Children can use others’ emotional expressions to infer their knowledge and predict their behaviors in classic false belief tasks

Yang Wu (yangwu@mit.edu)  
Jennah A. Haque (jhaque@mit.edu)  
Laura E. Schulz (lschulz@mit.edu)

Department of Brain and Cognitive Sciences, MIT  
77 Massachusetts Avenue, Cambridge, MA 02139 USA

Abstract
In this study, we investigate whether emotional expressions provide cues to knowledge sufficient for predicting others’ behavior based on their true and false beliefs. We adapted the classic Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985) such that children (N = 62, mean: 5.58 years, range: 4.05-6.98 years) were not told whether Sally saw Anne move the object or not. However, when Sally came back looking angry, even four-year-olds inferred that she had seen Anne move her toy; when she came back looking happy, children inferred that she had not seen the transfer. Based on these inferences, five and six-year-olds, although not four-year-olds, were able to predict where Sally would look for her toy.

Keywords: emotion understanding; emotional expressions; theory of mind; false beliefs; knowledge state

Introduction
Researchers have proposed that children construct an intuitive theory of others’ mental states in which representations of beliefs, desires, and emotions are causally linked (see Harris, 2008 and Wellman, 2014 for discussion and review). Considerable research has looked at children’s ability to infer and predict others’ emotional responses from knowledge of their beliefs, desires, and the broader context (e.g., Bradmetz & Schneider, 1999; Hadwin & Perner, 1991; Harris et al., 1989; MacLaren & Olson, 1993; Ruffman & Keenan, 1996; Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). However, there is a developmental lag between children’s understanding of beliefs and their ability to infer the emotional consequences of those beliefs (e.g., de Rosnay, Pons, Harris, & Morell, 2004; Hadwin & Perner, 1991; Harris et al., 1989; Pons, Harris, & de Rosnay, 2004; Ruffman & Keenan, 1996; Wellman & Bartsch, 1988). Thus for instance, four- and five-year-olds may know that Red Riding Hood falsely believes her grandmother is in bed, and nonetheless conclude that Red Riding Hood is frightened (Bradmetz & Schneider, 1999).

However, in the real world, we are more likely to have access to someone’s emotional reactions to events than to the beliefs and desires that contributed to that reaction. Nonetheless, relatively few studies have looked at children’s ability to reason backwards from others’ emotional reactions to their desires and beliefs, and most of this work has focused on inferences about others’ desires. Thus for instance, eighteen-month-olds, but not fourteen-month-olds, can use an agent’s emotional expressions (along with verbal cues like “Yummy!” and “Yucky!”) to decide if she wants a food different from what the child herself wants (Repacholi & Gopnik, 1997). By two and three, children can use someone’s emotional reaction to infer whether she is looking at something she does or does not want (Wellman, Philips & Rodriguez, 2000). By seven, children can use one person’s emotional expression in a social context to infer the desires of another based on their knowledge of social display rules (Wu & Schulz, 2017).

By contrast, before age six, children rarely refer to agents’ beliefs in explaining their emotional reactions (Rieffe, Terwogt & Cowan, 2005). One exception is that four and five-year-olds refer to others’ beliefs in explaining fearful or atypical reactions (e.g., saying “She thought it was a ghost” if a character is scared at hearing a noise, and “She thought it would be something else” if a character is sad upon opening a present; Rieffe, et al., 2005) and in explaining
others’ emotional reactions to surprising or unusual events (“she didn’t think there would be a giraffe” or “she didn’t know what was in the box”; Wellman & Banerjee, 1991). However, by preschool, children have learned many scripts connecting familiar events and emotions (e.g., Fabes, Eisenberg, McCormick, & Wilson, 1988; Gnepp, McKee, & Domanic, 1987; Harris, Olthof, Terwogt, & Hardman, 1987; Widen & Russell, 2010). Children might link fear with ghosts, or a disappointing gift with sadness (Rieffe et al., 2005) without necessarily representing the relationship between emotions and beliefs more broadly. Similarly, children might guess that an agent did not know about or expect unusual events because the events are atypical or mysterious rather than because they reason about the beliefs underlying emotional reactions generally.

Perhaps the strongest support for the idea that children recover beliefs from emotional responses comes from a recent study showing that five-year-olds, but not four-year-olds, infer false beliefs from a change of valence on someone’s facial expression between an anticipated and observed outcome (Wu & Schulz, 2018). In that study, children saw, for instance, that a character was happy when she anticipated an outcome and sad when she saw the actual outcome. By five, children could use a change in valence between the anticipated and observed outcome to infer that the character had originally held a false belief about the contents of the box.

However, much as people rarely have the opportunity to know others’ beliefs and desires before predicting their behavior, people rarely have the opportunity see others’ emotional reactions to both the anticipated and actual outcomes of an event. Here we look at a more minimal context in which children might be able to use others’ emotional expressions to recover their true and false beliefs. We exploit the fact that people can only have an emotional reaction to an event they know about; thus children might be able to use the presence or absence of an emotional reaction to determine whether someone has true or false beliefs about a scenario.

To test this, we adapt the tried and true classic unexpected transfer task (Wimmer & Perner, 1983; Baron-Cohen, Leslie, & Frith, 1985). In that task, Sally puts her marble in one location and Anne moves it during Sally’s absence (False Belief condition). By age four, children correctly predict that when Sally comes back, she will look for her marble in the place where she put it but not the current location. Three-year-olds, by contrast, incorrectly predict that Sally will look for her marble in its current location. Many studies also have a control condition (True Belief condition) in which Anne moves the marble when Sally is present; in that condition, children should predict that Sally will look for her marble in its current location, not the original one.

The extent to which the task itself is a fair assessment of children’s ability to attribute false beliefs is controversial. Although one of the best replicated tasks in the developmental literature (see Wellman, Cross, & Watson, 2001 for a review and meta-analysis), passing the task both requires information processing abilities other than theory of mind and fails to capture many interesting properties of mental state inference (see Bloom & German, 2000 for discussion). Moreover, a burgeoning recent literature suggests that the specific paradigm may under-estimate children’s abilities and much younger children successfully reason about others’ false beliefs given simpler tasks (see Scott & Baillargeon, 2017 for review).

Here however, we are not interested in establishing the onset of children’s false belief understanding but in looking at children’s ability to use emotional cues to infer others’ true and false beliefs. The previous literature suggests that the ability to integrate belief inferences with emotion understanding emerges relatively late in development and typically lags behind false belief understanding (e.g., Bradmeltz & Schneider, 1999; de Rosnay, Pons, Harris, & Morell, 2004; Wu & Schulz, 2018). Here we can see whether four-year-olds are capable of integrating emotion cues to true and false beliefs at least in the context of a well-established paradigm that four-year-olds typically pass.

In our modified version of the task, we establish that Sally really likes her toy and that Anne is her pesky little sister who likes to play tricks on Sally. We do not tell children whether Sally sees Anne move her toy or not. However, there is a window in the room; thus Sally might have seen what Anne has done. Critically, when Sally returns, she is either happy or angry. We predicted that children would be able to use Sally’s emotional expression to decide whether she saw the transfer and thus predict where she would look for her toy. Specifically, we predicted that children would infer that Sally had a false belief when she looked happy and a true belief when she looked annoyed.

**Method**

**Participants**

Sixty-two children between ages four and six (mean = 5.58 years; range: 4.05-6.98 years; 50% girls) were recruited from an urban children’s museum. There were 20 four-year-olds (mean = 4.57 years; range: 4.05-4.96 years), 22 five-year-olds (mean = 5.48 years; range: 5.03-5.90 years) and 20 six-year-olds (mean = 6.66 years; range: 6.08-6.98 years) in this sample. While most of the children were white and middle class, a range of ethnicities and socioeconomic backgrounds reflecting the diversity of the local population (47% European American, 24% African American, 9% Asian, 17% Latino, 4% two or more races) and the museum population (29% of museum attendees receive free or discounted admission) were represented throughout.

**Materials**

Each child saw two story presentations animated by Microsoft PowerPoint on a laptop. Two protagonists, Sally and Anne, were in one story, and another two protagonists, Tom and David, were in the other. One story (randomly chosen) presented the Angry condition, and the other presented the Happy condition. The order of the two
conditions was counterbalanced across participants. All the facial expressions used were selected from iStock photos (http://www.istockphoto.com/).

**Procedure**

Children were tested individually in a quiet room of the children’s museum. All sessions were videotaped. PowerPoint animations were used throughout to visually demonstrate the protagonists’ and objects’ changing locations. Each story was presented as follows (using the Sally-Anne story as an example).

![Angry Condition](Image67x283to176x344)

![Happy Condition](Image67x356to176x417)

<table>
<thead>
<tr>
<th>Scene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sally takes the book out of the drawer, and hides it waaaay under the couch. At Scene 3, the experimenter triggered Scene 4. The silhouette of Sally disappeared outside of the window and Sally appeared with either an angry or happy face. The experimenter said, “Look! Sally is back and she looks very angry/happy.” Finally, the experimenter asked two test questions in a fixed order. One question was about what Sally knew: “Do you think Sally saw what her sister did?” The other question was about what Sally would do next: “Now Sally wants to get her book back. Where do you think she will look for her book?” The experimenter coded children’s responses to the two test questions offline from videotapes. All responses were recoded by an independent coder; there was 97% agreement on children’s responses. Disagreement was resolved by discussion or choosing the conservative response (the one that went against our predictions).</td>
</tr>
<tr>
<td>2</td>
<td>The experimenter showed Scene 1 of the presentation (see Figure 1) and said, “This is Sally, and this is Sally’s favorite book. Today, Sally wants to go play outside, so she puts the book in a drawer, and leaves the room.” At the same time, the experimenter clicked a button to initiate the animation so that the book moved from Sally’s hand to inside the drawer and then Sally moved out of the room. A silhouette of Sally then appeared outside the window of the room. The experimenter drew the child’s attention to the silhouette, saying, “See! This is Sally playing outside!” Then the experimenter initiated Scene 2. She said, “This is Anne. She is Sally’s pesky little sister. She loves to play tricks on Sally, but Sally does not like that at all!” Then the experimenter said, “Today Anne decides to play a trick on Sally. Anne takes the book out of the drawer, and hides it waaaay under the couch. Then she leaves the room.” At the same time, a sequence of animation was triggered to show the book moving under the couch. At Scene 3, the experimenter asked two check questions to make sure that children correctly remembered the book’s current and original locations: “Where is the book now?” (Current Location question) and “Where was the book before?” (Original Location question). Incorrect responses were corrected throughout; however, if a child failed either check question their responses to this story were excluded from further analysis. Then the experimenter said, “Because there is a window in the room, Sally may have seen what was going on inside, or she may not have seen. We don’t know. Let’s see if we can find out!” The experimenter triggered Scene 4. The silhouette of Sally disappeared outside of the window and Sally appeared with either an angry or happy face. The experimenter said, “Look! Sally is back and she looks very angry/happy.”</td>
</tr>
<tr>
<td>3</td>
<td>The experimenter showed Scene 1 of the presentation (see Figure 1) and said, “This is Sally, and this is Sally’s favorite book. Today, Sally wants to go play outside, so she puts the book in a drawer, and leaves the room.” At the same time, the experimenter clicked a button to initiate the animation so that the book moved from Sally’s hand to inside the drawer and then Sally moved out of the room. A silhouette of Sally then appeared outside the window of the room. The experimenter drew the child’s attention to the silhouette, saying, “See! This is Sally playing outside!” Then the experimenter initiated Scene 2. She said, “This is Anne. She is Sally’s pesky little sister. She loves to play tricks on Sally, but Sally does not like that at all!” Then the experimenter said, “Today Anne decides to play a trick on Sally. Anne takes the book out of the drawer, and hides it waaaay under the couch. Then she leaves the room.” At the same time, a sequence of animation was triggered to show the book moving under the couch. At Scene 3, the experimenter asked two check questions to make sure that children correctly remembered the book’s current and original locations: “Where is the book now?” (Current Location question) and “Where was the book before?” (Original Location question). Incorrect responses were corrected throughout; however, if a child failed either check question their responses to this story were excluded from further analysis. Then the experimenter said, “Because there is a window in the room, Sally may have seen what was going on inside, or she may not have seen. We don’t know. Let’s see if we can find out!” The experimenter triggered Scene 4. The silhouette of Sally disappeared outside of the window and Sally appeared with either an angry or happy face. The experimenter said, “Look! Sally is back and she looks very angry/happy.” Finally, the experimenter asked two test questions in a fixed order. One question was about what Sally knew: “Do you think Sally saw what her sister did?” The other question was about what Sally would do next: “Now Sally wants to get her book back. Where do you think she will look for her book?” The experimenter coded children’s responses to the two test questions offline from videotapes. All responses were recoded by an independent coder; there was 97% agreement on children’s responses. Disagreement was resolved by discussion or choosing the conservative response (the one that went against our predictions).</td>
</tr>
<tr>
<td>4</td>
<td>The experimenter showed Scene 1 of the presentation (see Figure 1) and said, “This is Sally, and this is Sally’s favorite book. Today, Sally wants to go play outside, so she puts the book in a drawer, and leaves the room.” At the same time, the experimenter clicked a button to initiate the animation so that the book moved from Sally’s hand to inside the drawer and then Sally moved out of the room. A silhouette of Sally then appeared outside the window of the room. The experimenter drew the child’s attention to the silhouette, saying, “See! This is Sally playing outside!” Then the experimenter initiated Scene 2. She said, “This is Anne. She is Sally’s pesky little sister. She loves to play tricks on Sally, but Sally does not like that at all!” Then the experimenter said, “Today Anne decides to play a trick on Sally. Anne takes the book out of the drawer, and hides it waaaay under the couch. Then she leaves the room.” At the same time, a sequence of animation was triggered to show the book moving under the couch. At Scene 3, the experimenter asked two check questions to make sure that children correctly remembered the book’s current and original locations: “Where is the book now?” (Current Location question) and “Where was the book before?” (Original Location question). Incorrect responses were corrected throughout; however, if a child failed either check question their responses to this story were excluded from further analysis. Then the experimenter said, “Because there is a window in the room, Sally may have seen what was going on inside, or she may not have seen. We don’t know. Let’s see if we can find out!” The experimenter triggered Scene 4. The silhouette of Sally disappeared outside of the window and Sally appeared with either an angry or happy face. The experimenter said, “Look! Sally is back and she looks very angry/happy.” Finally, the experimenter asked two test questions in a fixed order. One question was about what Sally knew: “Do you think Sally saw what her sister did?” The other question was about what Sally would do next: “Now Sally wants to get her book back. Where do you think she will look for her book?” The experimenter coded children’s responses to the two test questions offline from videotapes. All responses were recoded by an independent coder; there was 97% agreement on children’s responses. Disagreement was resolved by discussion or choosing the conservative response (the one that went against our predictions).</td>
</tr>
</tbody>
</table>

![Figure 1. Example stimuli](Image67x501to176x562)

In general, children performed well on the check questions (Current Location question: mean accuracy = .99; Original Location question: mean accuracy = .97). Four children failed the check questions, one on the Angry story and three on the Happy Story; their responses to the test questions of the same story were excluded from further analyses.

We analyzed the Knowledge and Behavior test questions separately. For each question, we looked at the effects of the children’s Age (as a continuous variable) and Emotion (Happy vs. Angry) using a mixed-effects logistic regression model. For the Knowledge question, neither the effect of Age ($\beta = .27, SE = .37, z = .74, p = .461$) nor Emotion ($\beta = 1.98, SE = 3.13, z = .63, p = .526$) was significant. The interaction between Age and Emotion was not significant either ($\beta = -.24, SE = .55, z = -.44, p = .661$). For the Behavior question however, there was a significant effect of Age ($\beta = .68, SE = .34, z = 1.98, p = .048$). Neither the effect of Emotion ($\beta = 1.04, SE = 2.62, z = .40, p = .693$) nor the interaction between Age and Emotion ($\beta = -.22, SE = .47, z = -.47, p = .640$) was significant.

Because there was no effect of Emotion for either test question, we collapsed data across the Happy and Angry conditions in all following analyses. However, because there was a significant effect of Age for the Behavior question, we looked at both children’s performance overall and in age bins. We did this for both test questions to enable comparison between the questions as well as to support comparisons with the previous literature on theory of mind.
(which has largely reported data by age bins rather than as a continuous variable).

For both test questions, children as a whole performed significantly above chance (Knowledge: $M = .81, Z = 5.73, p < .001$; Behavior: $M = .68, Z = 3.89, p < .001$; Exact Wilcoxon-Pratt Signed-Rank Test). At all ages, children succeeded on the Knowledge question (four-year-olds: $M = .78, Z = 3.32, p < .001$; five-year-olds: $M = .82, Z = 3.50, p < .001$; six-year-olds: $M = .83, Z = 3.15, p = .002$). Five and six-year-olds, but not four-year-olds, succeeded on the Behavior question (four-year-olds: $M = .55, Z = .71, p = .727$; five-year-olds: $M = .66, Z = 2.65, p = .016$; six-year-olds: $M = .83, Z = 3.15, p = .002$).

Because four-year-olds showed above chance performance on the Knowledge question but chance performance on the Behavior question, we looked into individual child’s pattern of responses to provide more information. First, we looked at to what extent four-year-olds showed consistent responses between the Knowledge and Behavior questions. Forty-two percent of the responses were correct for the Knowledge and Behavioral questions of the same stories, 10% were incorrect for both, 34% were correct for the Knowledge question but incorrect for the Behavioral question, and 13% were incorrect for the Knowledge question but correct for the Behavioral question. These results suggest that successfully recovering the agent’s knowledge state is not sufficient for four-year-olds to pass the Behavioral question.

We also took a closer look at four-year-olds’ chance performance on the Behavioral question. We are interested in to what extent four-year-olds showed consistent responses to the Behavioral question across the angry and happy conditions. Four-year-olds got the behavioral questions of both conditions right, two got both wrong, six got the angry condition right and the happy condition wrong, and six got the angry condition wrong and the happy condition right. Thus there was no evidence of consistent responding.

These results suggest that children as young as four can use someone’s emotional expression to infer what she does and doesn’t see. However, the ability to use the information about the agent’s true and false beliefs to predict her subsequent behavior improves over development.

**Discussion**

This study suggests that children between ages four and six can use someone’s emotional expression to recover what she does and does not see, and five and six-year-olds, although not four-year-olds, can use this information to decide if she will act consistently with true or false beliefs. These results are consistent with a growing body of work (Repacholi & Gopnik, 1997; Rieffe, Terwogt, & Cowan, 2005; Wellman, Phillips & Rodriguez, 2000; Wu & Schulz, 2017, 2018) suggesting that emotional expressions provide an important entrée into others’ minds and offer a novel contribution to the literature on how children integrate their understanding of emotions with other mental state inferences.

It is intriguing that although four-year-olds could infer whether Sally had or had not seen Anne move the book (the Knowledge question), they failed to use this information to predict where Sally would look for her book (the Behavior question). One possibility is that four-year-olds used Sally’s emotional expression to determine whether or not she knew she had been tricked, but did not realize that this should lead Sally to update her beliefs or recognize that these updated beliefs should be the bases for predicting her future actions. However, like classic false belief tasks, our study makes.

---

1 The remaining two four-year-olds’ responses in one of the two conditions were excluded from further analysis because of failure to pass the check questions.
high demands on children’s information processing abilities. In order to answer the behavior question, children have to use the person’s emotional expression to infer her knowledge about the event (i.e., that Anne moved her book), use that to reason about whether she had a true or false belief about the location of her book, and then use that to predict her action. These task demands may have overwhelmed four-year-olds and masked their underlying competence. One reason to think this might be the case is that although four-year-olds failed the behavior question, they did not “fail” in the way that three-year-olds do in the classic version of the task: by systematically assuming she would look in the actual location of the toy. Instead, four-year-olds performed at chance in both the true and false belief conditions in this study. Although we cannot make anything much of children’s chance responding (they may have been confused by the behavior question), it is nonetheless curious that four-year-olds did not simply default to the actual location of the book. Future work is warranted to differentiate competence and performance account of four-year-olds’ failure to use emotional expressions to predict agent’s behavior.

However, at least for the older children, our results suggest that children can use emotional expressions to succeed on the classic unexpected transfer task. Our findings that children between ages four and six were increasingly likely to pass the Behavior question are largely consistent with previous work suggesting that there is a gradual development between children’s ability to pass the classic false belief tasks (around age four) and connect false beliefs to emotions (e.g., Bradmetz & Schneider, 1999; de Rosnay, Pons, Harris, & Morell, 2004; Hadwin & Perner, 1991; Harris, et al., 1989; Pons, Harris, & de Rosnay, 2004; Ruffman & Keenan, 1996; Wellman & Bartsch, 1988; Wu & Schulz, 2018).

It would be interesting for future research to study our question of interest with different paradigms appropriate for younger children, including violation-of-expectation, anticipatory-looking, and preferential-looking tasks. For example, if we show infants that Sally comes back with an angry face, will they predict that Sally will look for her toy in its current location by anticipatorily looking towards the current location of her toy, or by looking longer if Sally goes to its original location than if she goes to the current? By contrast, if Sally comes back with a happy face, will infants predict the opposite? Although the current study is not designed to investigate the onset of false belief understanding, our approach of adding emotional cues to false belief tasks can potentially provide novel insights into current debates about when young children start to represent false beliefs. There are currently a number of alternative explanations of the findings with infants. For example, infants may pass false belief tasks simply by using the behavioral rule that an agent will always look for her toy where she last saw it without representing her false beliefs (e.g., Ruffman & Perner, 2005). If this is true, infants will have the same predictions regardless of Sally’s facial expression when she comes back. However, if infants truly represent Sally’s mental states, they will make opposite predictions about Sally’s behavior depending on whether she looks angry or happy when she returns. Thus, future work could design infant-versions of our task, which may help confirm or disconfirm some low-level interpretations of previous findings with infants, and help resolve the longstanding debates about the onset of false belief understanding.

The current results however, suggest that at least by age five, children can use an agent’s emotional expression to recover whether she has seen an event or not, and then use this to predict whether the agent’s subsequent actions will be governed by true or false beliefs. Consistent with a small set of studies showing that young children can use observed emotional cues to gain mental state information that is otherwise unknown (Repacholi & Gopnik, 1997; Rieffe, Terwogt, & Cowan, 2005; Wellman, Phillips & Rodriguez, 2000; Wu & Schulz, 2017, 2018), our study suggests that emotional expressions provide a valuable entree into others’ minds.

**Acknowledgments**

This study is supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216. Warm thanks to the Boston Children’s Museum and participating parents and children. Thanks to Rachel Magid, Junyi Chu, Maddie Pelz, and Kary Richardson for helpful feedback, and to Anya Keomurjian for help with data collection and coding.

**References**


Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition, 130*(2), 204-216.


