



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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



Unintentional and intentional falsehoods: The role of morally relevant theory of mind



Alexander P. D'Esterre*, Michael T. Rizzo, Melanie Killen

Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:

Received 9 January 2018

Revised 11 July 2018

Keywords:

Moral judgment

Intentionality

Theory of mind

False claims

ABSTRACT

This study investigated how theory of mind (ToM) competence is related to children's ability to differentiate between intentional and unintentional false statements regarding claims to resources. Participants (4–10 years old; $N = 122$) heard about individuals who had different access to knowledge about resource ownership when making resource claims, and they were asked to make an evaluation, attribute intentions, assign punishment, and predict the teacher's assigned punishment. Two measures of ToM were assessed: a prototypic false belief ToM assessment and a contextually embedded, morally relevant false belief theory of mind (MoToM) assessment. Children's ToM competence reliably predicted more favorable evaluations of the individual who made the unintentional false claim than of the one who did so intentionally. Furthermore, the contextually embedded MoToM assessment predicted children's responses for all of the assessments above and beyond age and prototypic ToM competence. The findings indicate that children's contextually embedded MoToM competence bears on their moral assessments of the intentions of transgressors and underscores the importance of ToM in the ability to discriminate intentional and unintentional false statements.

© 2018 Elsevier Inc. All rights reserved.

Introduction

Navigating the social world is a complex process that requires the ability to understand others' beliefs, intentions, and desires. Understanding that others have beliefs, intentions, and desires that

* Corresponding author.

E-mail address: desterre@umd.edu (A.P. D'Esterre).

may be different from one's own—referred to as theory of mind (ToM) competence—reflects an important developmental milestone in children's understanding of their social world. This ability has been shown to be necessary for everyday communication as well as the acquisition of a vast range of social skills (Hughes & Devine, 2015; Wellman & Liu, 2004). Without the ability to understand that others may have different beliefs, for example, individuals approach problems assuming that others have the same access to information as they do, which is often not the case. Interpersonal exchanges require thinking about what others know, as well as what they do not know, and acting appropriately based on these beliefs and expectations. The ability to understand others' mental states emerges early in development, with a long trajectory that stems from early in infancy into adolescence (Sodian et al., 2016; Hughes & Devine, 2015).

Research examining the intersection of mental state knowledge and moral judgment has used everyday contexts to assess how mental state knowledge is related to the evaluation of moral transgressions. One example is research on age-related changes regarding deception and lying (Bussey, 1992; Evans & Lee, 2013). Deception requires knowing that one has access to information that is not available to others and reflects a form of ToM competence. However, in the case of deception and lying, there is also another element, a moral violation, caused by the speaker's intention to deceive others (Evans & Lee, 2013). Some research has suggested that preschool-aged children struggle to distinguish between lies and truthful statements and that second and fifth graders are significantly more accurate in this capacity (Bussey, 1992). However, other studies have shown that, even at a young age, children are able to accurately discriminate between intentional lies and unintentional false statements in the form of mistakes (Siegal & Peterson, 1996, 1998). In addition, a growing body of research has shown that young children not only have the ability to recognize lies when they occur but also can discriminate between prosocial lies (e.g., "white lies") and antisocial lies and that they evaluate prosocial lies as more acceptable than antisocial lies (Bussey, 1999; Talwar, Williams, Renaud, Arruda, & Saykaly, 2016). Thus, lying is an interesting case in which an understanding of intentions intersects with moral knowledge. Children need to have some form of mental state knowledge understanding in order to recognize both that one person has access to information that another individual does not (the informational asymmetry necessary to lie) and that one individual had an intention to deceive another individual.

In fact, research has demonstrated that children's ToM abilities are connected to their comprehension and production of lies (Cheung, Siu, & Chen, 2015; Fu, Sai, Yuan, & Lee, 2017; Talwar & Lee, 2008). In a study by Peterson (1995), children between 5 and 7 years of age were presented with stories in which a character experienced a memory lapse (made an untrue statement without an intent to deceive). Following this story, participants were asked whether they believed that the character was telling a lie, was telling the truth, or was trying to tell the truth. Those children who identified that the character tried to tell the truth stated that it was not a lie, whereas those who said the character was not trying to tell the truth misidentified the unintentional false statement as a lie. Thus, in some contexts, children have trouble in differentiating false statements that are intentional from false statements that are unintentional.

In this study, we proposed that unintentional false statements may be especially difficult for children to differentiate from intentional false statements when their ToM abilities are not fully developed. In these cases, children may interpret unintentional false statements as wrong and deserving of punishment because children do not have the ToM capacities required to recognize the unintentional nature of the false statements. No research to date has directly examined how children's mental state knowledge (as assessed by various ToM assessments) is related to their ability to distinguish between intentional and unintentional false statements. Moreover, research has not yet examined the role that mental state knowledge plays regarding children's ability to distinguish between these different types of false statements, nor has it investigated the ways in which mental state knowledge predicts differences in children's moral evaluations of these two types of false statements.

Prototypic and morally embedded ToM

One aim of the current study, then, was to measure children's false belief mental state knowledge and relate it to their evaluation of intentional and unintentional false statements, specifically

regarding a false statement that resulted in negative consequences to another person, creating a moral transgression. This goal necessitated testing a sample of children between 4 and 10 years of age to capture participants who both fail and pass prototypic false belief ToM tasks (Wellman & Liu, 2004).

Understanding that false statements can have moral consequences when perpetrators intentionally engage in deception has been shown to be complex for children. This is in part because children have a difficult time in connecting intentions and outcomes. Starting with Piaget's (1932) research, a large body of research has shown that young children often focus on outcomes (negative consequences of acts such as harm inflicted and a lack of access to resources) without systematically taking one's intentions into account. More recently, research has focused on the specific aspects of children's judgments that contribute to their tendency to focus on outcomes so extensively such as assumptions about negligence of an actor (Nobes, Panagiotaki, & Pawson, 2009) and failing to recognize that positive intentions may lead to negative outcomes such as in the case of an accidental transgressor (Cushman, Sheketoff, Wharton, & Carey, 2013; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011). However, much of this research has used only a single ToM assessment, which leaves open the question regarding how other levels of ToM competence are related to children's moral evaluations as well as whether children are able to apply their ToM competences in all contexts (see Rizzo & Killen, 2018a,b).

In fact, understanding how children process others' mental states *within a morally relevant context* provides an important insight into the bidirectional relationship between children's understanding of others' mental states and their moral evaluations (Killen et al., 2011; Leslie, Knobe, & Cohen, 2006; Smetana, Jambon, Conry-Murray, & Sturge-Apple, 2012). We propose that morally relevant theory of mind (MoToM) assessments will be fruitful for documenting the role of mental state knowledge for differentiating intentional and unintentional false statements. To compare different assessments of ToM competence, Killen et al. (2011) analyzed two types of ToM competence: prototypic and "morally embedded" (MoToM). The prototypic tasks were those identified by Wellman and Liu (2004) such as location change (where will a child look for an object when it has been displaced out of sight?) and false contents (what does a child think is in a box labeled *crayons* when it actually holds crackers?). Morally embedded false belief tasks involve assessing children's understanding of those same mental states in a context where the transfer of an object out of sight has moral meaning (e.g., one's special cupcake is displaced from a bag to the trash can) or the false contents is morally relevant (i.e. the paper bag looks like it has trash in it, but it actually holds someone's special cupcake). In this context, a false belief may lead to an accidental moral transgression (e.g., a classroom helper who unknowingly throws away another student's special cupcake). In the study by Killen et al. (2011) to examine these relationships, children passed the MoToM task at a later age than the they passed the prototypic false belief task, and understanding a false belief held by a transgressor was more difficult than understanding a false belief held by a victim. These findings suggest that children's ability to infer the mental states of others may be contextually dependent and that children's ability to demonstrate false belief ToM competence in one context does not guarantee that children will be able to demonstrate that competence in another morally relevant context.

Thus, it is important to understand (a) how children's underlying ToM *competence* (as assessed by prototypic ToM assessments) is related to children's developing understanding of morally relevant contexts and (b) whether children's *ability* to demonstrate that competence in morally relevant contexts provides a more precise account of how children perceive moral transgressions. That is, we propose that it is important to investigate children's ToM ability in social contexts where other factors (e.g., the presence of harm to a victim; Killen et al., 2011) are in play. No studies to date have provided a systematic comparison between MoToM and prototypic ToM and their ability to predict subsequent moral evaluations regarding intentional and unintentional false statements. Therefore, this study aimed to address this gap in the literature.

The current study

To investigate children's understanding of intentional and unintentional false claims in a resource allocation task, participants were told about two children, one who made an *unintentional* false statement and another who made an *intentional* false statement, regarding the legitimacy of their claim to

resources. Participants were asked to evaluate both characters' actions on several measures, which were designed to provide a comprehensive assessment of their perceptions about the intentional and unintentional false claims. These measures were (a) the child's *moral evaluation* of the behavior, (b) an *attribution of intentions* assessment for each claimant, (c) the *child's assigned punishment*, and (d) the *expected teacher's punishment* for the behavior. In addition, given the focus on how ToM ability may interact with children's understanding of these situations, we included both prototypic ToM and MoToM assessments.

Hypotheses

Based on previous research showing that, even at a young age, children consider the intentions of a character when generating moral judgments (e.g., Nobes et al., 2009), the first set of hypotheses (H1) tested the prediction that children would view the unintentional transgressor as more positive from a moral perspective than the intentional transgressor. We predicted that children would view the unintentional transgressor as less wrong than the intentional transgressor and that they would also attribute more positive intentions to the unintentional false claimant than to the intentional false claimant. In line with these predictions, we also hypothesized that, on average, participants would judge that the unintentional transgressor should receive less punishment for the actions than the intentional transgressor and that participants would predict that an authority figure would punish the unintentional transgressor less than the intentional transgressor.

Next, we predicted an interaction between children's performance on the MoToM measure and the intentions of the characters on their moral judgments (H2). Specifically, we hypothesized that participants who passed the MoToM assessment (demonstrating an ability to *apply* their ToM competence in the specific context) would perceive unintentional transgressors more positively than intentional transgressors, whereas participants who failed the MoToM assessment would not differ in their perceptions. That is, we expected children who passed the assessment to (a) report more positive evaluations, (b) attribute more positive intentions, (c) assign less punishment, and (d) expect less punishment from the teacher for unintentional transgressors compared with intentional transgressors. Relatedly, we also predicted that those who passed the MoToM measure would view the unintentional transgression more favorably than the intentional transgression, but no such prediction was made for those who failed the MoToM measure. These predictions were based on previous work showing that children's MoToM ability is a key predictor of children's evaluation of unintentional transgressions (Killen et al., 2011).

Finally, we also tested the hypothesis (H3) that children's performance on the MoToM assessment would predict moral evaluations above and beyond both age and prototypic ToM performance. Specifically, we expected that, consistent with previous research, prototypic ToM performance would predict differences in the moral assessment of unintentional transgressors as measured by the moral evaluation, attribution of intentions, child's assigned punishment, and expected teacher's punishment measures above and beyond age (Killen et al., 2011). Furthermore, we expected that MoToM performance would be able to predict these differences in assessments of the unintentional character beyond age and prototypic ToM performance. This hypothesis stems from our theoretical model that developmental mechanisms such as mental state knowledge are significant predictors of change, more than age alone.

Method

Participants

Participants were children aged 4–10 years ($N = 122$; $M_{\text{age}} = 6.93$ years, $SD = 1.77$). All participants were recruited from preschool and day camps serving lower-middle- to upper-middle-income families in the mid-Atlantic United States. Participants were evenly split by gender (62 girls) and were ethnically diverse (63% European American, 19% Asian American, 15% African American, and 3% Hispanic).

Procedure

A trained research assistant individually administered the task to all participants. Interviews were conducted in a quiet space in children's schools and day camps and lasted approximately 15–20 min. The research assistant read the children stories about an unintentional transgressor and an intentional transgressor from a script, which was accompanied by a PowerPoint presentation on a laptop. All participants heard stories about both an unintentional transgressor and an intentional transgressor, creating a within-participants design. To control for order effects, the presentation order of these stories was counterbalanced across participants.

Before participants were introduced to the stories, it was explained to all children that their participation was voluntary, and they were asked to provide their assent before the researcher continued. Following assent from participants, the children were introduced to a 6-point Likert-type scale and were trained on its use. Once children demonstrated sufficient comprehension of the scale and were able to reliably indicate the positively and negatively valenced ends of the scale, as well as the mid-points, the researcher introduced the first story.

Children were randomly assigned to hear the story regarding either the unintentional transgressor or the intentional transgressor first. Both stories began with an introduction to two characters (Sam/Alex in the unintentional story and Taylor/Jesse in the intentional story) who were planning on attending a field trip. The characters were portrayed as approximately each participant's own age and were matched to the gender of the participant. Participants were then told that the characters were instructed by their teacher to pack drinks for the field trip because everyone needed to bring their own. In both stories, it was also made known to participants that one character (Alex/Jesse) did not have access in the home to drinks that could be brought in, whereas the other character (Sam/Taylor) had drinks readily available to bring in. This format created a context in each story where the false claim of one character (Sam/Taylor) is evaluated alongside the legitimate claim of another character (Alex/Jesse). (See Appendix A for full text.)

Intentionality manipulation

In the unintentional transgressor story, children heard that the child who had access to the necessary drinks for the field trip (Sam) forgot to place the drinks into the backpack the night before the field trip and went to sleep. It was then shown and explained that Sam's brother came in and placed the needed drinks into Sam's backpack while Sam slept. In the morning, Sam was running late and needed to grab the backpack and run to school without having a chance to check inside of it. At this point, children were provided with a manipulation check in which they were asked, "Did Sam see inside [his/her] backpack before [he/she] left for school?" Children who correctly answered that Sam had not seen inside the backpack proceeded with the remainder of the story, whereas children who incorrectly answered that Sam had seen inside the backpack had the manipulation repeated to them until they were able to respond correctly. All participants reported were able to pass the memory check after two or fewer repetitions of the story.

In the intentional transgressor story, children heard that the character with access to the necessary drinks (Taylor) packed the drinks into the backpack and then went to bed. The following morning, Taylor grabbed the backpack and ran off to school. At this point, children were provided with a manipulation check in which they were asked, "Did Taylor pack drinks inside [his/her] backpack before [he/she] left for school?" Again, children who correctly answered that Taylor had packed the drinks in the backpack proceeded with the remainder of the story, whereas children who incorrectly stated that Taylor had not packed the drinks had the manipulation repeated to them until they answered correctly. All participants were able to correctly answer the memory check following two or fewer repetitions of the story.

False claim to communal resources

Following the manipulation, both stories once again followed the same trajectory. In both scenarios, participants heard that the students arrived at the school and the teacher announced that she brought two extra drinks and would give them to any students who did not have drinks of their own. As previously stated, each story contained one child who did not have drinks at home

(Alex/Jesse) and one character who did have drinks available at home (Sam/Taylor). In both stories, the character who did not have any drinks (Alex/Jesse) stated first that he or she did not have any drinks (legitimate claim to communal resources), and then the character with drinks in the backpack (Sam/Taylor) also stated that he or she did not have any drinks (false claim to communal resources). Importantly, the description of Sam and Taylor's behavior was described identically for both scenarios, with the only difference being their knowledge, or lack of knowledge, of the contents of the backpack.

Moral assessment of behavior

Following the false claims of Sam/Taylor, children were given four assessments regarding the behavior they witnessed. These assessments were *evaluation*, *attribution of intentions*, *child's assigned punishment*, and *expected teacher's punishment*.

For the evaluation assessment, participants were asked, "Do you think it was OK or not OK for [Sam/Taylor] to say [he/she] does not have drinks? How OK/not OK?" Participants gave their responses on a Likert-type scale ranging from 1 (*really not OK*) to 6 (*really OK*).

The attribution of intentions assessment involved the researcher asking children, "Okay, so you think that it's [children's previous rating], but what about [Sam/Taylor]? Did [Sam/Taylor] think [he/she] was doing something OK or not OK by saying that [he/she] does not have drinks? How OK/not OK?" Again, participants gave their response on a Likert-type scale ranging from 1 (*really not OK*) to 6 (*really OK*).

Next, participants were given the child's assigned punishment assessment, which consisted of the researcher asking, "Do you think [Sam/Taylor] should get in trouble for saying [he/she] had no drinks?" If participants responded that they thought the character should get in trouble, they were then given the follow-up question, "Do you think they should get in a little trouble or a lot of trouble?" This resulted in a 3-point scale ranging from 0 (*no trouble*) to 2 (*a lot of trouble*).

Before participants received the expected teacher's punishment assessment, they were told that the teacher had opened the backpacks of the children with the legitimate claim (Alex/Jesse) and with the false claim (Sam/Taylor) in order to give each a drink. In each story, one of the characters did not have drinks in the backpack, whereas the other character did. Participants were told that the teacher saw the drinks in the backpack of the character with the false claim (Sam/Taylor). At this point, the researcher administered the expected teacher's punishment assessment and asked participants, "Based on what the teacher knows *now*, do you think [Sam/Taylor] will get in trouble for saying [he/she] had no drinks?" If participants stated that they thought the character would get in trouble, they were then given the follow-up question: "Do you think they will get in a little trouble or a lot of trouble?" Children's responses were recorded on a 3-point scale ranging from 0 (*no trouble*) to 2 (*a lot of trouble*).

Embedded ToM measure

Following the child's assigned punishment assessment, but prior to the continuation of the narrative resulting in the teacher's knowledge of the contents of each character's backpack, participants were assessed on their perception of the characters' knowledge states. To measure participants' morally embedded ToM abilities (Killen et al., 2011), children's responses to first-order beliefs in the unintentional story (Sam) were used to create an *embedded ToM assessment*. Specifically, participants were asked, "What does Sam think that [he/she] has in [his/her] backpack?" and "What does the teacher think that Sam has in [his/her] backpack?" Participants who correctly indicated that both Sam and the teacher believed that there were not drinks in the backpack were indicated as having passed, whereas those who stated that either character knew there were drinks in the backpack did not pass.

The decision to include only beliefs relevant to the unintentional transgressor story was theoretically motivated by the desire to include only items that required children to recognize that their knowledge state differed from that of the characters in the story. When asked about the beliefs of a character who believes the same thing as participants, as was the case in the intentional story, children could correctly respond either by correctly inferring the beliefs of the character or by providing their own belief (Baron-Cohen, Leslie, & Frith, 1985). Therefore, it was determined that asking about the false beliefs present in the unintentional story would provide the most appropriate measure of

ToM ability. To maintain balance between the two stories, these questions were also asked for the intentional story, but analyses were not conducted on these data because we did not predict that these answers would provide useful predictive ability.

Prototypic ToM measure

After participants had heard stories about both the intentional and unintentional transgressors, all children then heard a modified version of the contents false belief task (Gopnik & Aslington, 1988). In this task, children were introduced to two children, Sally and Anne, who are in art class together. Before recess, Sally put her crayons in the crayon box and then went to play outside. Once she was outside, Anne came over and removed Sally's crayons from the box before placing her own crackers in the box instead. To test for comprehension, participants were then asked what is really in the box now and whether Sally could see Anne putting the crackers in the box. If participants failed to correctly answer these memory check questions, they heard the story again, up to two times, and all participants were able to answer both questions correctly within this limit. Following successful completion of the memory check questions, children were then asked the crucial question, "What does Sally think is in the box?" This question was selected as an appropriate comparison with the MoToM measure because both items required participants to report on their understanding of first-order false beliefs regarding the contents of a container.

Data analytic plan

Given that the current study employed a repeated-measures design, the order in which participants heard the stories was included as a covariate in all of our analyses. To investigate the relationship between the intentions of the claimant and children's moral judgments, a 2×2 repeated-measures multivariate analysis of covariance (MANCOVA), with age and order as covariates, was used. Age was included as a covariate in this model because participants were found to pass the morally embedded ToM measure more often as they got older. Therefore, age was entered as a covariate in these analyses to determine the importance of performance on the morally embedded ToM measures above and beyond age. Planned post hoc analyses to investigate differences between those who passed the MoToM measure and those who failed were conducted with independent-samples *t* tests, whereas planned post hoc analyses to determine differences between the unintentional and intentional transgressions for those who passed and failed the MoToM measure were conducted using paired-samples *t* tests. Finally, to test for the ability of MoToM performance to predict differences in children's moral judgments, a series of hierarchical linear regressions were conducted controlling for order in Step 1, prototypic ToM performance entered as Step 2, and MoToM performance entered as Step 3.

Results

Moral judgment of unintentional and intentional false claims

The hypothesis that participants would evaluate unintentional false claims more positively than intentional false claims (H1) was assessed through a repeated-measures MANCOVA controlling for age and order. As expected, this analysis revealed a significant effect of intentions for evaluation, $F(1, 120) = 78.02, p < .001$, attribution of intentions, $F(1, 120) = 36.33, p < .001$, child's assigned punishment, $F(1, 120) = 64.115, p < .001$, and expected teacher's punishment, $F(1, 120) = 38.95, p < .001$ (see Table 1). Participants viewed the unintentional false claim more positively than the intentional false claim, provided a more positive attribution of intentions for the unintentional false claim than for the intentional false claim (Fig. 1), assigned less punishment for the unintentional transgressor than for the intentional transgressor, and expected the teacher to assign less punishment for the unintentional transgressor than for the intentional transgressor (Fig. 2). This finding supported our hypothesis (H1) that children would be more favorable in their rating of the unintentional false claimant relative to the intentional false claimant.

Table 1
Participants' responses to unintentional and intentional false claims.

Measure	Intentionality of false claim			
	Unintentional		Intentional	
	M	SE	M	SE
Evaluation	3.26	0.143	1.76	0.110
Attribution of intentions	4.38	0.150	3.13	0.161
Child's assigned punishment	0.540	0.068	1.24	0.066
Expected teacher's punishment	0.850	0.067	1.31	0.064

Note. Evaluations and attribution of intentions are on a 6-point Likert-type scale ranging from 1 = *really not OK* to 6 = *really OK*. Child's assigned punishment and expected teacher's punishment are on a 3-point scale ranging from 0 = *no trouble* to 3 = *lots of trouble*. Differences between columns were significant for each row ($ps < .001$).

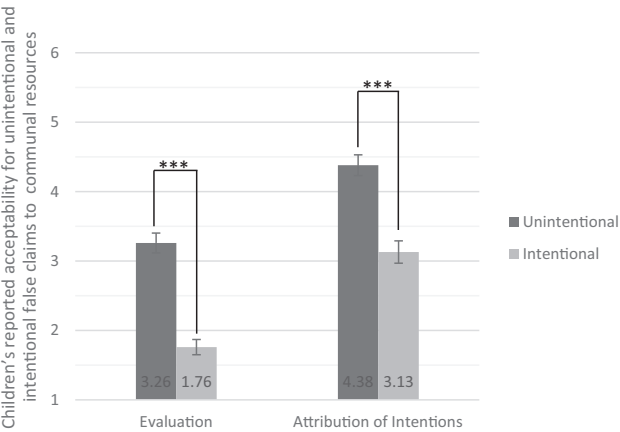


Fig. 1. Children's rating of evaluations and attribution of intentions for unintentional and intentional false claims. Scores range from 1 (*really not OK*) to 6 (*really OK*). ***Significant at $p \leq .001$.

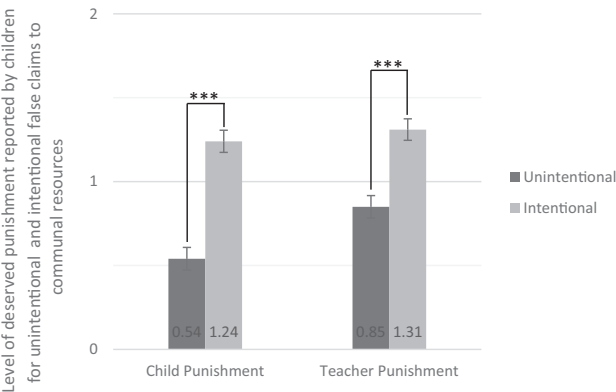


Fig. 2. Child's assigned punishment and expected teacher's punishment for unintentional and intentional false claims. Scores range from 0 (*no punishment*) to 2 (*lots of trouble*). ***Significant at $p \leq .001$.

Differences in moral assessment by MoToM performance

To test the relationship between MoToM performance and intentionality of the claimants on children's moral judgments (H2), a 2 (MoToM: pass or fail) \times 2 (unintentional or intentional false claim) MANCOVA was used with repeated measures on each of the four assessments (evaluation, attribution of intentions, child's assigned punishment, and expected teacher's punishment). The relationship between participant's age and performance on the MoToM measure was investigated through a linear regression. As expected, older children were more likely to pass the morally embedded ToM measure when age was entered as a continuous variable ($B = 0.12$, $p < .001$, $R^2 = .18$), and therefore age was used as a covariate for this analysis. Results revealed significant interactions between participants' performance on the MoToM assessment and the intentions of the false claimant for the child's evaluation, $F(1, 118) = 4.28$, $p = .041$, $\eta_p^2 = .04$, attribution of intentions, $F(1, 118) = 11.73$, $p = .001$, $\eta_p^2 = .09$, child's assigned punishment, $F(1, 118) = 7.54$, $p = .007$, $\eta_p^2 = .06$, and expected teacher's punishment, $F(1, 118) = 13.29$, $p < .001$, $\eta_p^2 = .10$ (see Table 2).

These interactions were investigated through planned follow-up t tests. To control for Type I error, Bonferroni-adjusted alpha levels of .006 were used for each test (.05/8). Paired-samples t tests revealed that participants who passed the MoToM measure gave significantly higher ratings to the unintentional false claim than to the intentional false claim for the evaluation ($p < .001$), attribution of intentions ($p < .001$), child's assigned punishment ($p < .001$), and expected teacher's punishment ($p < .001$) measures. No differences were found for any of the four assessments for participants who failed the MoToM assessment: evaluation ($p = .067$), attribution of intentions ($p = .499$), child's assigned punishment ($p = .050$), and expected teacher's punishment ($p = .860$) (Figs. 3–6). Likewise, independent-samples t tests showed that participants who passed the MoToM measure viewed the unintentional transgression significantly more positively than those who failed (evaluation [$p = .001$], attribution of intentions [$p < .001$], child's assigned punishment [$p = .004$], and expected teacher's punishment [$p < .001$]) but did not differ for the intentional transgressions (evaluation [$p = .258$], attribution of intentions [$p = .383$], child's assigned punishment [$p = .149$], and expected teacher's punishment [$p = .553$]). Taken together, these findings indicate that the children who passed the MoToM measure evaluated the unintentional transgression more positively than the intentional transgression, whereas those who failed the MoToM measure did not differentiate between the two transgressions on any of the four measurements. In addition, participants who passed the MoToM measure gave more positive evaluations of the unintentional false claim than those who failed the measure, whereas there were no significant differences between these two groups for the intentional false claim.

Predictive power of MoToM and prototypic ToM

To assess the ability of MoToM performance to predict moral judgments of the unintentional character, above and beyond prototypic ToM performance, a series of hierarchical linear regressions were

Table 2

Moral judgments of participants by ToM performance and intentionality of claim.

Measure	Morally embedded ToM performance							
	Passed MoToM measure				Failed MoToM measure			
	Unintentional		Intentional		Unintentional		Intentional	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Evaluation	3.51	0.18	1.75	0.14	2.76	0.25	1.79	0.20
Attribution of intentions	4.82	0.18	3.00	0.21	3.57	0.26	3.35	0.29
Child's assigned punishment	0.43	0.08	1.32	0.09	0.74	0.12	1.10	0.12
Expected teacher's punishment	0.67	0.08	1.33	0.08	1.20	0.11	1.28	0.12

Note. Evaluation and attribution of intentions are on a 6-point Likert-type scale ranging from 1 = *really not OK* to 6 = *really OK*. Child's assigned punishment and expected teacher's punishment are on a 3-point scale ranging from 0 = *no trouble* to 2 = *lots of trouble*. All differences between columns for individuals who passed the ToM measure are significant for each row ($ps < .001$). All differences between columns for individuals who failed the ToM measure are not significant ($ps \geq .05$).

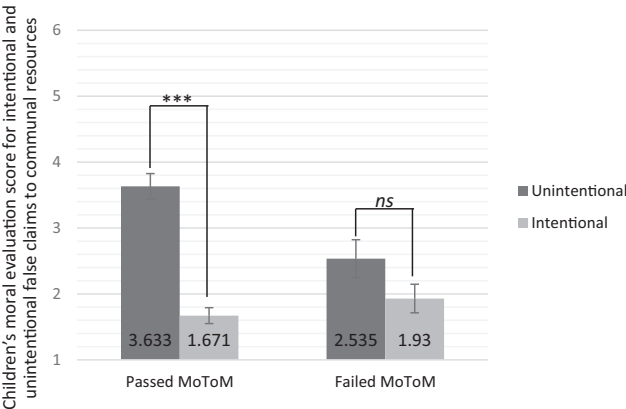


Fig. 3. Children's evaluation rating split by MoToM performance. Scores range from 1 (*really not OK*) to 6 (*really OK*).
***Significant at $p \leq .001$; ns, not significant.

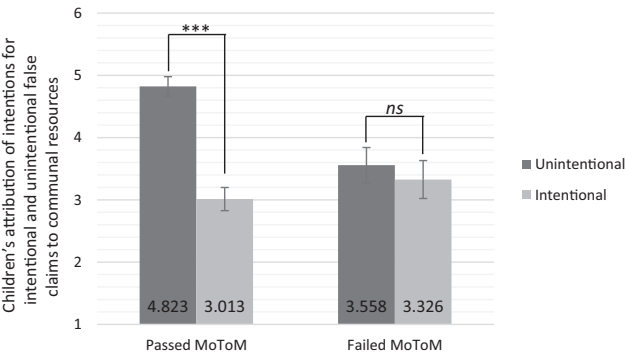


Fig. 4. Children's attribution of intentions split by MoToM performance. Scores range from 1 (*really not OK*) to 6 (*really OK*).
***Significant at $p \leq .001$; ns, not significant.

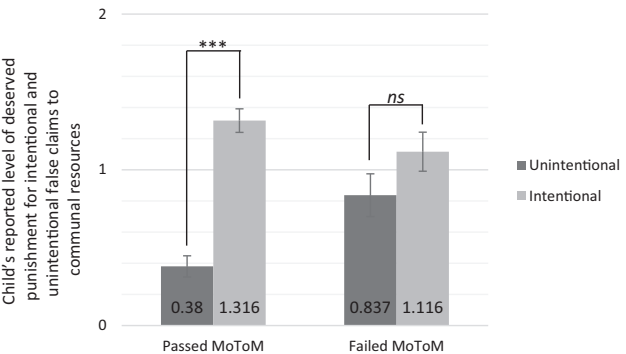


Fig. 5. Child's punishment rating split by MoToM performance. Scores range from 0 (*no trouble*) to 2 (*lots of trouble*).
***Significant at $p \leq .001$; ns, not significant.

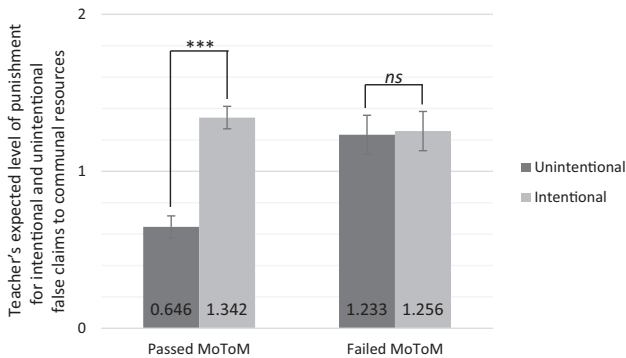


Fig. 6. Expected teacher's punishment rating split by MoToM performance. Scores range from 0 (*no trouble*) to 2 (*lots of trouble*). ***Significant at $p \leq .001$; ns, not significant.

performed. For each of these analyses, order was entered as a covariate in Step 1, prototypic ToM performance was entered in Step 2, and MoToM performance was entered in Step 3. This process was used in order to determine the ability of prototypic ToM performance to predict differences in each of the moral assessment measures while controlling for possible order effects and to determine the ability of MoToM performance to predict performance above and beyond prototypic ToM performance on each of the moral assessment measures. This pattern was repeated for each of the measures, and the results of the hierarchical linear regressions are discussed in sequence with all of the results reported in Table 3.

When predicting responses on the evaluation measure, the results of Step 1 showed that order did account for a significant amount of the variance on its own, $R^2 = .279$, $F(1, 120) = 46.43$, $p < .001$, that the addition of prototypic ToM performance in Step 2 resulted in a significant increase in the amount of variance explained, $\Delta R^2 = .03$, $\Delta F(1, 119) = 5.52$, $p = .02$, and the addition of MoToM performance in Step 3 resulted in a significant increase in the amount of explained variance in moral evaluation,

Table 3
Hierarchical regression analyses evaluating ToM predictors of moral judgments for the unintentional transgressor.

Measure	<i>R</i>	<i>R</i> ²	ΔR^2	ΔF	<i>df</i>	<i>B</i>
<i>Evaluation</i>						
Order	.528	.279	.279	46.43***	1, 120	1.83***
Prototypic ToM	.558	.311	.032	5.52*	1, 119	0.84
MoToM	.587	.344	.034	6.03**	1, 118	0.75**
<i>Attribution of intentions</i>						
Order	.161	.026	.026	3.20	1, 120	0.37
Prototypic ToM	.353	.125	.099	13.42***	1, 119	1.46*
MoToM	.440	.194	.069	10.102**	1, 118	0.969**
<i>Child's punishment</i>						
Order	.098	.010	.010	1.16	1, 120	−0.08
Prototypic ToM	.307	.094	.085	11.15**	1, 119	−0.647*
MoToM	.368	.136	.041	5.62*	1, 118	−0.336*
<i>Teacher's punishment</i>						
Order	.180	.032	.032	4.01*	1, 120	−0.19
Prototypic ToM	.345	.119	.087	11.71**	1, 119	−0.58*
MoToM	.446	.199	.080	11.78**	1, 118	−0.47**

Note. For each analysis, order was entered as a binary variable in Step 1, prototypic ToM was entered in Step 2, and MoToM was entered in Step 3. All *R* and *F* values are representative of the model at the stage in which these variables were entered, and all *B* values are taken from the final model in which all three variables are included at once. All *B* values are standardized. Values with * are significant at $p < .05$, values with ** are significant at $p < .01$, and values with *** are significant at $p \leq .001$.

$\Delta R^2 = .03$, $\Delta F(1, 118) = 6.03$, $p = .015$. Thus, adding MoToM to the model in the final step significantly increased the predictive power of the overall model, suggesting that MoToM had significant predictive power even above and beyond prototypic ToM performance. The final regression equation indicated that order ($p < .001$) and MoToM performance ($p = .015$) were statistically significant predictors of moral evaluation but that prototypic ToM performance was not significant in the final model ($p = .132$).

For analyses regarding the attribution of intentions measure, the results of Step 1 show that order did not account for a significant amount of the variance, $R^2 = .161$, $F(1, 120) = 3.20$, $p = .076$, that the addition of prototypic ToM performance in Step 2 resulted in a significant increase in the amount of variance in attribution of intentions explained, $\Delta R^2 = .10$, $\Delta F(1, 119) = 13.42$, $p < .001$, and that the addition of MoToM performance in Step 3 resulted in a significant increase in the amount of explained variance in attribution of intentions, $\Delta R^2 = .07$, $\Delta F(1, 118) = 10.10$, $p = .002$. Thus, the results of this analysis are in line with our hypothesis and show that MoToM performance has predictive power even above that of prototypic ToM performance. The final regression equation indicated that the significant predictors of attribution of intentions were MoToM performance ($p = .002$) and prototypic ToM performance ($p = .010$) but that order was not a significant predictor ($p = .187$).

The next hierarchical linear regression was set to examine the ability of MoToM to predict child's punishment decision beyond the predictive power of prototypic ToM performance. Step 1 of the analysis showed that order was not able to explain a significant amount of the variance by itself, $R^2 = .010$, $F(1, 120) = 1.16$, $p = .284$, that adding prototypic ToM performance into the model resulted in a significant increase in the amount of variance in child's punishment decision explained, $\Delta R^2 = .085$, $\Delta F(1, 119) = 11.15$, $p = .001$, and that the addition of MoToM performance in Step 3 resulted in a significant increase in the amount of explained variance in child's punishment decision, $\Delta R^2 = .041$, $\Delta F(1, 118) = 5.62$, $p = .019$. The final regression equation indicated that the only significant predictors of child's punishment decision for this model were MoToM performance ($p = .019$) and prototypic ToM performance ($p = .014$), whereas order was not significant ($p = .530$).

The final set of analyses was focused on the expected teacher's punishment measure, and these analyses found that entering order into the model at Step 1 did account for a significant amount of the variance, $R^2 = .032$, $F(1, 120) = 4.01$, $p = .048$, that entering prototypic ToM performance in Step 2 resulted in a significant increase in the amount of variance in expected teacher's punishment that was explained, $\Delta R^2 = .087$, $\Delta F(1, 119) = 11.71$, $p = .001$, and that entering MoToM performance in Step 3 resulted in a significant increase in the amount of explained variance in expected teacher's punishment, $\Delta R^2 = .080$, $\Delta F(1, 118) = 11.77$, $p = .001$. The final regression equation indicated that the significant predictors of expected teacher's punishment were MoToM performance ($p = .001$) and prototypic ToM performance ($p = .021$) but that order was not a significant predictor ($p = .122$). Taken together, these findings provide support for our hypothesis (H3) regarding the ability of MoToM to provide a strong contribution to the predictive power of our models even above that provided by prototypic ToM.

Discussion

By providing a direct comparison between unintentional and intentional false statements and investigating the ways in which ToM performance predicts moral assessment, the current study provides novel findings that further our understanding of children's comprehension of intentionality regarding moral judgments. First, young children distinguished between unintentional and intentional false statements that have moral implications, and they recognized that intentional false statements are more deserving of moral condemnation than those that are unintentional. These findings extend previous research, which has shown that children view unintentional harm as more morally permissible than intentional harm (Nobes et al., 2009) by providing multiple measures to demonstrate how this distinction manifests and by demonstrating this finding in a resource allocation context across four distinct moral assessment measures.

This study demonstrated that the distinction between unintentional and intentional false claims is robust for those children who understand it and that these moral judgments can be assessed through a

number of different measures. The evaluation measure employed in this study asked children to state how OK or not OK they found an action, and (as expected) children successfully indicated through this measure that they evaluated the unintentional transgression more positively than the intentional transgression. However, children also showed this distinction in their attributions of intentions, a more nuanced measure designed to assess children's belief that an individual wanted to do good or to do harm. In this way, it allowed for participants to state their views on the motivation behind the claim rather than the content of the claim itself or its outcome. Although this assessment was not an assessment of moral praise/blame directly, it did show the same consistent pattern as the other more direct moral assessments.

In addition to the attribution of intentions measure, the child's punishment decision and expected teacher's punishment measures were used to investigate a different aspect of moral assessments, in this case the severity of a reprimand required for the action. Here again, children showed a strong and consistent pattern where they thought unintentional false claimants deserved less punishment and expected that authority figures would share that viewpoint. Although judgments about punishment are certainly related to an evaluation or attribution of intentions, an action can be seen as wrong or poorly motivated but still not necessarily deserving of punishment, and therefore these items measure a different aspect of children's moral assessment. Importantly, for the first two measures, the unintentional false statement received *higher* ratings, whereas ratings on the second two measures were *lower* for the unintentional false statement. This pattern rules out the possibility of valence matching and provides strong evidence that children were reasoning about each of these questions independently. Furthermore, the finding that all four assessments followed a similar pattern provides a cohesive narrative regarding children's understanding of intentional and unintentional false statements.

Importantly, not all children demonstrated the ability to differentiate between the two types of false statements, and this study further showed that only those who were able to understand the difference in beliefs between the false claimants successfully distinguished between the two types of claims. That is, participants who successfully passed the MoToM measure rated the unintentional false statement more positively than the intentional false statement across each of the four measures. Interestingly, those who failed the MoToM measure did not show significant differences in their moral evaluation, attribution of intentions, child's assigned punishment, or expected teacher's punishment measures for the unintentional and intentional false statements. This suggests that those participants who failed to fully appreciate the belief state of the unintentional transgressor did not view the act as different from that of the intentional transgressor.

Critically, this study also served to replicate and extend previous research showing the ability of MoToM performance to predict moral assessments (Fu, Xiao, Killen, & Lee, 2014; Killen et al., 2011; Li, Rizzo, Burkholder, & Killen, 2017). In line with these studies, participants who passed the MoToM measure viewed the unintentional false statement more positively than those who failed the MoToM measure. Taken together with the other findings, the data provide convergent evidence that children's understanding of beliefs and intentions has a significant impact on their subsequent moral judgments for evaluating unintentional transgressions.

The current study also extended these findings by investigating the ways in which MoToM affected moral judgments of an *intentional* false statement. Importantly, these stories were identical in every capacity with the exception of the intentions and belief state of the claimant. Although there was a strong pattern of differences between those who passed and those who failed the MoToM measure when responding to the unintentional false claim, no significant differences were found for the intentional false claim. If MoToM was tapping into cognitive faculties other than a working understanding of beliefs (such as executive functioning and working memory), then these same differences between those who passed and those who failed the measure should be found for the intentional transgression. However, this was not the case, and the lack of significant differences for the moral judgment of the intentional character between those who passed and those who failed the MoToM measure further provides evidence that MoToM is tapping directly into children's ability to comprehend the mental states of individuals in a moral context. Furthermore, it is this ability that is driving the differences in children's moral assessments. By showing that the differences between these groups is restricted solely to scenarios in which there is unintentional harm, we demonstrate the unique and significant

importance of children's ability to fully appreciate the mental states of the individual in question when making moral judgments.

Prototypic ToM and MoToM

This study also provided a significant addition to the literature on children's ToM ability and its connection to their evaluations by rigorously testing the predictive power of the standard false belief ToM assessment and the morally embedded ToM assessment (MoToM). For three of the four moral assessments, MoToM performance was able to explain a significant amount of the variance in the measures even above age and prototypic ToM performance. Although MoToM was not able to explain a significant amount of the variance above and beyond age and prototypic ToM performance in the child's assignment of punishment measure, it was still found to be able to explain a significant and unique amount of the variance beyond that explained by prototypic ToM alone. These findings provide a case for the predictive power of MoToM measures and for the importance of considering this type of measure for assessments of mental state knowledge in socially complex contexts.

Limitations of the study

The fact that the MoToM measure is embedded directly into the relevant story, whereas the prototypic ToM measure is conducted at the end of the interview, raises some interesting issues and potential confounds, and it is prudent to consider them carefully. First, because the MoToM measure posed a question that was relevant to the stories, it is possible that this is partially acting as a comprehension check, and therefore it could be that it serves not only as a test of mental state understanding but also as a form of story comprehension more broadly. It is worth noting, however, that all participants passed the manipulation checks before they could proceed, and these manipulation checks were specifically targeted at ensuring that participants understood whether the characters witnessed the drinks being placed in their backpacks. This manipulation check ensured that all participants had comparable understanding of the story, specifically that they understood and remembered the relevant information to draw mental state inferences. Therefore, the MoToM measure did not investigate any aspects of story comprehension beyond what the manipulation checks measured, but rather they measured children's ability to apply that relevant information to draw an accurate inference about the mental states of the characters in question.

Another implication of the MoToM measure being embedded within the stories is that the MoToM measure may be more directly related to the moral evaluations simply by being assessed in the same context of the moral evaluation questions. The prototypic ToM question was posed at the end of the interview and measured mental state understanding in a different scenario than the moral evaluation questions, leading to a potential interpretation that the distinctly different context contributed to the lack of predictive power for ToM in contrast to MoToM. We assert that this is likely the case and that this further underscores our main argument, namely that successfully passing mental state assessments in one context does not necessarily guarantee success in another context and that measuring this cognitive competence in the relevant context will provide a stronger measure of a participant's cognitive processes as applied to that situation.

Finally, due to the way in which theory of mind was assessed in the morally embedded and nonembedded measures, there was an asymmetry in the number of items that children needed to correctly answer in order to pass. The MoToM measure required children to accurately report two first-order false beliefs, whereas the prototypic ToM measure required children to accurately report one first-order belief. This was done due to the fact that the ToM is a more widely used assessment. However, future research investigating morally embedded and nonembedded ToM measures should balance the number of items needed to pass each assessment.

Conclusion

Children are frequently in situations where they need to determine the culpability of others in order to determine how to react to perceived slights. When determining whether a false statement

is a mistake or a lie, the evaluator needs to know the intentions of the speaker to make a decision about the act (Evans & Lee, 2013). When children were presented with one character who had an intention to deceive and another character who did not, those participants who were able to differentiate these two types of intentions consistently judged the unintentional false statement more positively, whereas participants who were unable to make this distinction treated the two types of false statements as intentional transgressions. Even children who were old enough to pass a standard false belief ToM assessment made this mistake, which suggests that it is possible that children make these errors in real-life situations even later in development than previous studies would suggest.

Critically, this study also showed that a child's ability to understand that others may have beliefs that differ from their own in a neutral context was not the same as a child's performance within a socially embedded context. When comparing performance on a standard false belief task with performance on a similar task embedded within a morally relevant context, it was found that the morally embedded measure was a stronger predictor of children's evaluations of the scenarios. This suggests that a child's ability to identify the mental states of others in a socially simplified context is one way, but not the only way, to measure children's ability to identify mental states in more socially complex situations. In the standard ToM task, children are asked to provide their understanding of one character's beliefs and intentions, and then this response is used to predict how those children will respond in another situation involving another character and another set of beliefs. In contrast, the MoToM task assesses the beliefs and intentions of the same character for which children are also providing reasoning, and in this way the measure is more directly tapping into children's understanding of the relevant mental states. This logic also suggests that the morally embedded measure of ToM may be a beneficial methodology for research in which children's ability to use ToM in a complex context is a central concern. It is important to investigate morally relevant forms of mental state knowledge in complex social contexts in order to get a better understanding of children's application of their ToM abilities (Rizzo & Killen, 2018a,b). These contexts can easily be extended to include additional factors that may affect children's ability to apply their understanding of mental states to moral evaluations such as group identity and competitive/cooperative situations. Future research should continue to examine this issue by examining the role of morally embedded and nonembedded ToM assessments in a range of contexts, which would also increase the generalizability of the current results. In addition, this study was conducted using two very similar scenarios designed to assess children's understanding of false claims within the context of false claims to communal resources. The MoToM paradigm would benefit from additional research designed to establish its reliability across a greater number of scenarios.

Although the moral context is one such area in which ToM measures can be embedded, it is certainly not unique in that regard, and such an approach may prove to be useful in other domains as well. This approach has already proven to be successful within the realm of intergroup dynamics and the assessment of children's "theory of social mind" (Abrams, Rutland, Pelletier, & Ferrell, 2009), and further research should be done to explore embedded ToM assessments in other relevant domains and the ways in which these domains intersect.

In addition, we found that the order in which participants heard these morally embedded stories affected a few of their responses, leading participants to be more forgiving of the accidental transgressor when they had heard about the intentional transgressor first rather than the unintentional transgressor. Although this was not the focus of the current study, it suggests that children's definition of a serious transgression may be relative to that which they have seen previously. This seems reasonable given that a transgression such as stealing an eraser may be viewed as "really not OK," but when that situation is compared with pushing someone off the swings it may seem less severe. This order effect did not negatively affect the significance of findings in this study, but future research should be conducted to explore this possibility.

Another avenue for future research would be to administer a range of cognitive assessments to determine whether participants who passed the MoToM measures were more advanced in other cognitive abilities such as their executive function, verbal comprehension, and working memory (Doeniyas, Yavuz, & Selcuk, 2018; Lecce, Bianco, Devine, & Hughes, 2017). For example, one study found that children's second-order false belief understanding was related to MoToM competence after partialing out the effect of age (Fu et al., 2014). Investigating the relationships between a range of

cognitive abilities and performance on morally embedded ToM tasks has not been systematically explored and therefore would be a fruitful area for future research.

Children's understanding of the beliefs and intentions of others is crucial to their moral evaluations in many situations. However, this ability to accurately infer intentions, especially within a complex social scenario, is still developing throughout early childhood, and thus it is easy for children to mistakenly arrive at incorrect conclusions. The findings of the current study shed light on children's developing abilities to understand what makes false claims permissible or impermissible and the role that mental state knowledge plays in this important benchmark in child development.

Acknowledgments

We thank the members of the Social and Moral Development Lab at the University of Maryland for their invaluable contributions to this project and thank Adam Hartstone-Rose for the illustrations. We gratefully recognize the research assistants who helped with data collection for this project: Amanda Burkholder, Ashley Foster, Victoria Gordon, and Clare Sarsony. The first author was supported by a University of Maryland graduate fellowship, the second author was supported by a National Institute of Child Health and Human Development grant (T32HD007542), and the third author was supported by a National Science Foundation grant (1728918). We are very appreciative of the schools, students, and parents who participated in this study.

Appendix A

Unintentional scenario

This is Sam, and this is Alex. Their class is going on a field trip, and all of the students were told to bring water for the field trip. Where they are going, it is very hot and there won't be anything to drink! If Sam and Alex don't bring any water, they will start to feel sick. But if they have water, they will stay healthy and have fun on the trip! Look, Sam has some water bottles at home that she can take with her on the field trip. And look, Alex does not have any water bottles at home that she could bring with her on the field trip. The night before the field trip, Sam went to bed without putting water in her backpack. While Sam was asleep, her big brother put two water bottles in Sam's backpack and zipped it up.

In the morning, Sam sees that her backpack is zipped up. Sam grabs her backpack and runs off to school. At school, before the class leaves for the field trip, the teacher says, "Does anyone not have water? I have two extra bottles for anyone who does not have any water." Alex says to the teacher, "I do not have any water bottles." Sam also says to the teacher, "I do not have any water bottles."

Intentional scenario

This is Taylor, and this is Jesse. Their class is going on a field trip, and all of the students were told to bring water for the field trip. Where they are going, it is very hot and there won't be anything to drink! If Taylor and Jesse don't bring any water, they will start to feel sick. But if they have water, they will stay healthy and have fun on the trip! Look, Taylor has some water bottles at home that she can take with her on the field trip. And look, Jesse does not have any water bottles at home that she could bring with her on the field trip. The night before the field trip, Taylor put the water in her backpack.

In the morning, Taylor grabs her backpack and runs off to school. At school, before the class leaves for the field trip, the teacher says, "Does anyone not have water? I have two extra bottles for anyone who does not have any water." Jesse says to the teacher, "I do not have any water bottles." Taylor also says to the teacher, "I do not have any water bottles."

References

- Abrams, D., Rutland, A., Pelletier, J., & Ferrell, J. M. (2009). Children's group nous: Understanding and applying peer exclusion within and between groups. *Child Development*, 80, 224–243.
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–46.

- Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development*, 63, 129–137.
- Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development*, 70, 1338–1347.
- Cheung, H., Siu, T. S., & Chen, L. (2015). The roles of liar intention, lie content, and theory of mind in children's evaluation of lies. *Journal of Experimental Child Psychology*, 132, 1–13.
- Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, 127, 6–21.
- Doenyaş, C., Yavuz, H. M., & Selcuk, B. (2018). Not just a sum of its parts: How tasks of the theory of mind scale relate to executive function across time. *Journal of Experimental Child Psychology*, 166, 485–501.
- Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology*, 49, 1958–1963.
- Fu, G., Sai, L., Yuan, F., & Lee, K. (2017). Young children's self-benefitting lies and their relation to executive functioning and theory of mind. *Infant and Child Development*. Early online publication. doi:10.1002/icd.2051.
- Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology*, 50, 2083–2092.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Development*, 59, 26–37.
- Hughes, C., & Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives*, 9, 149–153.
- Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, 119, 197–215.
- Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Development*, 163, 69–86.
- Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect. *Psychological Science*, 17, 421–427.
- Li, L., Rizzo, M. T., Burkholder, A., & Killen, M. (2017). Theory of mind and resource allocation in the context of hidden inequality. *Cognitive Development*, 43, 25–36.
- 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.
- Peterson, C. C. (1995). The role of perceived intention to deceive in children's and adults' concepts of lying. *British Journal of Developmental Psychology*, 13, 237–260.
- Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.
- Rizzo, M. T., & Killen, M. (2018a). Theory of mind is related to children's resource allocations in gender stereotypic contexts. *Developmental Psychology*, 54, 510–520. <https://doi.org/10.1037/dev0000439>.
- Rizzo, M. T., & Killen, M. (2018b). How social status influences our understanding of others' mental states. *Journal of Experimental Child Psychology*, 169, 30–41.
- Siegal, M., & Peterson, C. C. (1996). Breaking the mold: A fresh look at children's understanding of questions about lies and mistakes. *Developmental Psychology*, 32, 322–334.
- Siegal, M., & Peterson, C. C. (1998). Preschoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology*, 34, 332–341.
- Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. (2012). Reciprocal associations between young children's developing moral judgments and theory of mind. *Developmental Psychology*, 48, 1144–1155.
- Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development*, 87, 1221–1232.
- Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, 79, 866–881.
- Talwar, V., Williams, S. M., Renaud, S., Arruda, C., & Saykaly, C. (2016). Children's evaluations of tattles, confessions, prosocial, and antisocial lies. *International Review of Pragmatics*, 8, 334–352.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75, 523–541.