PAPER

False-belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks

Rose M. Scott,[1] Zijing He,[2] Renée Baillargeon[3] and Denise Cummins[3]

1. *School of Social Sciences, Humanities and Arts, University of California Merced, USA*
2. *Department of Psychology, Sun Yat-sen University, China*
3. *Department of Psychology, University of Illinois, USA*

Abstract

Recent research indicates that toddlers and infants succeed at various non-verbal spontaneous-response false-belief tasks; here we asked whether toddlers would also succeed at verbal spontaneous-response false-belief tasks that imposed significant linguistic demands. We tested 2.5-year-olds using two novel tasks: a preferential-looking task in which children listened to a false-belief story while looking at a picture book (with matching and non-matching pictures), and a violation-of-expectation task in which children watched an adult 'Subject' answer (correctly or incorrectly) a standard false-belief question. Positive results were obtained with both tasks, despite their linguistic demands. These results (1) support the distinction between spontaneous- and elicited-response tasks by showing that toddlers succeed at verbal false-belief tasks that do not require them to answer direct questions about agents' false beliefs, (2) reinforce claims of robust continuity in early false-belief understanding as assessed by spontaneous-response tasks, and (3) provide researchers with new experimental tasks for exploring early false-belief understanding in neurotypical and autistic populations.

Introduction

As adults, we recognize that others' behavior is sometimes guided by incorrect representations of reality, or false beliefs. Investigations of young children's ability to attribute false beliefs to others have uncovered substantial developmental differences across tasks, with children succeeding at some tasks long before others. In particular, *elicited-* and *spontaneous-response* tasks have yielded highly divergent results. In both types of task, children are presented with a scene involving a mistaken agent who holds a false belief about some aspect of the scene. In elicited-response tasks, children are asked a direct question about how the mistaken agent will act; in spontaneous-response tasks (such as violation-of-expectation and anticipatory-looking tasks), children's spontaneous responses to the unfolding scene are measured instead. When given elicited-response tasks, children often do not succeed until about age 4 (e.g. Baron-Cohen, Leslie & Frith, 1985; Gopnik & Astington, 1988; Perner, Leekam & Wimmer, 1987; Wellman & Bartsch, 1988; Wellman, Cross & Watson, 2001; Wimmer & Perner, 1983). In contrast, positive results have been obtained in spontaneous-response tasks with toddlers in the third year of life (e.g. He, Bolz & Baillargeon, 2011; Southgate, Senju & Csibra, 2007) as well as with infants in the first two years of life (e.g. Kovács, Téglás &

Endress, 2010; Onishi & Baillargeon, 2005; Scott, Baillargeon, Song & Leslie, 2010; Song & Baillargeon, 2008; Surian, Caldi & Sperber, 2007; Träuble, Marinović & Pauen, 2010; for a review, see Baillargon, Scott & He, 2010).

The many positive results that have been obtained in spontaneous-response tasks with toddlers and infants seem to paint a highly consistent picture of early false-belief understanding. One possible exception, however, comes from *verbal* spontaneous-response tasks. In contrast to the spontaneous-response tasks cited above, all of which either involved very simple language or were entirely non-verbal, verbal spontaneous-response tasks impose significant linguistic demands, often comparable to those in elicited-response tasks. Although verbal spontaneous-response tasks have generally produced positive findings with 3- to 4-year-olds (e.g. Clements & Perner, 1994; Low, 2010; Ruffman, Garnham, Import & Connolly, 2001), results with toddlers under age 3 have been far more mixed (e.g. Clements & Perner, 1994; Clements, Rustin & McCallum, 2000; Garnham & Perner, 2001; Garnham & Ruffman, 2001). These mixed results bear critically on theories of early false-belief understanding: if toddlers fail at verbal spontaneous-response tasks (though they possess sufficient linguistic skills to comprehend the language used in the tasks), this could be taken (1) to signal an important discontinuity in

early false-belief knowledge as assessed by spontaneous-response tasks and (2) to support the notion that verbal tasks tap a more advanced form of false-belief understanding than do non-verbal tasks. In the following section, we describe the results that have been obtained to date with toddlers in verbal spontaneous-response tasks. We then introduce the present research, which investigated the question of whether toddlers can succeed at verbal spontaneous-response tasks by testing 2.5-year-olds in two novel tasks.

Prior findings with verbal anticipatory-looking tasks

In a series of experiments, Garnham (née Clements) and her colleagues tested 2- to 4-year-olds in verbal spontaneous-response false-belief tasks that measured anticipatory looking (e.g. Clements & Perner, 1994; Clements *et al.*, 2000; Garnham & Perner, 2001; Garnham & Ruffman, 2001). In a typical experiment, children heard a story, enacted with props, in which a character first hid an object in one location and then left; in the character's absence, the object was moved to a different location. When the character returned to look for the object, an experimenter uttered a self-addressed prompt such as 'I wonder where she's going to look'. Children's spontaneous looks following the prompt were recorded and coded. In her first experiment, Garnham tested 29- to 54-month-olds and found a sharp age effect (Clements & Perner, 1994): 88% of the children aged 35 months and older looked at the object's *original* location, thus correctly anticipating where the character's false belief would lead her to search; in contrast, 77% of the children under 35 months looked at the object's *current* location, suggesting that they did not represent the character's false belief. In her second experiment (Clements *et al.*, 2000), Garnham used a slightly different procedure and again obtained an age effect: whereas 71% of the children aged 43 to 47 months looked mainly at the object's original location following the anticipatory prompt, only 26% of the children aged 34 to 42 months did so (no further details were provided about the responses of the unsuccessful children). In her next two experiments, Garnham did not report age effects. In one experiment, 29- to 55-month-olds ($M = 39$ months) were tested in two similar conditions, again using slightly different procedures (Garnham & Perner, 2001); overall performance was marginally significant in one condition, with 62% of the children showing correct anticipation, but was at chance in the other condition, with only 53% of the children doing so.[1] In the other experiment, 25- to 49-month-olds ($M = 37$ months) were tested using a

slightly different procedure, and 72% of the children looked more at the object's original location (Garnham & Ruffman, 2001).

The preceding results make it difficult to arrive at any firm conclusions about whether toddlers in the third year of life can succeed at verbal spontaneous-response tasks. On the one hand, the fact that significant or nearly significant findings were obtained with two samples that included children under age 3 provides suggestive evidence that toddlers – when assessed on their own – might succeed at verbal spontaneous-response tasks. On the other hand, the fact that inconsistent results were obtained across experiments suggests that either (1) toddlers' performance at verbal spontaneous-response tasks is far from robust or (2) slight procedural variations in verbal anticipatory-looking tasks can greatly increase their demands and render them overly challenging for toddlers.

The present research

Can toddlers in the third year of life succeed at verbal spontaneous-response false-belief tasks? The present research was designed to address this question. Given the interpretive difficulties associated with the experiments reviewed in the last section, we chose to test only toddlers (all children were between 28 and 34 months of age), and we devised two novel verbal spontaneous-response tasks: a *preferential-looking* task in which children listened to a false-belief story while looking at a picture book (with matching and non-matching pictures), and a *violation-of-expectation* task in which children watched an adult 'Subject' answer (correctly or incorrectly) a false-belief question.

We reasoned that evidence that toddlers failed at the present tasks would undermine claims of consistency in early false-belief reasoning as assessed by spontaneous-response tasks. Prior research indicates that children with more advanced verbal skills perform better at elicited-response tasks (e.g. de Villiers & Pyers, 2002; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Low, 2010; Milligan, Astington & Dack, 2007; Peterson, Wellman & Liu, 2005; Slade & Ruffman, 2005); if toddlers were found to often perform poorly at verbal spontaneous-response tasks, this might again point to significant interactions in development between language and false-belief understanding. Such results could be taken to support proposals that false-belief knowledge is initially rudimentary and undergoes important conceptual or representational changes – fueled in part by linguistic advances – in the course of early childhood (e.g. Apperly & Butterfill, 2009; Low, 2010; Perner, 2010; Ruffman & Perner, 2005).

On the other hand, evidence that toddlers succeeded at the present tasks, despite their substantial linguistic demands, (1) would support the distinction between spontaneous- and elicited-response tasks by showing that toddlers succeed at verbal false-belief tasks that do not

---

[1] Garnham and Perner (2001) conducted a second experiment with 71 29- to 51-month-olds ($M = 39$ months). Of these, 24 were tested with an anticipatory-looking task, and 67% showed correct anticipation. Unfortunately, the specific age range of these 24 children was not mentioned in the report, so that it is unclear whether the sample included children under age 3.

require them to answer direct questions about mistaken agents' likely actions, (2) would reinforce claims of robust continuity in early false-belief reasoning as assessed by spontaneous-response tasks, and finally (3) would provide researchers with new experimental tasks for exploring early false-belief understanding in typical and atypical populations.

Experiment 1

The first verbal spontaneous-response false-belief task we devised was a *preferential-looking* task: it took advantage of children's and adults' well-established tendency to look spontaneously at scenes that match the utterances they hear (e.g. Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). In Experiment 1, 2.5-year-olds listened to a typical false-belief story in which a character named Emily hid an apple in one of two containers; in her absence, the apple was moved to the other container. The story was accompanied by a large picture book (we adapted a setup used by Cummins, 2007, to study toddlers' understanding of social rules). Each double-page in the book showed two pictures: one matched the narrative and one was irrelevant. In the final double-page, one picture showed Emily searching for her apple where she falsely believed it to be (original-location picture), and the other picture showed Emily searching for her apple in its current location (current-location picture). While viewing this final double-page, children heard 'Emily is looking for her apple'. We reasoned that if children could follow the story and represent Emily's false belief, they should look reliably longer at the original-location picture, which matched the final line of the story. Such a result would indicate that toddlers could succeed at a spontaneous-response false-belief task even if it imposed significant linguistic demands.

We also included a knowledge condition that was identical to the false-belief condition except that Emily watched Sarah move the apple. This condition helped to rule out the possibility that children simply formed an association between Emily and the container she manipulated. If the children simply formed such an association, then the children in both the false-belief and the knowledge conditions should look longer at the original-location picture in the test trial. If, however, the children were actually reasoning about Emily's belief about the apple's location, then the children in the false-belief condition should look longer at the original-location picture, whereas those in the knowledge condition should look longer at the current-location picture.

Method

Participants

Participants were 28 native English-speaking toddlers, 14 male and 14 female (29 months, 26 days to 33 months, 19

days, M = 31 months, 17 days). Another 10 children were tested but excluded, three because they were overly active, two because the difference between their test looking times was over 2.5 standard deviations from the mean of their condition, one because she did not look at the pictures in which the apple was hidden and moved, one because he looked more than 85% of the time at the left picture during the setup trials, one because she failed to look at the matching picture on 4/5 setup trials, one because her hair obscured her eyes, and one because of parental interference. Half the children were randomly assigned to the false-belief condition (M = 31 months, 24 days) and half to the knowledge condition (M = 31 months, 10 days).

Apparatus

Children sat on a parent's lap at a table with a picture book raised at a 25° angle in front of them; an experimenter sat to their left. A first camera captured the children's eye movements, a second camera captured the stimuli, and a third camera captured the experimenter.

The picture book (see Figure 1) rested on a plastic bookstand (30 cm wide × 24 cm tall). Attached to the bookstand was a piece of foam board (61 cm × 22.5 cm) that provided a stable surface for the double-pages to rest on when turned towards the child. The double-pages were attached to the top of the bookstand with three binder rings. Each double-page (55 cm × 22.5 cm) was composed of two clear plastic photo sheets attached in the center; color photos (20 cm × 25 cm) were inserted in the sheets.

The picture book consisted of eight double-pages. In the first two, one side displayed a picture and the other side displayed white paper. In the subsequent six double-pages, one side displayed a picture that matched the story, and the other side displayed an irrelevant picture.

Procedure

To start, the book's double-pages were face down behind the bookstand, away from the children. In each trial,

Figure 1   *Picture book used in Experiment 1 (left) and a child participating in the task (right).*

the experimenter recited a line of the story, flipped a double-page towards the children so that the pictures became visible, and then repeated the line of the story. The experimenter paused for about 3 s before reciting the next line of the story. During each trial, the experimenter gazed at a neutral location in front of her so that children could not use her gaze as a cue for finding the matching picture.

The story was organized into introduction, setup, and test trials (see Appendix for pictures and script). The introduction trials presented the main character, Emily (introduction-1), and her friend Sarah (introduction-2); only one picture was shown in each trial. In the setup trials in the *false-belief condition*, children were told that Emily had an apple (setup-1) and that she hid it (setup-2) in one of two containers, a basket and a box (hiding container was counterbalanced across children); Sarah watched as Emily hid her apple. Emily then went to take a nap (setup-3). While Emily was sleeping, Sarah moved the apple to the other container (setup-4) and then went outside to play (setup-5). Side of the matching picture in the introduction and setup trials was counterbalanced across trials and across children: in each trial, about half the children in each hiding-container condition saw the matching picture on the left, and about half saw it on the right.

The story ended with a test trial in which Emily looked for her apple; Emily searched in the basket in one picture and in the box in the other picture. Side of the matching picture in the test trial was counterbalanced across sex, hiding-container condition, and across side condition in the introduction and setup trials.

The *knowledge condition* was identical to the false-belief condition except that the script was different for three trials (see Appendix): instead of going to take a nap, Emily sat down to read a book (setup-3), watched Sarah move the apple (setup-4), and became hungry (test trial). The new matching picture for setup-4 showed Emily watching Sarah move the apple.

Coding

In each two-picture trial, we coded where the children looked (left picture, right picture, away) frame-by-frame from silent video. All children were coded independently by two coders, who agreed on the children's direction of gaze for 94% of coded video frames. Trials in which agreement was less than 90% (24/168 trials) were resolved through discussion.

For the five setup trials, we coded the first 6 s that the pictures were visible to the children; this 6-s window ended during the pause after the story line was repeated, prior to the story line for the next trial. Children easily followed the story: they looked reliably longer at the matching than at the non-matching picture during the five setup trials, $F(1, 26) = 71.12$, $p < .00025$. This was true in the false-belief condition (matching: $M = 3.52$, $SD = 0.49$; non-matching: $M = 2.21$, $SD = 0.56$), $F(1, 26)$

$= 25.56$, $p < .0001$, with 13/14 children showing this pattern, Wilcoxon signed-ranks $T = 1$, $p < .00025$, and in the knowledge condition (matching: $M = 3.78$, $SD = 0.51$; non-matching: $M = 2.00$, $SD = 0.46$), $F(1, 26) = 47.19$, $p < .0001$, with 14/14 children showing this pattern, $T = 0$, $p < .0001$.

The test trial included a 2-s preview followed by a 6-s window. Previews are commonly used in preferential-looking tasks with children and adults, as they allow participants to inspect the visual scene before responding to the test sentence (e.g. Dahan & Tanenhaus, 2005; Waxman, Lidz, Braun & Lavin, 2009). Because our two test pictures were highly similar, differing only in the direction of Emily's reach, we reasoned that children might need a few seconds to inspect them. All children looked at both pictures at some point during the course of the test trial. Analysis of the 2-s preview indicated that children in both conditions looked about equally at the original- (false-belief: $M = 0.89$, $SD = 0.81$; knowledge: $M = 0.61$, $SD = 0.70$) and current-location (false-belief: $M = 0.95$, $SD = 0.85$; knowledge: $M = 0.96$, $SD = 0.80$) pictures, both Fs < 1. In the 6-s window, we again compared how long children in each condition looked at the two test pictures.

Preliminary analyses revealed no interaction of condition and test picture with sex, hiding-container condition, side condition in the introduction and setup trials, or side condition in the test trial; the data were therefore collapsed across these factors in subsequent analyses.

Results and discussion

Children's looking times during the 6-s test window were analyzed using an analysis of variance (ANOVA) with condition (false belief or knowledge) as a between-subjects factor and picture (original- or current-location) as a within-subject factor. Only the interaction of condition and picture was significant, $F(1, 26) = 23.12$, $p < .0001$: children in the false-belief condition looked reliably longer at the original- ($M = 3.62$, $SD = 1.45$) than at the current-location ($M = 1.79$, $SD = 1.09$) picture, $F(1, 26) = 11.01$, $p < .005$, whereas those in the knowledge condition looked reliably longer at the current- ($M = 3.73$, $SD = 0.95$) than at the original-location ($M = 1.80$, $SD = 1.05$) picture, $F(1, 26) = 12.24$, $p < .0025$. In the false-belief condition, 10/14 children looked longer at the original-location picture, $T = 14$, $p < .025$; in the knowledge condition, 12/14 children looked longer at the current-location picture, $T = 8.5$, $p < .005$.

In both conditions, children looked reliably longer at the picture that matched the final line of the story. When Emily did not see her apple moved, children attributed to her the false belief that her apple was still in its original location: upon hearing the final line of the story, 'Emily is looking for her apple', they looked longer at the picture in which Emily acted on her false belief and searched for her apple in its original location. In contrast, when

Emily saw her apple moved, children looked longer at the picture in which she searched for her apple in its current location.

Further results

To confirm the findings of the false-belief condition in Experiment 1, we replicated this condition with an additional set of 14 native English-speaking toddlers, eight male and six female (29 months, 26 days to 34 months, 1 day, M = 31 months, 17 days). Another three children were tested but excluded, two because they did not look at the pictures in which the apple was hidden and moved, and one because the difference between his test looking times was over 2.5 standard deviations from the mean. The apparatus and procedure were identical to those in the false-belief condition except that new pictures featuring different women were used. Analysis of the setup trials showed that the children again looked reliably longer at the matching (M = 3.47, SD = 0.64) than at the non-matching (M = 2.40, SD = 0.69) picture across trials, $F(1, 13)$ = 9.30, $p < .01$, and 12/14 children showed this pattern, T = 11.5, $p < .01$. The test trial again included a 2-s preview followed by a 6-s window. All children looked at both pictures during the test trial. During the 2-s preview, children looked about equally at the original- (M = 1.19, SD = 0.80) and current-location (M = 0.66, SD = 0.69) pictures, $F(1, 13)$ = 1.90, $p = .19$.

Children's looking times during the 6-s test window were analyzed using an ANOVA with picture (original- or current-location) as a within-subject factor. The main effect of picture was significant, $F(1, 13)$ = 8.29, $p < .025$: children looked reliably longer at the original- (M = 3.60, SD = 1.03) than at the current-location (M = 2.18, SD = 0.94) picture, and 11/14 children showed this pattern, T = 10.5, $p < .01$.[2] These results thus confirm those of the false-belief condition in Experiment 1. Children attributed to Emily the false belief that her apple was still in its original location, and they looked longer at the picture in which she acted on this false belief.

Adult stories

Could children have succeeded at our task without listening or attending to the story they were told? Could they have gleaned the key elements of the story from the pictures themselves? This possibility seemed to us unlikely: recall that 39 of the 42 2.5-year-olds we tested looked longer overall at the matching than at the non-matching picture in the five setup trials; such a consistent

pattern strongly suggests that children's responses were guided by the story they heard rather than by educated guesses about what the story might be.

To evaluate how easy or difficult it might be to deduce the stories we told from the pictures alone, we showed the pictures from Experiment 1 to 18 naïve adults (15 female, 18 to 57 years, M = 22.6); half saw the false-belief pictures, and half saw the knowledge pictures. In each condition, participants were asked to tell the story they thought accompanied the pictures; narratives were video recorded, transcribed, and coded. Only 9/18 adults implied that Emily was looking for an apple in the test trial; the remaining subjects described the test pictures in other, irrelevant ways (e.g. 'she's hiding something in the basket', 'the day is done and she's putting everything away', 'maybe she's teaching them about colors'). Of the nine adults who intimated that Emily was looking for an apple in the test trial, five were in the false-belief condition and hinted that Emily was either mistaken (n = 3), ignorant (n = 2), or correct (n = 1) about the apple's location; the other four adults were in the knowledge condition and suggested that Emily was either correct (n = 2) or ignorant (n = 2) about the apple's location. Adults' highly varied responses suggest that the key elements of each story were not obvious from the pictures themselves, and make it unlikely that the 2.5-year-olds in Experiment 1 could have looked preferentially at the matching pictures in the setup and test trials without listening to the story.

Comparison to previous verbal false-belief tasks

Could children have succeeded at our task because its linguistic demands were weaker than those of previous verbal (spontaneous- or elicited-response) false-belief tasks (e.g. Baron-Cohen *et al*., 1985; Clements & Perner, 1994; Low, 2010; Ruffman *et al*., 2001; Wellman & Bartsch, 1998; Wimmer & Perner, 1983)? Comparison of the language used in our task and in previous tasks suggests that this is unlikely. For example, Wellman and Bartsch (1988) found that 3.5-year-olds failed at an elicited-response task in which they were presented with the following scenario: 'Jane wants to find her kitten. Jane's kitten is really in the playroom, but Jane thinks the kitten is in the kitchen. Where will Jane look for her kitten?' Most children answered that Jane would look for her kitten in its current location, the playroom. It would be difficult to argue that this scenario is more linguistically demanding than the (considerably longer) story used in our task.

While the global linguistic demands of our task were not weaker than those of previous verbal false-belief tasks, it is possible that our task was nevertheless easier for children due to a critical difference in the language used in the test trial. Csibra and Southgate (2006) suggested that when young children hear sentences such as 'Where is she going to look?' in verbal false-belief tasks, they prematurely interpret the word 'where' as asking

---

[2] One could also ask whether children's performance differed significantly from chance. If we assume chance level to be looking equally at the two pictures, then all three conditions in Experiment 1 differed significantly from chance: false-belief condition: $t(13)$ = 3.10, $p = .008$; knowledge condition: $t(13)$ = 4.09, $p = .001$; further results: $t(13)$ = 2.96, $p = .011$).

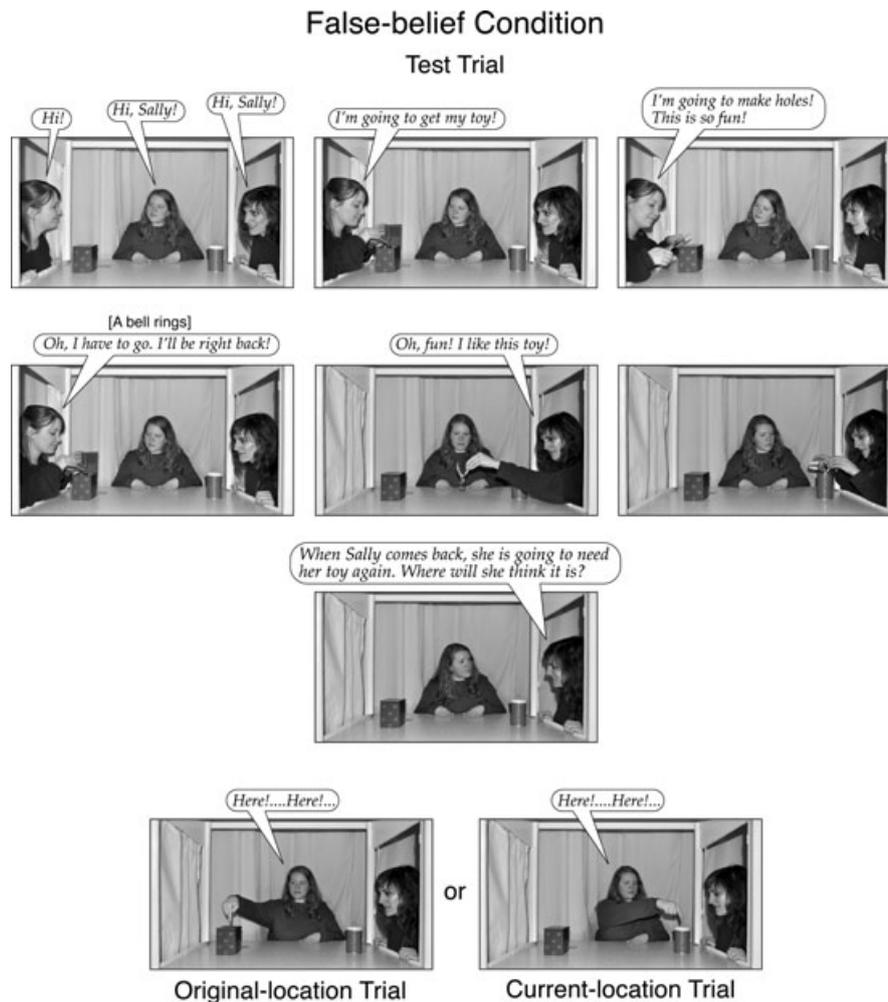## False-belief Condition
Test Trial

Figure 2   *Summary description of the events shown in the test trial of the false-belief condition in Experiment 2. The three assistants served as Sally (on the left), the Experimenter (on the right), and the Subject (at the back). The familiarization trial began in the same manner as the test trial, but ended after Sally returned her toy (a hole-puncher) to the left container and left the room. The familiarization and test events shown in the knowledge condition were similar with one exception: in the test trial of the false-belief condition, the Experimenter moved Sally's toy to the right container after Sally left the room; in the knowledge condition, Sally watched the Experimenter place the toy in the right container before she left the room.*

about the true location of the object; this pragmatic difficulty with the word 'where' causes them to indicate the object's actual location, thus responding incorrectly (for a different account of how young children's limited pragmatic understanding might contribute to their poor performance in verbal false-belief tasks, see Hansen, 2010). In our task, children did not hear the word 'where' in the test trial: the test sentence simply stated that Emily was looking for her apple. It could be that removing the word 'where' substantially reduced the pragmatic linguistic difficulty of the test trial, thus enabling children to succeed. We addressed this possibility in Experiment 2.

Experiment 2

The second verbal spontaneous-response false-belief task we devised was a *violation-of-expectation* task. In

Experiment 2, 2.5-year-olds watched live events in which an adult answered, correctly or incorrectly, a typical false-belief test question; the task examined whether children would detect a violation (as indexed by reliably longer looking times) when the adult answered incorrectly.

Children received a familiarization and a test trial involving three female assistants: one served as *Sally* (affectionately named after the main character in the classic Sally-Anne task of Baron-Cohen *et al.*, 1985), one served as the *Experimenter*, and one served as the *Subject* (see Figure 2). During the familiarization trial, while the Experimenter and the Subject watched, Sally played with a toy, hid it in one of two containers, and then left the room. The test trial began in the same manner; after Sally left, the Experimenter moved the toy to the other container and then asked the Subject where Sally would look for her toy when she returned. At that point, the

Subject pointed to either the original-location container (original-location trial) or the current-location container (current-location trial). If children (1) represented Sally's false belief, (2) expected the Subject to also represent Sally's false belief, and (3) correctly interpreted the question that was addressed to the Subject despite the fact that it involved the word 'where', then they should expect the Subject to answer the question by pointing to the original-location container, where Sally's false belief would lead her to search, and they should detect a violation when she pointed to the current-location container instead. Children who received the current-location trial should thus look reliably longer than those who received the original-location trial.

As in Experiment 1, we also included a knowledge condition to ensure that children were not simply responding based on an association between Sally and the container she manipulated. This condition was identical to the false-belief condition except that Sally witnessed her toy being moved to the other location before she left the room. Children in this condition should expect the Subject to point to the current-location container, where the toy was now hidden, and they should detect a violation when she pointed to the original-location container instead. Children who received the original-location trial should therefore look reliably longer than those who received the current-location trial. Opposite looking patterns were thus predicted for the false-belief and knowledge conditions.

Method

Participants

Participants were 28 native English-speaking toddlers, 14 male and 14 female (28 months, 18 days to 33 months, 20 days, M = 30 months, 22 days). Another two children were tested but excluded, one because she was active and one because of parental interference. Half the children were randomly assigned to the false-belief condition (M = 30 months, 7 days) and half to the knowledge condition (M = 31 months, 6 days). Within each condition, half the children were randomly assigned to receive the original-location test trial, and half the current-location test trial.

Apparatus

The apparatus consisted of a wooden display booth (127.5 cm high × 101 cm wide × 73.5 cm deep) mounted 76 cm above the floor of a brightly lit test room. The child sat on a parent's lap and faced a large opening (44 cm × 93.5 cm) in the front of the apparatus; between trials, a muslin-covered frame (61 cm × 99.5 cm) was lowered to hide this opening. The back and side walls of the apparatus were painted white; the floor extended 11.5 cm behind the back wall and was covered with granite-patterned contact paper.

Sally, the Experimenter, and the Subject (who all spoke during the trials) were native English speakers. Sally wore a blue shirt; at the start of each trial, she knelt at a window (58 cm × 48 cm) in the left wall of the apparatus, behind a muslin curtain that could be drawn to the side to open the window. The Experimenter wore a green shirt and knelt at an open window (57 cm × 48 cm) in the right wall of the apparatus. The Subject wore a red shirt and sat at an open window (60 cm × 101 cm) in the back wall. A large white curtain behind the apparatus hid the test room.

In each trial, a box and a can stood on the apparatus floor, 32 cm from the front edge of the apparatus. One container stood on the left, 8 cm from the left window, while the other stood on the right, 8 cm from the right window. The side of the box and can was counterbalanced across sex, condition, and test trial. The box (12.5 cm × 9 cm × 12 cm) was made of foam board, covered in green contact paper, and decorated with red hearts; a lid was attached to the rear of the box. The can (12.5 cm × 8 cm in diameter) was covered with blue contact paper and decorated with red stars; it was closed by a yellow lid with a wooden knob (1.5 cm in diameter) affixed to its top surface. Sally's toy was a silver hole-puncher (13 cm × 7 cm × 2 cm at its largest points). A small piece of orange paper (10 cm × 7 cm) rested on the apparatus floor behind the left container.

Trials

Each trial began with the raising of the curtain. During each trial, the assistants spoke in a clear, child-directed manner. They did not look at the child but simply followed each other's actions.

False-belief condition. At the start of the familiarization trial (which lasted about 61 s), the Experimenter and the Subject were present at their windows, facing the box and can; whichever container was on the left, near Sally's curtained window, held the hole-puncher, and behind the container was the small piece of paper. Sally opened her window and was greeted by the Experimenter and the Subject in turn, 'Hi, Sally!' Sally acknowledged each greeting ('Hi!') and then announced, 'I am going to get my toy!' She removed the hole-puncher from the left container, picked up the piece of paper, and said, 'I am going to make holes! This is so fun!' Next, Sally proceeded to make five holes in the paper. At that point, a bell rang and Sally announced, 'Oh! I have to go! I'll be right back.' She placed the hole-puncher back into the left container, closed her window, and left the room (children heard the room door open and close and so could infer that she had departed). At this point, the trial ended.

The test trial was divided into an initial and a final phase. The initial phase (which lasted about 91 s) began like the familiarization trial but continued as follows. After Sally left the room, the Experimenter reached across the apparatus and removed the hole-puncher from

the left container. She stated, 'Oh fun! I like this toy!' and placed it into the right container. The Experimenter then turned to the Subject and said, 'When Sally comes back, she is going to need her toy again. Where will she think it is?' In response to this prompt, the Subject pointed to either the left container (original-location trial) or the right container (current-location trial) and said, 'Here!' During the final phase of the trial, the Subject continued pointing and proclaimed 'Here!' once every 2 seconds until the trial ended.

Knowledge condition. The trials in the knowledge condition were identical to those in the false-belief condition, with the following exceptions. In the test trial, as Sally began putting the hole-puncher away, the Experimenter asked, 'Can I see it?' Sally then handed her the hole-puncher. The Experimenter said, 'Oh fun! I like this toy!' and placed it into the right container, telling Sally, 'I put it in here' as she did so. Sally acknowledged this by saying 'OK' and then added 'I'll be right back!' and departed. The Experimenter then turned to the Subject to ask the test question; from that point on, the trial proceeded as in the false-belief condition.

Procedure

Two observers monitored the child's looking behavior through peepholes in large cloth-covered frames on either side of the apparatus; neither observer could reliably guess which test trial children received. Each observer depressed a button linked to a computer when the child looked at the events shown during a trial; looking times were computed using the primary observer's responses. Sessions were also video recorded and checked offline for accuracy.

Children in both conditions were highly attentive during the 61-s familiarization trial and looked for 96% of the time on average. During the test trial, looking times during the initial and final phases of the trial were computed separately. Children were again highly attentive during the 91-s initial phase of the trial and looked for 98% of the time on average. The final phase of the trial ended when children either (1) looked away for 0.5 consecutive seconds after having looked for at least 12 cumulative seconds or (2) looked for 60 cumulative seconds without looking away for 0.5 consecutive seconds.

To assess interobserver agreement, each child's session was divided into 100-ms intervals, and the computer determined in each interval whether the two observers agreed that the child was or was not looking at the event. Percent agreement was calculated by dividing the number of intervals in which the observers agreed by the total number of intervals in the session. Agreement was measured for all 28 participants and averaged 97% per child.

Preliminary analyses of the test data revealed no significant effects of sex or container; the data were therefore collapsed across these factors in subsequent analyses.

Results and discussion

Children's looking times during the final phase of the test trial were analyzed using an ANOVA with condition (false-belief or knowledge) and trial (original- or current-location) as between-subjects factors. Only the interaction of condition and trial was significant, $F(1, 24) = 19.37$, $p < .00025$. Children in the false-belief condition looked reliably longer if they received the current-location trial ($M = 40.8$, $SD = 17.7$) as opposed to the original-location trial ($M = 20.1$, $SD = 7.4$), $F(1, 24) = 10.22$, $p < .005$. This pattern reversed in the knowledge condition, where children looked reliably longer if they received the original-location trial ($M = 41.6$, $SD = 13.4$) as opposed to the current-location trial ($M = 21.9$, $SD = 6.4$), $F(1, 24) = 9.17$, $p < .01$. Wilcoxon rank-sum tests confirmed these findings: children looked reliably longer if they received the current-location trial in the false-belief condition, $Ws = 34$, $p < .025$, and the original-location trial in the knowledge condition, $Ws = 31$, $p < .01$.

In the false-belief condition, children were able to keep track of Sally's false belief, and they expected the Subject to do so as well. When the Subject was asked where Sally would search for her toy, children expected the Subject to point to the original-location container, and they detected a violation if she pointed to the current-location container instead. In the knowledge condition, children realized that Sally had observed the transfer of the toy to its new location, and they expected the Subject to be aware of this as well. Thus, they expected the Subject to point to the current-location container in answer to the test question, and they detected a violation if she pointed to the original-location container instead.

When asked a *direct* question about where an agent who has a false belief about a toy's location will search for the toy, young children often fail: they simply point to the toy's current location (e.g. Wellman *et al.*, 2001). The present results indicate that, when the same question is addressed to *someone else*, even 2.5-year-olds give evidence that they know the correct answer to the question: children in the false-belief condition expected the Subject to correctly predict where Sally's false belief would lead her to search, and they looked reliably longer when this expectation was violated. The results of Experiment 2 thus extend those of the verbal preferential-looking task in Experiment 1 and indicate that toddlers can also demonstrate false-belief understanding in a verbal violation-of-expectation task. Finally, the present results cast doubt on the suggestion that young children fail at verbal false-belief tasks because they prematurely interpret the word 'where' in the test question as asking about the true location of the toy. Although children in the false-belief condition of Experiment 2 heard a 'where' question, they expected the Subject to point *not* to the container that currently held the toy, but rather to the container where Sally falsely believed her toy was hidden.

General discussion

The present experiments examined whether 2.5-year-old toddlers could succeed at two novel verbal spontaneous-response false-belief tasks.

In Experiment 1, children were tested in a *verbal preferential-looking task*: they viewed a picture book while listening to a story about a character named Emily. Emily hid her apple in one of two containers and then left; in her absence, the apple was moved to the other container. In the test trial, one picture showed Emily searching for her apple where she had hidden it, and one picture showed Emily searching for her apple where it was moved in her absence. Children looked reliably longer at the original- than at the current-location picture, suggesting that, despite the linguistic demands of the task, they were able to keep track of Emily's false belief.

In Experiment 2, children were tested in a *verbal violation-of-expectation task*: they watched a live test event involving three adult female assistants who played the roles of Sally, the Experimenter, and the Subject. Sally stored a toy in one of two containers and then left the room; in her absence, the Experimenter moved the toy to the other container and then asked the Subject where Sally would search for her toy when she returned. Children expected the Subject to point to the original- as opposed to the current-location container, and they looked reliably longer when this expectation was violated.

The results of the present experiments are generally consistent with two sets of prior findings. First, our results support the positive findings (described in the Introduction) obtained by Garnham and her colleagues with 2- to 4-year-olds using verbal anticipatory-looking tasks (Garnham & Ruffman, 2001; Garnham & Perner, 2001). By the same token, the present results suggest that the negative findings obtained by Clements and Perner (1994) and Clements *et al.* (2000) most likely stemmed from procedural details that created inadvertent but substantial difficulties for toddlers. Future experiments can hopefully shed light on the differential demands of various anticipatory-looking tasks and on the challenges they can pose for young children.

Second, the present results extend the positive findings of Southgate *et al.* (2007) and He *et al.* (2011) by showing that toddlers in the third year of life can demonstrate false-belief understanding in verbal as well as non-verbal spontaneous-response tasks, supporting claims of robust continuity in early false-belief understanding. Of course, it might be objected that discontinuities could still emerge if children younger than those in the present research (say older infants between 18 and 24 months of age) were tested with verbal tasks. Our intuition is that this is unlikely to be true. A number of false-belief tasks with very simple language have obtained positive results with older infants (e.g. Buttelmann, Carpenter & Tomasello, 2009; Scott *et al.*, 2010; Song, Onishi, Baillargeon & Fisher, 2008; Southgate, Chevallier & Csibra, 2010); these findings, together with the present results, suggest that, as long as infants and toddlers can comprehend the language spoken in the task, they can demonstrate false-belief understanding – provided they are not asked a direct question about the agent's false belief. We take up this last point in the next section.

Why the discrepancy between elicited- and verbal spontaneous-response false-belief tasks?

The present experiments support the distinction between elicited- and spontaneous-response tasks: they indicate that toddlers succeed at verbal false-belief tasks that do not require them to answer a test question about a mistaken agent's likely actions. Indeed, toddlers give evidence of false-belief understanding when the test question is addressed not to them but to someone else, as shown in Experiment 2. Why are elicited-response false-belief tasks so challenging? If children can succeed at various verbal spontaneous-response tasks by age 2.5, why do they fail at elicited-response tasks until about age 4? There are at least two broad possibilities.

One possibility is that the early false-belief understanding revealed in verbal spontaneous-response tasks is in critical respects more rudimentary than that revealed by elicited-response tasks (e.g. Apperly & Butterfill, 2009; Low, 2010; Perner, 2010; Ruffman *et al.*, 2001). Within this perspective, one influential suggestion has been that, whereas success in verbal spontaneous-response tasks can be achieved with only an *implicit* form of false-belief understanding, success in elicited-response tasks requires an *explicit* form of false-belief understanding (e.g. Clements & Perner, 1994; Garnham & Ruffman, 2001; Perner, 2010; Ruffman *et al.*, 2001). In line with this suggestion, Low (2010) recently argued for 'separate implicit and explicit representational systems' (p. 611). He administered a battery of tasks to 3- and 4-year-olds and found that (1) performance in a verbal spontaneous-response anticipatory-looking false-belief task was related to performance in various elicited-response false-belief tasks; (2) performance in a language (complementation) task and an executive-control (cognitive-flexibility) task was related to performance in the elicited- but not the spontaneous-response false-belief tasks; and (3) performance in the language and executive-control tasks was correlated, giving rise to the possibility that the two tasks in fact measured overlapping competences. Low concluded that, with the development of language and executive control, infants' and toddlers' implicit false-belief understanding gradually metamorphoses, through a process of representational redescription (e.g. Karmiloff-Smith, 1992), 'into a higher order format that supports conscious and verbally correct false-belief judgments' (p. 612).

Another possibility is that verbal spontaneous-response tasks and elicited-response tasks tap the *same* false-belief understanding, but elicited-response tasks

generally are more challenging because they engage additional processes (e.g. Bloom & German, 2000; German & Hehman, 2006; Roth & Leslie, 1998; Russell, Mauthner, Sharpe & Tidswell, 1991). In a recent *processing-load* account, we proposed that elicited-response tasks not only require children to represent the agent's false belief, but also involve at least two executive-function processes (e.g. Baillargeon *et al.*, 2010; Scott & Baillargeon, 2009; Scott *et al.*, 2010). One is an *inhibition* process: when children are asked the test question (and thus shift from merely observing the test scene to participating in a conversation about it), their own perspective on the scene naturally becomes prominent and must be inhibited to allow them to adopt the agent's perspective (e.g. Birch & Bloom, 2003; Leslie & Polizzi, 1998; Leslie, German & Polizzi, 2005; see also Nilsen & Graham, 2009). The other process is a *response-selection* process: children must address the test question and select a response (e.g. Setoh, Scott & Baillargeon, 2011; see also Mueller, Brass, Waszak & Prinz, 2007; Saxe, Schulz & Jiang, 2006). According to the processing-load account, spontaneous-response tasks – such as those used in the present experiments – are thus easier than elicited-response tasks because (1) children observe the false-belief scene as bystanders, so that their own perspective is less salient, leaving them free to reason about the scene from the agent's perspective, and (2) children respond spontaneously, so that the response-selection process is not engaged.

Future research will need to determine which (if either) of the two possibilities described above is correct. For example, one approach might be to compare toddlers' performance in similar verbal spontaneous-response tasks and elicited-response tasks, to ascertain what critical elements lead to success or failure. Another approach might involve focusing on a key prediction from the processing-load account: if young children fail at elicited-response tasks because of high processing demands, then manipulations that reduce these demands should result in better performances. Prior research has provided some support for this prediction with 3- to 4-year-olds (e.g. Goetz, 2003; Kovács, 2009; Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Surian & Leslie, 1999; Yazdi, German, Defeyter & Siegal, 2006), and new experiments could attempt to extend these results to toddlers. If subtle manipulations aimed at reducing processing demands allow toddlers to succeed at elicited-response tasks, then it may become difficult to argue that success at these tasks crucially depends on the gradual redescription of implicit knowledge into explicit knowledge.

Concluding remarks

In the present research, 2.5-year-olds succeeded at two novel verbal spontaneous-response false-belief tasks: a preferential-looking and a violation-of-expectation task.

These results (1) support the distinction between elicited- and spontaneous-response false-belief tasks, (2) reinforce prior claims of robust continuity in early false-belief reasoning as assessed by spontaneous-response tasks, and (3) provide researchers with additional tasks for investigating early false-belief reasoning. The preferential-looking task used here, in particular, could easily be adapted for use with children from different cultures and from atypical populations. A recent report found that children with autism spectrum disorder failed at an anticipatory-looking false-belief task (Senju, Southgate, Miura, Matsui, Hasegawa, Tojo, Osanai & Csibra, 2010). Since the response demands of our preferential-looking task differ from those of anticipatory-looking tasks (children do not need to anticipate where the agent will go, but follow a story line by finding the matching picture depicting where the agent did go), determining how children with autism perform at this task could shed additional light on the nature and development of their false-belief reasoning ability.

Acknowledgements

References

Apperly, I.A., & Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**, 953–970.

Baillargeon, R., Scott, R.M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, **14**, 110–118.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, **21**, 37–46.

Birch, S.A.J., & Bloom, P. (2003). Children are cursed: an asymmetric bias in mental-state attribution. *Psychological Science*, **14**, 283–286.

Bloom, P., & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, **77**, B25–B31.

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

Clements, W.A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, **9**, 377–395.

Clements, W.A., Rustin, C., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Developmental Science*, **3**, 81–92.

Csibra, G., & Southgate, V. (2006). Evidence for infants' understanding of false beliefs should not be dismissed: response to Ruffman and Perner. *Trends in Cognitive Sciences*, **10**, 4–5.

Cummins, D.D. (2007, February). The evolution and development of social cognition. Developmental Colloquium, University of Illinois, Champaign, IL.

Dahan, D., & Tanenhaus, M.K. (2005). Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, **12**, 453–459.

de Villiers, J.G., & Pyers, J.E. (2002). Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development*, **17**, 1037–1060.

Garnham, W.A., & Perner, J. (2001). Actions really do speak louder than words – but only implicitly: young children's understanding of false belief in action. *British Journal of Developmental Psychology*, **19**, 413–432.

Garnham, W.A., & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking related to understanding of false belief? *Developmental Science*, **4**, 94–100.

German, T.P., & Hehman, J.A. (2006). Representational and executive selection resources in 'theory of mind': evidence from compromised belief-desire reasoning in old age. *Cognition*, **101**, 129–152.

Goetz, P.J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, **6**, 1–15.

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.

Hale, C.M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science*, **6**, 346–359.

Hansen, M.B. (2010). If you know something, say something: young children's problem with false beliefs. *Frontiers in Psychology*, **1** (23), 1–7.

He, Z., Bolz., M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: evidence from change-of-location and unexpected-contents violation-of-expectation tasks. *Developmental Science*, **14**, 292–305.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kovács, Á.M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, **12**, 48–54.

Kovács, Á.M., Téglás, E., & Endress, A.D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, **330**, 1830–1834.

Leslie, A.M., German, T.P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology*, **50**, 45–85.

Leslie, A.M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: two conjectures. *Developmental Science*, **1**, 247–253.

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact? *Child Development*, **61**, 1514–1519.

Lohman, H., & Tomasello, M. (2003). The role of language in the development of false-belief understanding: a training study. *Child Development*, **74**, 1130–1144.

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, **81**, 597–615.

Milligan, K., Astington, J.W., & Dack, L.A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, **77**, 622–646.

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, **39**, 107–127.

Mueller, V.A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulated zone in internally selected actions. *NeuroImage*, **37**, 1354–1361.

Nilsen, E.S., & Graham, S.A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, **58**, 220–249.

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

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P.A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science* (Vol. 1, pp. 241–261). Hove, UK: Psychology Press.

Perner, J., Leekam, S.R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: the case for a conceptual deficit. *British Journal of Developmental Psychology*, **5**, 125–137.

Peterson, C.C., Wellman, H.M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness and autism. *Child Development*, **76**, 502–517.

Roth, D., & Leslie, A.M. (1998). Solving belief problems: toward a task analysis. *Cognition*, **66**, 1–31.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, **80**, 201–224.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences*, **9**, 462–463.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, **9**, 331–349.

Saxe, R., Schulz, L.E., & Jiang, Y.V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Social Neuroscience*, **1**, 284–298.

Scott, R.M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about identity at 18 months. *Child Development*, **80**, 1172–1196.

Scott, R.M., Baillargeon, R., Song, H., & Leslie, A.M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, **61**, 366–395.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., & Csibra, G. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, **22**, 353–360.

Setoh, P., Scott, R.M., & Baillargeon, R. (2011, March). False-belief reasoning in 2.5-year-olds: evidence from an elicited-response low-inhibition task. Paper presented at the Biennial meeting of the Society for Research in Child Development, Montreal, Canada.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: a longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, **23**, 117–141.

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, **44**, 1789–1795.

Song, H., Onishi, K., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected through an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, **109**, 295–315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, **13**, 907–912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, **18**, 587–592.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science*, **18**, 580–586.

Surian, L., & Leslie, A.M. (1999). Competence and performance in false belief understanding: a comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, **17**, 141–155.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632–1634.

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: do infants understand others' beliefs? *Infancy*, **15**, 434–444.

Waxman, S.R., Lidz, J., Braun, I., & Lavin, T. (2009). Twenty-four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, **59**, 67–95.

Wellman, H.M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, **30**, 239–277.

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

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–128.

Yazdi, A.A., German, T.P., Defeyter, M.A., & Siegal, M. (2006). Competence and performance in belief–desire reasoning across two cultures: the truth, the whole truth, and nothing but the truth. *Cognition*, **100**, 343–368.

Appendix: Pictures and script used in Experiment 1
False-belief condition

Introduction-1

'This is a story about a girl named Emily.'

'There's Emily! Do you see Emily?'

(Pause 3 s)

'Emily has a friend named Sarah.'

Introduction-2

'There's Sarah! Do you see Sarah?'

(Pause 3 s)

'Emily has an apple.'

Setup-1

'See? Emily has an apple.'

(Pause 3 s)

'Emily puts her apple in a box.'

Setup-2

'Look! Emily is putting her apple in a box.'

(Pause 3 s)

'Emily is tired. She goes upstairs to take a nap.'

Setup-3

'See? There she is taking a nap.'

(Pause 3 s)

'While Emily is sleeping, Sarah takes the apple out of the box and puts it in the basket.'

Setup-4

'Look! She's putting the apple in the basket.'

(Pause 3 s)

'And then Sarah goes outside to play.'

Setup-5

'See? She's outside playing with her frisbee.'

(Pause 3 s)

'Emily wakes up and she's hungry! She looks for her apple.'
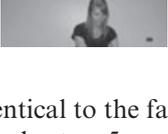
'See? Emily is looking for her apple.'
(Pause 3 s)

Knowledge condition

The knowledge condition was identical to the false-belief condition except for the hiding picture used in setup-4 and the script used in setup-3, setup-4, and setup-5.

'Emily sits down to read her book.'

Setup-3

'See? There she is reading her book.'

(Pause 3 s)

'While Emily is watching, Sarah takes the apple out of the box and puts it in the basket.'

Setup-4

'Look! She's putting the apple in the basket.'

(Pause 3 s)

'And then Sarah goes outside to play.'

Setup-5
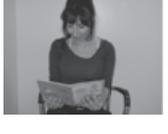
'See? She's outside playing with her frisbee.'

(Pause 3 s)

'Emily is hungry. She goes and looks for her apple.'