moral judgment, but we also demonstrated how moral judgment bears on false belief theory of mind judgments. Recent research by Leslie, Knobe and colleagues (Knobe, 2005; Leslie et al., 2006) as well as recent neuroscience research (Saxe, Whitfield-Gabrieli, Scholz, & Pelphrey, 2009; Young, Cushman, Hauser, & Saxe, 2007) has pointed to new connections between theory of mind and moral judgments, with the novel proposition that moral judgments bear on intentionality judgments. This proposition is different from the traditional view which is that theory of mind is necessary for making a moral judgment. Our findings provide new data for both directions of influence, and how both forms of knowledge are brought to bear on social judgments and evaluations.

What was new in the present study, in fact, was that in addition to assessing children's false belief theory of mind competence and moral judgment, we analyzed whether incorporating moral components into the standard false belief task changed either the theory of mind competency or the moral judgment. Our task was referred to as MoToM (morally-relevant theory of mind), and involved an accidental transgression scenario in which one child accidentally destroys a desired object (cupcake) of another child. Incorporated into the task was the possibility of measuring a child's perception of the transgressor's false belief (false contents; what did X think was in the bag?) and the victim's false belief (location change: where will Y look for the bag?). We included a range of measures that differentiated attributions of intentions of the accidental transgressor from the participants' moral evaluation of the act as well as false belief theory of mind knowledge and emotion attributions. The same assessments were administered across three tasks (false belief theory of mind, moral transgression, and "morally-relevant theory of mind") which enabled us to compare competencies across tasks, generating a number of new findings. These measures enabled us to investigate the relationships between theory of mind (false belief) knowledge and moral judgment not previously examined, and to test children's false belief knowledge in a morally relevant social context, one that involved a potential transgressor and victim.

Most centrally, children found it challenging to identify a protagonist's false belief when it led to a moral violation, and more so than when it pertained to action predictions made outside of the moral context. This was indicated by the finding that children who passed the false contents assessment (did he/she know what was in the bag/box?) in the standard false belief ToM task were less likely to pass it in the MoToM task, which involved a transgressor. Children who passed the location change assessment (did he/she know where to look for the cupcake/markers?) in the standard false belief ToM task, however, also did so in the MoToM Task, which involved a victim. One primary difference between the morally relevant and traditional false belief tasks is the role of the victim and transgressor in the MoToM tasks. In the MoToM task, the false contents question was asked about the transgressor, and the location change question was asked about the victim.

Why would false belief knowledge be more difficult when applying it to a transgressor than to a victim? It may be that children more readily identify with a victim than with a transgressor, and this identification enables them to more easily apply their false belief knowledge to a victim. Alternatively, it could be that the judgments about location change did not differ between the two tasks because moral judgments were not activated in either task. Accessing false contents knowledge in the MoToM story (e.g., what did the boy, who threw out the bag think was in the bag?) required the extra step of inhibiting a moral judgment which may interfere with the ability to correctly attribute mental states to the transgressor. In the traditional false contents task the item in the box holds very little salience but in the morally relevant context, the transgressor (who holds the false belief) is discarding a highly desired object. In addition, there may be a positive/negative asymmetry at work in that children may not be as concerned with an act that produces little negative outcome (traditional false contents: what arbitrary object is in the container?) but may be quite concerned with an act that produces a clear-cut negative outcome (morally-relevant false contents: what desired object is in the container?) (Leslie et al., 2006; Vaish, Grossmann, & Woodward, 2008). These alternative hypotheses require further investigation.

Morally relevant contexts are those that typically generate conflict and misattributions of intentions in actual daily life. In fact, moral judgments often influence how it is that we do or do not coordinate perspectives, that is, attributions of blame or wrongdoing can determine how an individual takes perspectives of others, and the data in the present study provide support for this directionality of influence in young children. What may often happen is that the coordination of multiple perspectives (transgressor and victim) is influenced by attributions of wrongdoing, that is, moral judgments.

Children who lacked false belief knowledge were more likely to attribute negative intentions to an "accidental transgressor" than were children who had acquired false belief knowledge. What was surprising was that being able to evaluate a peer encounter which required both false belief knowledge as well as moral knowledge was difficult for all children, even those who had false belief competence as measured by the standard task. Thus, while children who lacked false belief competence were more likely to evaluate the act (throwing the cupcake into the trash) and the intentions of the actor (what did the classroom helper intend to do?) negatively, the middle and oldest group, who had false belief theory of mind competence, differentiated the intentions of the "transgressor" ("he didn't mean to throw it away"), viewing the intentions positively, from the transgressor's act ("it was a bad thing to do"), but still maintained that the act itself was bad.

Leslie, Knobe and colleagues have conducted research on the Side-Effect Effect, which suggests that moral judgment may influence theory of mind in important ways (Knobe, 2005; Leslie et al., 2006; Pettit & Knobe, 2009). Essentially, the researchers have shown that individuals use information about the outcome of actions to make judgments about intentions. If the outcome of an action is negative then individuals are more likely to assume that the action was done intentionally; when the outcome of an action is positive then individuals are more likely to assume that the motive behind the action was neutral. While the accidental transgressor in the MoToM scenario does not have foreknowledge of the outcome, as do the actors in the above described research, the conflict for children between understanding that the accidental transgressor's intentions were positive, but still identifying the act as wrong, suggests that false belief competence may not be enough for full moral judgments. Despite this, in the present study, children with false belief theory of mind competence differentiated punishment acceptability in the MoToM condition while those without false belief theory of mind did not, suggesting that false belief theory of mind competence aided children's ability to make moral judgments.

Our findings indicate that weighing intentions involves integrating diverse forms of information. As Saxe and her colleagues have demonstrated, in the neuroscience area, the ability to focus on intentions is potentially contingent on activation in the Right Temporo-Parietal Junction (RTPJ) (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young et al., 2007). RTPJ activity may indeed become more specialized with development, thus enabling the child to more thoroughly integrate different actors' intentions when evaluating morally relevant scenarios (Saxe et al., 2009).

As well, these findings shed light on moral judgment findings and reveal the process by which children begin to connect acts and consequences, a central aspect of moral evaluations (Killen & Smetana, 2006, 2008). While all children viewed pushing someone off a swing as wrong, only the older children coordinated their evaluations of the act with their attributions of intentions for the accidental transgressor in the MoToM task. The middle group evaluated an accidental transgression as wrong but recognized that the target did not have negative intentions. Not until 7.5 years of age did the participants' act evaluations shift due to the understanding that the act was not intentional. To further disentangle the evaluation of the act from attributions of intentions for the middle and oldest group, the results from Experiment 2 revealed that children who viewed the act negatively did not judge that punishment was warranted, providing an indication that their negative evaluation of the act may have focused on the transgressor's negligence and that this focus was distinct from their moral evaluation of the transgressor's intention. Nonetheless, we were surprised that the 6–8 year old group were not at ceiling on all of the measures. For future research it would be interesting to include children who were older than 8 years of age to examine late-developing aspects of false belief theory of mind in the context of morally relevant peer encounters.

These findings provide insight into why it is that children may often have conflicts in peer exchanges that involve ambiguity about intentions. When children do not coordinate their evaluations of acts with their attributions of intentions then they are likely to blame a peer for wrongdoing in a situation in which there was an absence of negative intentions. Being the recipient of *unfair* negative attributions not only generates interpersonal conflicts, as has been shown extensively in the social information processing literature (Crick & Dodge, 1994), but also creates mixed messages about whether what makes an act wrong has to do with intentions or consequences. The accidental transgressor may infer that the consequences are viewed as more central then their neutral or good intentions by their peers when experiencing accusations of wrongdoing for acts that were not malevolently motivated. New lines of social cognitive developmental research could explore the coordination of attributions of intentions along with the evaluations of acts in other morally relevant contexts, such as the allocation of resources, infliction of psychological harm, or contexts of discrimination.

A number of new findings from this study provide a close examination of how children weigh different components of a morally relevant context. Overall, participants' evaluations of the transgressors' actions were linked to their ability to keep in mind what it is that the transgressor knows. In this study, participants made a distinction between transgressors who were aware of what they are doing (as measured in the prototypic moral transgression by pushing someone off a swing) and transgressors who were unaware of what they were doing (as measured in the MoToM task by an actor who unknowingly throws out a ''hidden'' cupcake in a paper bag when cleaning up the room). This was shown both in the judgment data (children who passed false belief in the MoToM scenario were less willing to punish in the MoToM scenario than they were in the moral transgression scenario) as well as the justification data (children who passed false belief in the MoToM scenario cited a lack of negative intentions when justifying their judgment more than did children who did not pass false belief in the MoToM scenario).

With age, children used different justifications to explain their judgments about the intentions of an accidental transgressor and their evaluations of the act. Older children referred to the absence of negative intentions when discussing the accidental transgressor's motives as well as when evaluating the act itself. References to harm decreased with age for both attributing intentions of the actor as well as evaluating the act. Referring to an absence of negative intentions is a form of perspective – taking in that the participant is recognizing the relevance of mental states when interpreting the moral status of an act towards another person.

While children did not refer to negligence when discussing the actor's intentions, with age, a minority of children referred to negligence when evaluating the act itself. Children who do not reference the absence of negative intentions, then, are not necessarily solely focused on outcomes. They could be evaluating the actor as doing something wrong because the actor failed to fully understand the parameters of the situation, including what else they might have been able to do to avoid the negative consequences of the act (Shultz & Wright, 1985). For example, judging that a transgressor was negligent (that he/she did not do something that he/she should have done to avoid a transgression) might provide yet another basis by which an individual can evaluate an act as unacceptable. When do children take into account what someone should have done to avoid committing a moral transgression? Adults have been shown to consider negligence in action (Finkel & Groscup, 1997; Shultz & Wright, 1985), as have children (Chandler et al., 2001; Siegal & Peterson, 1998). Additionally, recent research suggests that children and adults do use information about an actor's negligence in making moral judgments and assessments of punishment acceptability (Nobes et al., 2009). What role, if any, do children ascribe to victims whose carelessness puts them in harm's way? This question, similar to the questions addressed in this article, revolve around the intersection between theory of mind (e.g., what does someone know?) and moral reasoning (e.g., what should someone find out before acting?) and deserves further study.

The findings for the prototypic moral transgression were novel as well. While the results replicated previous studies on children's evaluations of acts of harm towards another (''It's wrong''), with age, children expected that the transgressor thought what he/she was doing was not wrong. Children who had false belief ToM viewed the transgressor's intentions as less wrong than did children who did not pass false belief in a MoToM scenario (who did not differentiate between their own evaluation of the act and their attributions of intentions of the actor). Perhaps children at this age were identifying with the transgressor and trying to figure out whether there could be a positive interpretation of their actions. It may be that children think that the transgressor would not push someone off the swing unless they had positive intentions. Part of this novel finding has to do with the assessment that was introduced for the first time in the prototypic moral transgression task (to provide comparisons across all of the tasks). Asking a participant whether the transgressor "thought what he/she was doing was all right or not all right?" provides a direct probe of the participants' expectations about a transgressor's motives.

What is interesting is that, at first glance, the results for the prototypic moral transgression appear to be similar to the "happy victimizer" effect (Keller, Lourenco, Malti, & Saalbach, 2003), in which children view a victimizer as "happy" due to the gain of a desired outcome (such as getting to swing on the swing after pushing someone off). There are several important differences between these findings and those of the "happy victimizer" effect, however. In the present study, participants were asked whether the transgressor thought he was doing something positive or negative in contrast to asking participants how the transgressor will feel, which has been done in prior studies. In this study, the age trend was in the reverse. Instead of a decrease in attributing positive emotions to a victimizer with age, as has been shown in the prior "happy victimizer" studies, here, with age, children thought that the transgressor might have had positive intentions. Thus, there was an increase in speculating that the transgressor meant to do something positive and this was related to false belief ToM competence.

This may be a belief in the consistency between thought and action (e.g., Onishi & Baillargeon, 2005: "why would anyone do something that they thought was bad?"). It could also be related to research on informational assumptions in that there is an expectation that we act in a way that supports our belief in what is right (Wainryb, 1991). Moreover, the "happy victimizer effect" comes about with an assessment of attribution of emotions, rather than an assessment of attribution of belief in the rightness of an action. It is possible that the fixed order of the tasks produced a carry-over effect whereby participants were still anticipating positive intentions having just received a scenario with explicit positive intentions. Further research is warranted to clarify the discrepancies revealed between extant research and the results revealed in this study.

This study employed a verbal protocol that did not allow for the assessment of nonlinguistic participants (Schick, de Villiers, de Villiers, & Hoffmeister, 2007). It remains an important task to evaluate whether this link between theory of mind and moral judgment can be assessed non-linguistically through habituation paradigms or spontaneous-response paradigms (Baillargeon, Scott, & He, 2010; Hamlin, Wynn, & Bloom, 2007). Research testing false belief knowledge and helping behavior has revealed that children as young as 15 months old are able to apply false belief understanding to social situations (Buttelmann, Carpenter, & Tomasello, 2009). Whether a nonverbal task can be developed that taps both theory of mind and moral judgment is a logical next step in evaluating the linkages between the two constructs. Additionally, our oldest participants evaluated the accidental transgression as wrong, though they did advocate for less punishment than did younger children. However, this suggests that further work with this methodology could be conducted with a sample of even older children and perhaps an adult comparative sample. This would provide greater information into the developmental changes in understanding accidental transgressions.

In sum, the studies reported revealed a number of new findings about moral knowledge, false belief theory of mind, and the intersection of the two forms of knowledge. Both forms of knowledge pertain to intentionality, a foundation of all social cognition. In many cases attributing intentions to others' actions is not as straightforward as it might appear at first glance. Many aspects of the context or situation are incorporated to accurately interpret others' intentions. For children as well as adults, misattribution of intentions frequently results in conflicts and tensions that contribute to problematic social interactions and social relationships. Research in social psychology, for example, has shown that stereotyping increases in ambiguous situations (Dovidio & Gaertner, 2004), and research in developmental psychology has shown that attributions of intentions based on group membership are revealed in interracial peer encounters in which the intentions of the potential perpetrator are ambiguous (Killen, Kelly, Richardson, & Jampol, 2010). We demonstrated that applying false belief theory of mind knowledge to situations with morally relevant information is complex, and children's ability to figure this out may have important implications for their social relationships and positive peer interactions, both of which are central to the acquisition of social knowledge and social cognition throughout development.

Acknowledgements

The first author was supported by a grant from the National Science Foundation (#BCS-0840492). We acknowledge the undergraduate research assistance from Elizabeth Blumenfeld, Samantha Cibelli, Betsy Cunnane, Mara Duvra, Laurie Eisenband, Wei Fan, Shawnese Gilpin, Sonia Giron, Phyllis Guthua, Riva Kantor, Stephanie Rivero, Erika Salomon, Dennis Tom-Wigfield, Jayne Trageser, Lisa Weinberg, and Erica Zippert. We thank the graduate students, Aline Hitti, Shelby Cooley, and Megan Clark Kelly, for their research assistance, and the students, parents, and teachers who participated in this study.

Appendix A. Assessments, stimuli, and measurement item for three tasks

A.1. Morally-relevant theory of mind (false belief) task

	Morally-relevant theory of mind (MoToM)						
Assessment name	MoToM false contents (transgressor)	MoToM location change (victim)	MoToM intentions of the actor	MoToM intention justifications	MoToM evaluation of the act	MoTom act justifications	MoToM feelings about losing cupcake/MoToM feelings about transgressor
Content	Cupcake, trash	Trash, table	Cupcake destroyed	Cupcake destroyed	Cupcake destroyed	Cupcake destroyed	Cupcake destroyed
Question to participant	What is in the bag?	Where will Y look for object?	Did transgressor think it was all right to throw out the bag?	Why?	Did you think it was all right to throw out the bag?	Why?	How will X feel about losing cupcake?/How will X feel towards Y?
Construct	Belief	Belief	Intention	Intention	Act evaluation	Act evaluation	Victim emotions

A.2. Prototypic moral transgression

	Prototypic moral transgression (moral)				
Assessment name	Prototypic moral transgression – intentions of the actor	Prototypic moral transgression intention justification	Prototypic moral transgression evaluation of the act	Prototypic moral transgression act justification	Prototypic moral transgression – feelings about harm/feelings about Transgressor
Content	Pushed off swing	Pushed off swing	Pushed off swing	Pushed off swing	Pushed off swing
Question to participant	Did the transgressor think it was all right to push X off the swing?	Why?	Do you think it was all right to push X off the swing?	Why?	How will X feel about being pushed? How will X feel about Y who pushed her/him?
Construct	Intention	Intention	Act evaluation	Act evaluation	Victim emotions

A.3. Theory of mind (false belief) task

	Theory of mind (ToM) – prototypic			
Assessment name	ToM prototypic – false contents	ToM – prototypic false contents	ToM prototypic – location change	ToM prototypic – location change
Content	Crayons, crackers	Crayons, crackers	Art table, cabinet	Art table, cabinet
Question to participant	What will the children who have been out of the room think is in the box?	What is really in the box?	Where will X look for the markers?	Where are the markers really located?
Construct	Belief	Belief	Belief	Belief

References

Ardila-Rey, A., & Killen, M. (2001). Middle class Colombian children's evaluations of personal, moral, and social-conventional interactions in the classroom. *International Journal of Behavioral Development, 25*, 246–255.

Astington, J. W. (2004). Bridging the gap between theory of mind and moral reasoning. *New Directions for Child and Adolescent Development, 2004*, 63–72.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110–118. doi:10.1016/j.tics.2009.12.006.

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.

Carpendale, J., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686–1706.

Carpendale, J., & Lewis, C. (2006). *How children develop social understanding*. Oxford, UK: Blackwell Publishing.

Chandler, M. J., Sokol, B. W., & Hallett, D. (2001). Moral responsibility and the interpretive turn: Children's changing conceptions of truth and rightness. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.

Chandler, M. J., Sokol, B. W., & Wainryb, C. (2000). Beliefs about truth and beliefs about rightness. *Child Development, 71*, 91–97.

Crick, N., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–52). San Diego, CA: Academic Press.

Finkel, N. J., & Groscup, J. L. (1997). When mistakes happen: Commonsense rules of culpability. *Psychology, Public Policy, and Law, 3*, 65–125. doi:10.1037/1076-8971.3.1.65.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief-desire reasoning: Inhibition or general difficulty. *Developmental Science, 8*, 218–225.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–560.

Helwig, C. C., Tisak, M. S., & Turiel, E. (1990). Children's social reasoning and context. *Child Development, 61*, 2060–2078.

Keller, M., Lourenco, O., Malti, T., & Saalbach, H. (2003). The multifaceted phenomenon of "happy victimizers": A cross-cultural comparison of moral emotions. *British Journal of Developmental Psychology, 21*, 1–18.

Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science, 16*, 32–36.

Killen, M., Kelly, M. C., Richardson, C., & Jampol, N. S. (2010). Attributions of intentions and fairness judgments regarding interracial peer encounters. *Developmental Psychology, 46*, 1206–1213. doi:10.1037/a0019660.

Killen, M., & Smetana, J. G. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development, 70*, 486–501.

Killen, M., & Smetana, J. G. (2006). *Handbook of moral development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Killen, M., & Smetana, J. (2008). Moral judgment and moral neuroscience: Intersections, definitions, and issues. *Child Development Perspectives, 2*, 1–6. doi:10.1111/j.1750-8606.2008.00033.

Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences, 9*, 357–359.

Lagattuta, K. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development, 76*, 713–733.

Leslie, A., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science, 17*, 421–427.

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.

Nucci, L. P. (2001). *Education in the moral domain*. Cambridge, England: Cambridge University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.

Perner, J., Frith, U., Leslie, A., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60*, 688–700.

Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind and Language, 24*, 586–604. doi:10.1111/j.1468-0017.2009.01375.

Piaget, J. (1932). *The moral judgment of the child*. New York: Free Press.

Posada, R., & Wainryb, C. (2008). Moral development in a violent society: Colombian children's judgments in the context of survival and revenge. *Child Development, 79*, 882–898.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development, 80*, 1197–1209. doi:10.1111/j.1467-8624.2009.01325.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development, 78*, 376–396. doi:10.1111/j.1467-8624.2007.01004.

Shultz, T. R., & Wright, K. (1985). Concepts of negligence and intention in the assignment of moral responsibility. *Canadian Journal of Behavioral Science, 17*, 97–108.

Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34*, 332–341. doi:10.1037/0012-1649.34.2.332.

Slomkowski, C., & Killen, M. (1992). Young children's conceptions of transgressions with friends and nonfriends. *International Journal of Behavioral Development, 15*, 247–258.

Smetana, J. G. (1995). Morality in context: Abstractions, ambiguities, and applications. In R. Vasta (Ed.). *Annals of child development* (Vol. 10, pp. 83–130). London: Jessica Kinglsey Publishers.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: Lawrence Erlbaum Associates.

Turiel, E. (1998). The development of morality. In *Handbook of child psychology*. In W. Damon (Ed.). *Social, emotional, and personality development* (5th ed., Vol. 3, pp. 863–932). New York: Wiley.

Turiel, E. (2006). The development of morality. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 789–857). Hoboken, NJ: Wiley.

Turiel, E. (2008). Thought about actions in social domains: Morality, social conventions, and social interactions. *Cognitive Development, 23*, 136–154.

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.

Wainryb, C. (1991). Understanding differences in moral judgments: The role of informational assumptions. *Child Development, 62*, 840–851.

Wainryb, C. (2006). Moral development in culture: Diversity, tolerance, and justice. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 211–240). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Metaanalysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 502–517.

Wellman, H. M., & Miller, J. G. (2008). Including deontic reasoning as fundamental to theory of mind. *Human Development, 51*, 105–135.

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.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. Malle, L. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 149–169). Cambridge, MA: MIT Press.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporal-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy Sciences, 107*, 6753–6758.

Young, L., Cushman, F. A., Hauser, M. D., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*, 8235–8240.

Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47*, 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology, 24*, 358–365. doi:10.1037/0012-1649.24.3.358.

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