Moral Judgment and Its Relation to Second-Order Theory of Mind

Genyue Fu
Zhejiang Normal University

Wen S. Xiao
Zhejiang Normal University and University of Toronto

Melanie Killen
University of Maryland

Kang Lee
Zhejiang Normal University and University of Toronto

Recent research indicates that moral judgment and 1st-order theory of mind abilities are related. What is not known, however, is how 2nd-order theory of mind is related to moral judgment. In the present study, we extended previous findings by administering a morally relevant theory of mind task (an accidental transgression) to 4- to 7-year-old Chinese children (N = 79) and analyzing connections with 2nd-order theory of mind understanding. Using hierarchical multiple regression analyses, we found that above and beyond age, children’s 1st-order theory of mind and 2nd-order theory of mind each significantly and uniquely contributed to children’s moral evaluations of the intention in the accidental transgression. These findings highlight the important roles that 1st- and 2nd-order theory of mind play in leading children to make appropriate moral judgments based on an actor’s intention in a social situation.

Keywords: moral judgment, theory of mind, first-order theory of mind, second-order theory of mind, accidental transgression

When making moral judgment about an individual’s action, one needs to recruit and integrate information about both the actor’s intent and the consequence of the action (Piaget, 1932; Turiel, 2006; Zelazo, Helwig, & Lau, 1996). In some situations, consequences and intentions may conflict with each other. For example, someone with a benevolent intention to help may end up causing harm by accident. This person’s action should be judged as less morally negative than that of another person with a malevolent intention whose action unintentionally results in a beneficial outcome. Following early research testing Piaget’s theories about intentions and outcomes in early social judgments, there has been increased recognition that reasoning about intentions and beliefs, which is referred to as theory-of-mind (ToM) understanding, may be interrelated with moral judgments (e.g., Chandler, Sokol, & Hallett, 2001; Leslie, Knobe, & Cohen, 2006; Wainryb & Brehl, 2006; Wellman & Miller, 2008; Young, Cushman, Hauser, & Saxe, 2007).

In these studies, it has been suggested that a more advanced understanding of others’ mental states is necessary for more mature moral judgments, but little research has tested this hypothesis directly or examined how development in theory-of-mind (ToM) understanding contributes to the development of moral judgment. Building upon the existing findings, the present study aimed to bridge a significant gap in the literature by examining the interrelationships between preschool children’s moral judgments of accidental and prototypic transgressions and their mental state knowledge in general and second-order ToM understanding specifically.

Mental state knowledge covers a vast area of social cognitive competence. For example, children assign blame and make less favorable evaluations of people who have caused negative outcomes even though the actions they caused are known to be unintended, unforeseeable, or due to negligence (Nelson-Le Gall & Gumerman, 1984; Nobes, Panagiotaki, & Pawson, 2009; Shultz & Wright, 1985; Shultz, Wright, & Schleifer, 1986; Yuill & Perner, 1988). Moreover, when young children judge an action’s acceptability, their judgments often depend on whether the outcome is positive or negative regardless of the intention underlying the act (Helwig, Zelazo, & Wilson, 2001; Zelazo et al., 1996).

All these findings taken together suggest that young children’s evaluations of social situations are different from older children’s judgments with respect to the use of intention. As such, the realization that actions should not only be judged according to their consequences but also by the intentions underlying the actions is a significant developmental milestone. As children mature, they become better able to use intention information and consider both outcomes and intentions when making moral judgments (Kil-
len & Smetana, in press; Zelazo et al., 1996). With increased age, children show greater reliance on mental-state information in their judgments of the wrongfulness of acts than in their judgment of deserved punishment (Cushman, Shetkoff, Wharton, & Carey, 2013). Thus, integrating a transgressor’s intention into moral judgment is particularly challenging for children.

Very few studies have directly investigated the developmental relationship of moral judgments and mental-state understanding, however, with a few exceptions (Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Smetana, Jambon, Conry-Murray, & Sturge-Apple, 2012). Smetana et al. (2012) measured the two abilities separately and found that advanced ToM abilities were associated with children’s judgments of moral transgressions as more wrong independent of authority, less permissible, and less independent of rules. The moral judgments investigated in their study, however, did not require that children draw psychological inferences about the wrongdoer’s intentions or the transgressor’s and victim’s desires. The role of children’s developing ToM in their constructions of moral judgments when the outcome of an individual’s action conflicts with the individual’s intention was unclear.

Killen et al. (2011) used a novel embedded task in which children had to use ToM knowledge in a context that involved an “accidental transgressor” (a morally relevant ToM task, or MoToM). The MoToM task adopts the traditional location-change false-belief paradigm but with several important differences. In this new task, the object being displaced by the “mover” is one that is highly desirable to the owner (e.g., a child’s special cupcake brought from home). During the displacement in the story, the mover (without knowing about the existence of the object) accidentally destroys an object that is highly desirable to the owner, thus creating a potential transgressor (e.g., the classroom helper who threw away the special cupcake by mistake) and a potential victim (e.g., the cupcake owner who lost his or her special cupcake while out of the room). Participants were first asked to predict where the owner would search for the object (the owner false-belief question), and then whether the mover knew of the existence of the object before it was moved (the mover false-belief question). They found that, with age, children who successfully completed the standard false belief tasks were able to correctly attribute the transgressor’s accidental intentions in the morally relevant false belief task, and therefore viewed the accidental transgression as less wrong and less deserving of punishment. These same children, however, were less successful at predicting the mover’s (or potential transgressor’s) false belief (what was in the bag). Younger children found it more challenging to use their false-belief understanding to reason about another’s mental state in a morally relevant context and therefore, had greater difficulty making appropriate moral judgments than did older children.

The present study aimed to extend the existing studies by linking children’s second-order theory of mind understanding to their moral judgments. Reasoning about a second-order mental state is concerned with the understanding that one person’s mental states can be embedded within mental states such as “He thinks that she thinks that . . . ” (Perner & Wimmer, 1985; for a review, see Miller, 2009, 2012). Several empirical studies implied that second-order understanding may be involved in children’s judgments about responsibility for an accident (Yuill & Perner, 1987), acts of commission and omission with negative outcomes (Hayashi, 2007), promising and lying (Maas, 2008), and children’s evaluative judgments attributed to an observer (Shiverick & Moore, 2007). In addition, Baird and Astington (2004) found that children’s scores on the second-order false-belief tasks significantly correlated with their teachers’ reports of their actual moral behaviors at school.

However, no study has directly examined whether second-order ToM understanding plays a role in children’s moral beliefs regarding accidental transgressors. Previous findings have shown that not until about 8 years of age do children successfully understand that the actions of the mover are not wrong (because the classroom helper who discarded the bag with the cupcake inside did not know the contents of the bag; Killen et al., 2011). To achieve such understanding, not only the intention of the transgressor but also the second-order mental state of the victim’s understanding about the transgressor’s belief might be involved. For example, if the owner thinks that the mover thinks there is trash inside the bag, then the transgression should be judged as less negative than the case where the owner thinks the mover knows there is a cupcake inside the bag.

In the present study, we used Killen et al.’s (2011) accidental transgression task in which a mover accidentally threw away a desirable item to measure children’s moral judgments. Also, for comparison purposes, we used the prototypic moral transgression task where a story character intentionally transgressed. We measured children’s first-order false-belief understanding in both prototypic ToM (Wellman & Liu, 2004) and MoToM tasks, similar to Killen et al. (2011). In addition, we measured children’s second-order false-belief competence. To achieve a thorough understanding of what aspects of children’s ToM abilities were recruited and applied to their moral judgments, we used hierarchical multiple regression to investigate the unique and common contributions of first- and second-order ToM to children’s moral judgments of accidental and prototypic moral transgressions.

**Method**

**Participants**

The final sample consisted of 79 children drawn from classrooms serving children who were 4–5 years old (n = 20, 10 boys; M = 4.45, SD = 0.31, range 4.00–4.80), 5–6 years old (n = 31, 15 boys; M = 5.37, SD = 0.32, range 4.75–6.00), and 6–7 years old (n = 28, 15 boys; M = 6.47, SD = 0.26, range 5.83–6.83). They were recruited from one kindergarten and one elementary school in a southeastern city in People’s Republic of China. Additional children participated but were not included due to two reasons: incorrect responses on memory check questions (n = 15, 10 in the 4- to 5-year-old group, four in the 5- to 6-year-old group, and one in the 6- to 7-year-old group) or because they had missing values on one or two of the measures due to experimenter error or lack of response (n = 19, six in the 4- to 5-year-old group, four in the 5- to 6-year-old group, and nine in 6- to 7-year-old group). All children were Han Chinese from families reflecting a range of socioeconomic status. No reliable sources were available to indicate the socioeconomic status of the children’s families. Children attended the school determined by the location of the neighborhood where their family lived. The neighborhood of the sample was highly diverse since their parents held a wide range of occu-
pational positions (e.g., farmers, factory workers, merchants, teachers, and civil servants). Informed consent was obtained from the children’s parents or legal guardians. Children’s own verbal assent was also obtained prior to their participation.

Procedure

Trained experimenters tested children individually in a quiet room at their schools. For all tasks, children were first told a story and then asked questions about the story. We used picture cards to aid children’s comprehensions of the story and to ease their memory loads. The stories were developed in English and translated into Mandarin, and children were tested in Mandarin. Five tasks were administered to them in a fixed order as described in the following sections. The whole session lasted approximately 30 min.

Warm-up task. In the warm-up task, we trained children to use a 4-point Likert scale by asking them questions about their own individual likes and dislikes. All children were able to successfully use the scale to indicate their level of preference.

MoToM task. The MoToM task consisted of a short vignette based on a translation of the following story used in Killen et al. (2011) with the names changed to be appropriate for the Chinese context (the translation and back-translation were conducted by Chinese–English bilinguals):

Xiaoping and Xiao Hong are classmates. One day Xiaoping brings a cupcake to school and puts it in a paper bag because she wants to eat it after school. Then she goes outside to play. Xiao Hong comes in to help the teacher with cleaning the room and notices the bag left on the table. Xiao Hong throws the bag in the trash.

MoToM measures and scoring. After hearing the vignette, children were asked about the transgressor’s false belief about the contents: (a) “What did Xiao Hong think the cupcake was really located now?” (Correct answer: “In the cabinet”). (b) “What does Fen think she will get for her birthday?” (Correct answer: “A toy”). A second-order false-belief question: (c) “When Xiao Hong threw out the bag, did she think she was doing something that was all right or not all right?” (Correct answer: “Trash”). The highest possible score for the second-order false-belief task is 2.

Prototypic moral transgression task. The prototypic moral transgression story (Smetana, Jambon, & Ball, 2014) was told as follows: “Xiao Mei and Lei are playing at the park. Xiao Mei is having fun on the swing. Lei wants to get on it so she pushes Xiao Mei off the swing. Xiao Mei falls down and gets hurt.” Then children were asked to evaluate Lei’s intention and behavior on a 4-point Likert scale as in the accidental transgression judgment: (a) “When Lei pushed Xiao Mei, did Lei think she was doing something that was all right or not all right?” (the evaluation of the intention of the prototypic transgression), and (b) “Do you think she was doing something that was all right or not all right?” (the evaluation of the act of the prototypic transgression).

First-order false belief tasks. Two tasks were used to assess first-order false-belief understanding. The first task was an unexpected-content false-belief task (Gopnik & Astington, 1988; Wellman & Liu, 2004). Children were told a story about Bao Bao, who puts some crackers in an empty crayon box. Next, they were asked two target questions about the false contents: (a) “What does other children, those who have never opened the box, think is inside a crayon box?” (Correct answer: “Crackers”), and (b) “What does the crayon box actually contain?” (Correct answer: “Crackers”). The final question tested children’s memories of the story: “Did other children who were playing outside see Bao Bao put the crackers in the crayon box?” (Correct answer: “No”). The second task was a change-of-location false-belief task (Wimmer & Perner, 1983). Children were told a story about Da Wei who is using markers in an art class. During the break, Da Wei goes outside to play while the teacher, Ms. Lee, cleans the table and moves the markers into the cabinet. Two target questions were asked: (a) “Where will Da Wei look for the toy when she comes back?” (Correct answer: “On the table”), and (b) “Where are the markers really located now?” (Correct answer: “In the cabinet”). The third question tested children’s memories of the story: “Did Laura see where Ms. Lee put the markers?” (Correct answer: “No”). Only when children successfully answered the memory-check questions were their scores for the target questions included for further analysis. Correct responses to each target question were coded as 1 point. The highest possible score for the first-order false-belief measures was four. In traditional scoring for false-belief tasks, correct responses on the reality questions are used as inclusion criteria. In the current study, we included the reality questions as part of the target questions to indicate full false-belief competence. The analyses conducted using the traditional assessments for false belief by excluding participants who did not pass the reality questions generated the same results as using the full sample.

Second-order false belief tasks. Two second-order false-belief stories were adapted from Astington, Pelletier, and Homer (2002) and Sullivan, Zaitchik, and Tager-Flusberg (1994). One story involved a girl (“Fen”) and her Mom. Fen’s birthday is coming up, and her mom buys a puppy for a birthday gift. In order to surprise Fen, her mom hides the puppy in the basement and tells Fen she got her a toy but not a puppy as a gift. When Fen goes down to the basement to get her rollerblades, she finds the puppy with a card saying “Happy birthday to Fen!”

Children were asked three control questions: “Did Fen see the puppy?”; “What does Fen think she will get for her birthday?”; and “Does Mom think that Fen saw the puppy?” Children were asked one target question: “What does Mom think Fen will tell her friends she is getting for her birthday?” (Correct answer: “A toy”). The second story involved Dong and Yue. Dong puts his crayon on the table and then leaves the classroom. Yue wants to play a trick on him by moving his crayon from the table to a basket. Unbeknownst to Yue, Dong is watching her through the window. Children were asked three control questions: “Can Dong see Yue?”; “Where does Dong think the crayon is?”; and “Does Yue think that Dong can see her?” Children were asked one target question: “Where does Yue think Dong will look for his crayon
when he returns to the room?” (Correct answer: “On the table”). The highest possible score for the second-order false belief measure was two.

Results

Pearson Correlations

A matrix of Pearson and partial correlations (with and without age control) among the variables as well as the means for each task are shown in Table 1 and Table 2. Several features of these correlations are noteworthy. First, first-order and second-order false-belief performance were significantly correlated (Table 1). However, this correlation was accounted for by the effect of age since the partial correlation between first-order and second-order false-belief understanding was insignificant (Table 2). Second, MoToM and second-order false belief understanding were both positively related to the moral evaluation of intention in the accidental transgression. Third, moral evaluation of the intention in the accidental transgression was positively correlated with age, suggesting that older children tended to give less negative ratings of the accidental transgressor’s intention than younger children. However, moral evaluation of the act in the prototypic transgression was negatively correlated with age, indicating that older children judged the intentional transgression more negatively than did younger children.

Hierarchical Multiple Regression Analyses

We conducted separate hierarchical regression analyses to assess the relative contributions of first-order false belief, second-order false belief, and moral ToM to moral judgments in the accidental and prototypic transgression scenarios. Children’s evaluations of the intention and the act in the accidental and prototypic transgressions were the predicted variables, respectively, for the separate analyses. For each of the models, age was entered first as a covariate, followed by the first-order false belief, second-order belief, and MoToM measures.

As shown in Table 3, in the analysis predicting the moral evaluation of the intention of the accidental transgression, Step 1 (with age as the predictor) was significant, $F(1, 77) = 13.35, p < .001$. Older children evaluated the intention less negatively than did younger children. The second step with first-order false belief, second-order false belief, and MoToM entered as the predictors was also significant, $F(4, 74) = 6.82, p < .001$. Together, all of the variables accounted for 26.90% of the variance of moral judgment. Inspection of the model shows that both second-order and MoToM provided significant, unique contributions to moral judgment of the intention above and beyond the contributions of age and first-order false-belief performance. Children who performed better in the second-order false-belief task gave more positive ratings when evaluating the intentions of the accidental transgressor, $\beta = .24, p = .045$. Children who had better performance in the MoToM task were more likely to positively evaluate the transgressor’s intention, $\beta = .35, p = .008$. However, first-order false-belief competence failed to significantly predict moral judgment of the accidental transgression when other variables were taken into account.

As shown in Table 3, in the analysis predicting the moral evaluation of the intention of the prototypic transgression, Step 1 (with age) was not significant. This suggests that children’s moral evaluations of the prototypic transgression did not differ with age. The second step with first-order false belief, second-order false belief, and MoToM entered was significant, $F(4, 74) = 2.78, p = .033$. Together, all of the variables accounted for 13.00% of the variance of moral judgment. When examining the model, we found that only MoToM significantly predicted moral evaluation of the intention of the prototypic transgression above and beyond the effects of age and all other variables. Children with better perfor-

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
<th>Age</th>
<th>False belief</th>
<th>Accidental intention</th>
<th>Prototypical intention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>First-order</td>
<td>Second-order</td>
<td>MoToM</td>
</tr>
<tr>
<td>Age</td>
<td>5.53 (0.84)</td>
<td>.56**</td>
<td>.38**</td>
<td>.60**</td>
<td>.33**</td>
</tr>
<tr>
<td>First-order false belief (Max = 4)</td>
<td>3.34 (0.86)</td>
<td>.53**</td>
<td>.38**</td>
<td>.60**</td>
<td>.33**</td>
</tr>
<tr>
<td>Second-order false belief (Max = 2)</td>
<td>0.68 (0.81)</td>
<td>.53**</td>
<td>.38**</td>
<td>.60**</td>
<td>.33**</td>
</tr>
<tr>
<td>Morally relevant theory of mind (Max = 2)</td>
<td>1.37 (0.77)</td>
<td>.50**</td>
<td>.33**</td>
<td>.40**</td>
<td>.39**</td>
</tr>
<tr>
<td>Evaluation of the intention in accidental transgression (Max = 4)</td>
<td>2.20 (1.16)</td>
<td>.38**</td>
<td>.20</td>
<td>.38**</td>
<td>.40**</td>
</tr>
<tr>
<td>Evaluation of the act in accidental transgression (Max = 4)</td>
<td>1.59 (0.81)</td>
<td>.10</td>
<td>-.02</td>
<td>.18</td>
<td>.22*</td>
</tr>
<tr>
<td>Evaluation of the intention in prototypic transgression (Max = 4)</td>
<td>1.72 (0.97)</td>
<td>.01</td>
<td>-.18</td>
<td>-.22</td>
<td>-.25*</td>
</tr>
<tr>
<td>Evaluation of the act in prototypic transgression (Max = 4)</td>
<td>1.33 (0.50)</td>
<td>-.30**</td>
<td>-.10</td>
<td>.33**</td>
<td>-.07</td>
</tr>
</tbody>
</table>

Note. $N = 79$. Likert response format (1 = not all right; 4 = all right). Evaluation of the intention in accidental transgressor: “When Xiao Hong threw out the bag, did she think she was doing something that was all right or not all right?” Evaluation of the act in accidental transgression: “When Xiao Hong threw out the bag, do you think she was doing something that was all right or not all right?” Evaluation of the intention in prototypic transgression: “When Lei pushed Xiao Mei, did Lei think she was doing something that was all right or not all right?” Evaluation of the act in prototypic transgression: “Do you think Lei was doing something that was all right or not all right?” Although the mean evaluations of the accidental transgression (act) and the prototypical transgression (act and intention) were low, adequate variability was obtained in terms of percentage of children rating these as 1 (54.0%, 55.8%, 69.0%, respectively) and the skewness (1.36, 1.27, 1.14) and kurtosis (1.28, 0.57, -0.08) of the measures. MoToM = morally relevant theory-of-mind task; Max = maximum.

$p < .05$. $** p < .01$. 

---

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
mance in MoToM tended to have less negative evaluations of the transgressor’s intention even when the transgression seemed to be done intentionally, β = .38, p = .009. However, neither first- nor second-order false belief ToM abilities had significant, unique contributions to moral judgment of the prototypic transgression.

When it comes to the moral judgment of the transgression behavior itself, none of the ToM measures made a significant contribution to the evaluation of the act, regardless of whether it was an intentional or accidental transgression. Only the effect of age was significant for the action evaluation of the prototypic transgression in Step 1, $F(1, 77) = 7.53, p = .008$. Older children had more negative evaluations of the act than did younger children, β = .30, p = .008.

### Table 2

Partial Correlations With Age as Covariate for All Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>False belief</th>
<th>MoToM</th>
<th>Accidental Intention</th>
<th>Accidental Act</th>
<th>Prototypical intention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First-order</td>
<td>Second-order</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second-order false belief</td>
<td>.12</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morally relevant theory of mind</td>
<td>.45**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation of the intention in accidental transgression</td>
<td>.02</td>
<td>.23*</td>
<td></td>
<td>.27*</td>
<td></td>
</tr>
<tr>
<td>Evaluation of the act in accidental transgression</td>
<td>.07</td>
<td>.14</td>
<td></td>
<td>.04</td>
<td>.38**</td>
</tr>
<tr>
<td>Evaluation of the intention in prototypic transgression</td>
<td>.03</td>
<td>.21</td>
<td></td>
<td>.25*</td>
<td>.46**</td>
</tr>
<tr>
<td>Evaluation of the act in prototypic transgression</td>
<td>.01</td>
<td>.07</td>
<td></td>
<td>.12</td>
<td>.38**</td>
</tr>
</tbody>
</table>

Note. $N = 79$. * $p < .05$, ** $p < .01$.

### Table 3

Hierarchical Multiple Regression Analyses Predicting Moral Evaluation of the Intention and the Act in the Accidental Transgression and Prototypical Transgression

<table>
<thead>
<tr>
<th>Dependent variable/predictor</th>
<th>$\Delta R^2$</th>
<th>$F$ for change in $R^2$</th>
<th>$\beta$</th>
<th>Part correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of the intention in accidental transgression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>.15</td>
<td>13.35**</td>
<td>.38**</td>
<td>.38</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>4.11**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>.12</td>
<td></td>
<td>.38**</td>
<td>.38</td>
</tr>
<tr>
<td>First-order false belief</td>
<td></td>
<td>1.00</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>Second-order false belief</td>
<td></td>
<td>7.66</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Morally relevant theory of mind</td>
<td></td>
<td>3.70*</td>
<td>.24*</td>
<td>.20</td>
</tr>
<tr>
<td>Evaluation of the act in accidental transgression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>.03</td>
<td>0.77</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Step 2</td>
<td>.13</td>
<td>3.70</td>
<td>.14</td>
<td>.10</td>
</tr>
<tr>
<td>First-order false belief</td>
<td></td>
<td>1.00</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Second-order false belief</td>
<td></td>
<td>7.53**</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Morally relevant theory of mind</td>
<td></td>
<td>3.70*</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Evaluation of the intention in prototypical transgression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>.00</td>
<td>0.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Step 2</td>
<td>.02</td>
<td>0.55</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>First-order false belief</td>
<td></td>
<td>1.00</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Second-order false belief</td>
<td></td>
<td>7.53**</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Morally relevant theory of mind</td>
<td></td>
<td>3.70*</td>
<td>.24</td>
<td>.20</td>
</tr>
</tbody>
</table>

Note. $N = 79$. * $p < .05$, ** $p < .01$.

### Discussion

This study investigated whether and how children’s first- and second-order ToM understanding contributed to their moral devel-
opment. Specifically, we focused on the prototypic second-order false-belief understanding and first-order false-belief understanding in a morally relevant social context. Moral evaluations of the transgressions and the transgressors’ intentions were assessed among a sample of 4- to 7-year-old Chinese children.

Extending existing findings regarding the development of moral judgments and ToM obtained mainly in North America (e.g., Killen et al., 2011; Smetana et al., 2011) to a new population, our results showed that, with age, children were less likely to assign blame to an accidental transgressor. Chinese children’s evaluations of the accidental transgression behavior itself remained unchanged with age. Although children rated the action of throwing away the desired object as wrong, their ratings of the intention of the transgressor became less negative with increased age. We interpret this to indicate that, with age, children realized that the transgressor did not mean to throw away a desired object that belonged to another child. In contrast, when judging the intention underlying the prototypic transgression, all children consistently gave negative ratings regardless of age. In addition, the moral judgment that the intentional transgressor (in the prototypic task) did something wrong became more negative with increased age.

Further, children’s abilities to attribute first-order false beliefs in a morally relevant context (MoToM) significantly contributed to the variance of their evaluative judgments about the transgressor’s intention for both accidental and prototypic transgressions. Again consistent with recent findings (Killen et al., 2011; Smetana et al., 2011), our results indicate that children’s moral judgments were related to their understanding of first-order mental states. Notably, the significant correlation between MoToM and moral evaluation of the intention was robust even after controlling for the other effects. This result indicated that children who performed better in the false-belief task in a morally relevant story made judgments that were based on a less negative view of the transgressor’s intention than those who could not correctly attribute false beliefs to the victim and the transgressor.

Surprisingly, we also found that children who were advanced in MoToM gave less negative evaluation for prototypical transgressor. Prior research on prototypic moral judgments (whether an act of harm is wrong) has not typically revealed age differences (Smetana et al., 2014). However, previous research on prototypic moral judgments has rarely measured children’s attributions of the intentions of the transgressor (the focus is more often on the evaluation of the act). We interpret this finding as indicating that the children were trying to figure out whether there could be a positive interpretation of the intentions underlying an act of harm committed by a same-age peer (e.g., “Maybe she did not know she would fall off?”). In fact, a very recent study has demonstrated that children’s evaluations of acts of harm differ depending on their expectations of the transgressors’ intentions (Smetana et al., 2014). Further research should be conducted to examine children’s attributions of intentions of transgressors in a range of prototypic moral transgression contexts.

For the first time, it was shown that children’s second-order, but not first-order false belief understanding, was significantly related to their moral evaluations of the accidental transgressor’s intention. This novel finding was robust even after partia ling out the effect of age. Children with better second-order false-belief understanding were more likely to judge the intention of the transgressor less negatively. We interpret this finding as indicating that children’s considerations of the victim’s belief about the transgressor’s false belief enabled them to recognize that the potential transgressor had positive intention (e.g., that the victim thought that the transgressor thought that there was trash inside the bag). Previous findings have shown that children can attribute responsibility appropriately based on their understanding of second-order false beliefs (e.g., Hayashi, 2007; Maas, 2008; Yuill & Perner, 1987). This ability may be related to attribution of blame in a morally relevant context. Further research should be conducted to examine these potential connections in children’s early social and moral cognition.

It is noteworthy that the current study examined two types of moral judgments, prototypical moral transgressions and more complex transgressions involving benign intentions by the transgressor. These two contexts required psychological inference for the correct conclusion to be drawn. Our results showed that only for children’s moral evaluations of the accidental transgression did second-order false-belief understanding correlate significantly. In the accidental transgression where a transgressor’s actual intention (e.g., throwing away the trash) was in conflict with the seeming intent-in-action (e.g., throwing away the desired object of the victim), we speculate that children need to coordinate multiple perspectives (those of the transgressor and the victim) to make appropriate attributions of wrongdoing. However, when there was no such conflict, as in the prototypical transgression, second-order false belief ToM failed to predict the variance of moral evaluations. Again, we propose that this distinction may be due to the relatively straightforward set of judgments required for the prototypic transgression behavior to be viewed as wrong in contrast to the complex reasoning about the accidental transgressor’s mental state. Further research should be conducted to determine what aspects of the complex situation were difficult to process and interpret for children’s assessments. This requires using a wider range of potential transgressions, most likely, as well as varying other features of the context such as the history of interactions of the participants, their relationships, and the nature of the misdeed.

Our results also shed light on the relation between first-order and second-order false belief understanding. We found that the two types of false-belief understanding were significantly correlated when the effect of age was not partialled out. When considering age as a covariate, the partial correlation between first-order and second-order false belief performance was not significant. The current literature suggests that the ability to understand second-order false belief develops later than first-order false belief (Miller, 2009). This means that one achievement is a prerequisite for the other because a first-order false belief is embedded within the second-order question. However, there has been lack of empirical evidence that supports the correlation within individuals, and our data document that the correlation between first-order and second-order false belief performance was due to the effect of age.

Nevertheless, these findings reflect several caveats that need to be considered when interpreting the present findings. One limitation concerns the correlation-based nature of the present study, which constrains any conclusions regarding the cause and effect correlations between moral judgments and ToM understanding. Longitudinal data would be an ideal way to examine these rela-
tionships more thoroughly in future research. Another limitation is that we only examined two types of moral transgressions: accidental property damage and prototypic harm. Children’s understanding of morality includes many other aspects, such as their understanding of other moral situations involving resource allocation or social exclusion. Unlike first-order ToM understanding, second-order ToM understanding was only measured in a prototypic context, another caveat of the current study. To understand the issues better, it would be important for researchers in future studies to create a moral transgression story in which the second-order false-belief questions are also embedded within the story (e.g., “What does the victim think that the transgressor thought inside the bag?”). Through this way, one could concurrently examine whether and how the first- and second-order ToM understanding in different contexts interact with each other to influence children’s moral judgments.

In summary, our findings provide evidence for the relationship between second-order ToM and moral judgment in an accidental transgression context. Specifically, the present study was able to separately identify the effects of first-order false belief, second-order false belief, and MoToM in predicting the variations of moral evaluation of accidental transgression. Second-order false belief and MoToM both play an important role in young children’s appropriated attributions of the accidental transgressor’s intention. These findings shed important light on the potential developmental connections between children’s developing moral judgments and knowledge about theory of mind and intentional mental states.

References


Received May 2, 2013
Revision received April 30, 2014
Accepted May 5, 2014