British Journal of Developmental Psychology (2012), 30, 14–29

The British Psychological Society

www.wileyonlinelibrary.com

2.5-year-olds succeed at a verbal anticipatory-looking false-belief task

Zijing He[1]*, Matthias Bolz[2] and Renée Baillargeon[2]

[1]Sun Yat-sen University, Guangzhou, China
[2]University of Illinois, USA

Recent research suggests that infants and toddlers succeed at a wide range of non-elicited-response false-belief tasks (i.e., tasks that do not require children to answer a direct question about a mistaken agent's likely behaviour). However, one exception to this generalization comes from verbal anticipatory-looking tasks, which have produced inconsistent findings with toddlers. One possible explanation for these findings is that toddlers succeed when they correctly interpret the prompt as a self-addressed utterance (making the task a non-elicited-response task), but fail when they mistakenly interpret the prompt as a direct question (making the task an elicited-response task). Here, 2.5-year-old toddlers were tested in a verbal anticipatory-looking task that was designed to help them interpret the anticipatory prompt as a self-addressed utterance: the experimenter looked at the ceiling, chin in hand, during and after the prompt. Children gave evidence of false-belief understanding in this task, but failed when the experimenter looked at the child during and after the prompt. These results reinforce claims of robust continuity in early false-belief reasoning and provide additional support for the distinction between non-elicited- and elicited-response false-belief tasks. Three accounts of the discrepant results obtained with these tasks – and of early false-belief understanding more generally – are discussed.

The ability to understand that others may hold and act on false beliefs is generally viewed as one of the hallmarks of human psychological reasoning. The question of when children first develop this ability has recently been the subject of intense controversy, in large part because different tasks suggest different answers to the question. In particular, consider the discrepant results that have been obtained with elicited- and non-elicited-response tasks.

In *elicited-response* tasks, children first listen to a verbal narrative about an agent who holds a false belief about some aspect of a scene. Next, children are asked a direct question about the mistaken agent's likely behaviour. Results from various elicited-response tasks suggest that the ability to attribute false beliefs to others does not emerge

*Correspondence should be addressed to Zijing He, Department of Psychology, Sun Yat-sen University, 135 Xingang West Road, Guangzhou 510275, China (e-mail: hezij2@mail.sysu.edu.cn).

DOI:10.1111/j.2044-835X.2011.02070.x

until about age 4 (e.g., Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

In *non-elicited-response* tasks, children are not asked a direct question about a mistaken agent's likely behaviour; instead, their false-belief understanding is assessed by some other means. Non-elicited-response tasks include non-verbal spontaneous-response tasks, prompted-action tasks, and verbal spontaneous-response tasks; these tasks are described next.

In *non-verbal spontaneous-response* tasks, children watch a scene involving an agent who holds a false belief. As the name implies, these tasks involve little or no language, and children's false-belief understanding is assessed through their spontaneous responses to the scene. For example, researchers monitor *how long* children look when the mistaken agent's beliefs and actions are inconsistent (violation-of-expectation tasks), or they code *where* children look as they anticipate the location the mistaken agent will approach (anticipatory-looking tasks). Results obtained with these tasks suggest that the ability to attribute false beliefs to others is present not only in toddlers in the third year of life (e.g., He, Bolz, & Baillargeon, 2011; Southgate, Senju, & Csibra, 2007), but also in infants in the first 2 years of life (e.g., Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; Träuble, Marinović, & Pauen, 2010).

Prompted-action tasks represent a hybrid between non-verbal spontaneous-response tasks and elicited-response tasks. Like non-verbal spontaneous-response tasks, prompted-action tasks typically involve simple language and present children with a scene in which an agent holds a false belief. As in elicited-response tasks, however, children are then given a verbal prompt – but one that *only indirectly* taps their representation of the agent's false belief. For example, prompts may require children to help the mistaken agent achieve his goal of retrieving a hidden toy (e.g., 'Go on, help him!'; helping tasks), or they may require children to infer which of two objects a mistaken agent intends to refer to (e.g., 'Can you get the sefo for me?'; referential-communication tasks). Positive results have been obtained in these tasks with toddlers and infants (e.g., Buttelmann, Carpenter, & Tomasello, 2009; Southgate, Chevallier, & Csibra, 2010).

Finally, *verbal spontaneous-response tasks* are similar to non-verbal spontaneous-response tasks except that they impose significant linguistic demands, often comparable to those in elicited-response tasks; as might be expected, verbal spontaneous-response tasks are typically used with toddlers and older children rather than with infants. Positive results have recently been obtained with 2.5-year-old toddlers in two different verbal spontaneous-response tasks (Scott, He, Baillargeon, & Cummins, 2011): a verbal violation-of-expectation task in which children watched an adult 'Subject' answer (correctly or incorrectly) a false-belief question, and a verbal preferential-looking task in which children listened to a false-belief story while looking at a picture book (with matching and non-matching pictures). Verbal anticipatory-looking tasks have also been used with young children, but with conflicting results: although positive results have consistently been obtained with 3-year-olds (e.g., Clements & Perner, 1994; Low, 2010; Ruffman, Garnham, Import, & Connolly, 2001), results with children under age 3 have been far more mixed, with toddlers succeeding in some reports but failing in others (e.g., Clements & Perner, 1994; Clements, Rustin, & McCallum, 2000; Garnham & Perner, 2001; Garnham & Ruffman, 2001). These mixed results, which are reviewed in the next section, provided the impetus for the present research, which tested 2.5-year-old toddlers in a novel verbal anticipatory-looking task.

Prior results with verbal anticipatory-looking tasks

In a series of experiments, Garnham (née Clements) and her colleagues tested 2- to 4-year-olds using verbal anticipatory-looking false-belief tasks (Clements *et al.*, 2000; Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001). In a typical experiment, children listened to 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 then 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 34–42 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 conditions 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 findings make it difficult to arrive at any firm conclusions about whether toddlers in the third year of life can succeed at verbal anticipatory-looking 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 such tasks, just as they succeed at other verbal spontaneous-response tasks, at non-verbal spontaneous-response tasks, and at prompted-action tasks (e.g., Buttelmann *et al.*, 2009; He *et al.*, 2011; Scott *et al.*, 2011; Southgate *et al.*, 2007). On the other hand, the fact that highly inconsistent results were obtained across experiments suggests that slight procedural variations in verbal anticipatory-looking tasks can greatly affect toddlers' performance.

Procedural variations

What procedural variations might have contributed to the discrepant results of the experiments reviewed in the last section? Although these experiments differed in several ways (e.g., different false-belief stories were used, and children's anticipatory responses were coded using different scoring systems), one difference, in particular, seemed to

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

us potentially critical: this difference had to do with the experimenter's behaviour during and after the anticipatory prompt, which could have affected whether toddlers interpreted the prompt as a self-addressed utterance or as a direct question.

In Clements and Perner (1994), the experimenter looked *at the child* during and after the prompt: 'to avoid the possibility that the child followed the experimenter's gaze, the experimenter, facing the child, took care that she was looking at the child at this critical point' (p. 383). Garnham and Perner (2001) used a similar procedure, and Clements *et al.* (2000) presumably did so as well. However, a different procedure was used by Garnham and Ruffman (2001), who obtained positive results with a sample that included toddlers. First, the experimenter did not actually utter the anticipatory prompt: children listened to an audiotaped narrative, and the experimenter simply moved the appropriate props as the story was narrated. Second, following the audiotaped prompt, the experimenter looked *away from the child*: 'a period of approximately 10 s was allowed to elapse during which the experimenter looked down so that the child could not use the experimenter's gaze as a cue to where to look' (p. 96).

These procedural differences suggested the following possibility. It could be that when the experimenter looked at the child during and after the prompt, the younger children tended to interpret it as a *direct question* (because they were less skilled at processing the relevant lexical, prosodic, or pragmatic cues, e.g., Shatz, 1978); this had the unfortunate effect of transforming the task into an elicited-response task – since the children assumed that they were being asked a direct question about the mistaken character's likely actions – resulting in the poor performance typically observed in such tasks. In contrast, when the experimenter's behaviour made it easier for the younger children to interpret the prompt as a *self-addressed utterance*, they could then reveal in their spontaneous behaviour (as in other non-elicited-response tasks) their understanding of the character's false belief.

The present research

Would a sample composed exclusively of toddlers succeed at a verbal anticipatory-looking false-belief task that was designed to help them interpret the anticipatory prompt as a self-addressed utterance? To address this question, toddlers between 29 and 34 months of age were tested using a novel verbal anticipatory-looking task: while delivering the prompt, the experimenter stared continuously at the ceiling, chin in hand, as though thinking out loud; following the prompt, she remained in the same introspective position for a 5-s pause.

We reasoned that positive findings (1) would confirm, with a sample composed exclusively of toddlers, the positive results of Garnham and Ruffman (2001); (2) would extend the positive results that have been obtained with toddlers in other non-elicited-response tasks (e.g., Buttelmann *et al.*, 2009; He *et al.*, 2011; Scott *et al.*, 2011; Southgate *et al.*, 2007); (3) would support the distinction between elicited- and non-elicited-response tasks (by providing additional evidence that toddlers succeed at verbal false-belief tasks that do not require them to answer direct questions about mistaken agents' likely actions); (4) would reinforce claims of robust continuity in early false-belief understanding as assessed by non-elicited-response tasks; and finally (5) would bear on the debate concerning the nature of the understanding infants and toddlers demonstrate in these tasks.

EXPERIMENT

In our experiment, 2.5-year-old toddlers were assigned to a false-belief condition, a knowledge condition, or a question-false-belief condition. Instead of acting out a story using dolls and props, we created a realistic scenario in which two female experimenters (E1 and E2) interacted with the child sequentially.

In the *false-belief* condition, E1 was cutting out stickers for the child to take home when she was informed that she had a phone call; she then stored her scissors in one of two identical containers (we refer to this container as the *target* container), told the child she would continue cutting out the stickers when she returned, and left. At that point, E2 entered and began interacting with the child. She found the scissors in the target container and put them in her pocket while saying, 'Oh, scissors, I was looking for a pair just like these! I need them for a project I am working on. I am going to take them!' In pilot work, we found that having E2 take possession of the scissors – as opposed to simply moving them to the other container – was a salient manipulation for our young participants (this may have been due in part to self-interest: if E2 left with the scissors, E1 could not continue cutting out stickers for them when she returned). After E2 placed the scissors in her pocket, she looked at the ceiling, chin in hand, and said, 'But when [E1's name] comes back, she is going to need her scissors again . . . ' E2 then paused for 2 s. Next, she uttered the anticipatory prompt, 'Where will she think they are?', and again paused for 5 s. Since we were uncertain whether our toddlers would understand the words 'I wonder', often used by Garnham and her colleagues in their anticipatory prompts (e.g., 'I wonder where she's going to look'; Clements & Perner, 1994; Garnham & Ruffman, 2001), we chose this simpler anticipatory prompt instead. Children in the *knowledge* condition were tested using the same procedure except that, before E1 left the test room, she saw E2 put the scissors in her pocket. Finally, children in the *question-false-belief* condition were tested using the same procedure as in the false-belief condition except that E2 looked at the child (instead of at the ceiling) during and after the anticipatory prompt. Thus, whereas for children in the false-belief and knowledge conditions the prompt was delivered as a self-addressed utterance, making the task a non-elicited-response task, for children in the question-false-belief condition the prompt was delivered as a direct question, making the task an elicited-response task.

If we were right in supposing that toddlers could demonstrate false-belief understanding in a verbal anticipatory-looking task as long as they perceived the anticipatory prompt as a self-directed utterance rather than as a direct question, then children in the false-belief condition should expect E1 to falsely believe that her scissors were still in their original location, the target container; children in the knowledge condition should expect E1 to know the scissors' current location, E2's pocket; and children in the question-false-belief condition should fail to represent E1's false belief and should mistakenly attribute to her their own knowledge about the scissors' current location (this corresponded, of course, to the error typically observed in young children tested with elicited-response false-belief tasks).

The preceding reasoning led us to predict that, following the anticipatory prompt, (1) children in the knowledge and question-false-belief conditions would respond similarly, (2) children in the false-belief condition would look reliably longer at the target container – the scissors' *original* location – than children in the knowledge and question-false-belief conditions, and (3) children in the knowledge and question-false-belief conditions would look reliably longer at E2 – the scissors' *current* location – than children in the false-belief condition. Note that these predictions differed from those in previous anticipatory-looking experiments in one crucial respect: because E2 transferred the scissors from the target container to her pocket (rather than to the non-target container), and most children spent a great deal of time looking at E2 during the session, we could not directly compare, *in each condition*, how long children looked at the scissors' original and current locations: in all three conditions, children looked longer at E2 than at the target container both during and after the prompt. Instead, we compared, *for each location*, how long children looked across conditions. As stated above, for the scissors' original location (the target container), we predicted that children in the false-belief condition would look reliably longer than those in the other two conditions; for the scissors' current location (E2), we predicted the reverse pattern.

Method

Participants

Participants were 57 English-speaking 2.5-year-olds, 28 male and 29 female (ages 29 months, 0 day to 34 months, 1 day, $M = 30$ months, 19 days). Another 14 children were tested but excluded, five because they were uncooperative (e.g., they refused to select a sheet of stickers), three because they looked away throughout the 5-s pause, two because of parental interference, two because their looking times at the target container during the 5-s pause was over three standard deviations from the mean of their condition, one because she was distracted by her sticker, and one because he covered his eyes during the 5-s pause. Nineteen children were randomly assigned to the false-belief ($M = 30$ months, 24 days), knowledge ($M = 30$ months, 11 days), and question-false-belief ($M = 30$ months, 22 days) conditions.

Participants were recruited from purchased mailing lists and locally published birth announcements; parents were offered reimbursement for their transportation expenses but were not compensated for their participation.

Materials

Each child was tested individually in a quiet lab room. The child sat on a parent's lap at a test table (73 cm high × 116.5 cm wide × 127 cm deep), in a cut-out area (34 cm × 34 cm). Parents were instructed to remain silent and neutral throughout the session.

During the session, E1 or E2 stood across from the child on the opposite side of the table. Both experimenters were female native English-speakers; they wore plain sweatshirts (one blue and one red) and pants with pockets. Under the experimenters' side of the table (out of the child's view) was a large bin containing four warm-up toys (a toy truck, a rattle, a koosh ball, and a stuffed dog), a marker, sheets of coloured paper (21.5 cm × 28 cm), and a box (13 cm × 13 cm × 9.5 cm) with assorted sheets of stickers (e.g., butterflies, stars, or cars).

Two identical containers stood 55 cm apart on the left and right sides of the table, 42 cm from the experimenters' side of the table and out of the child's reach. Each container (16 cm × 7.5 cm in diameter) had a lid (2 cm × 8.5 cm in diameter) with a knob at its centre. The containers and lids were covered with blue contact paper and decorated with yellow dots. A pair of silver scissors (14 cm × 6.5 cm × 0.75 cm at the largest points) was placed inside one of the containers.

A video camera on the experimenters' side of the table captured the child's behaviour. A mirror behind the parent made it possible to also view the experimenters' actions in the recorded sessions; these actions were checked offline for accuracy.

Procedure

The procedure included a warm-up phase, a sticker-game phase, a belief-induction phase, and a test phase.

False-belief condition

At the start of the *warm-up phase* in the false-belief condition, E1 stood across from the child; the two containers rested on the table, with the scissors hidden in one of them (target container side was counterbalanced). E1 introduced herself and then used the warm-up toys to establish rapport with the child.

During the *sticker-game phase*, E1 announced that they would now play a game in which she would give the child stickers and the child would place them on a sheet of paper to take home. As the game developed, E1 removed and replaced props from the bin at her feet. To start, E1 presented the child with several sheets of coloured paper, asked the child to choose one, wrote the child's name on the paper with the marker, and reminded the child of her own name (it was important for the test phase that the child know E1's name). Next, she presented the child with several sheets of stickers and asked the child to choose a sheet. E1 then said, 'I'm going to need my scissors to cut out these stickers for you!' E1 first opened the non-target container, showed the child that it was empty, closed it, and replaced it on the table. Next, she opened the target container where she found her scissors, exclaiming 'That's right! This is where my scissors are. This is where I keep them!' After closing the container, E1 asked the child to choose a sticker; she then cut it out and helped the child affix it to the paper. The child then chose a second sticker; as E1 was about to cut it out, she coughed mildly; this served as a signal to E2 (who was waiting outside the room) to knock at the door. E1 said, 'Come in!', and E2 opened the door and told E1 she had a phone call. E1 replied, 'No problem, I will be right there!' E2 then closed the door and waited outside. E1 put the scissors away in the target container, put the sheet of stickers back into the bin, and informed the child that she would continue cutting out the stickers when she returned. E1 then opened the door, ending the sticker-game phase.

In the *belief-induction phase*, E1 asked E2 at the door (in the child's hearing) to stay with the child while she was gone. E2 answered, 'Sure, no problem!' She then walked into the room, closed the door, and greeted the child as she took E1's place at the table. To establish rapport, E2 spent a few minutes chatting with the child (e.g., about the child's coloured paper and sticker). When E2 felt that the child was comfortable, she asked casually, 'What else are you playing with? What's in these containers?' She first opened the non-target container, showed it to the child while saying, 'Oh, there is nothing in this one!', and closed it again. Next, she reached for the target container while saying, 'What about this one? Is there something in here?' She then opened the target container, took out the scissors, and closed the lid while saying, 'Oh, scissors, I was looking for a pair just like these! I need them for a project I am working on. I am going to take them!' She then put the scissors in her pants pocket, ending the belief-induction phase.

In the *test phase*, E2 looked at the ceiling, chin in hand, and said, as though thinking out loud, 'But when [E1's name] comes back, she's going to need her scissors again . . . ' E2 paused for 2 s and then delivered the anticipatory prompt, 'Where will she think they are?' E2 then paused for 5 s, in the same position (staring at the ceiling, chin in hand). The test session then ended, and E1 returned to give the child additional stickers.

Knowledge condition

The knowledge condition was similar to the false-belief condition except for the following changes in the belief-induction phase. When E1 asked E2 to stay with the child, E2 gave E1 a cell phone. After E2 entered the room, E1 told her she would take the call in the room and closed the door. E1 then knelt to the right of the test table (in the child's view), placed the phone to her ear, and began her phone conversation ('Hello!'). E1 watched E2's actions attentively while making quiet agreement noises into the phone now and then. After E2 placed the scissors in her pocket, E1 ended her phone call, announced that she had to return the phone to the office and would be right back, and left the room. The test phase then began and proceeded exactly as in the false-belief condition.

Question-false-belief condition

The question-false-belief condition was similar to the false-belief condition except that E2 looked at the child during the test phase. After putting the scissors in her pocket, E2 leaned forward, with her hands on her knees, and looked at the child while saying, 'But when [E1's name] comes back, she's going to need her scissors again . . . ' E2 paused for 2 s and then went on to deliver the anticipatory prompt, 'Where will she think they are?' E2 then paused for 5 s, still looking at the child.[2]

Coding

For each child, we coded frame-by-frame where the child looked during two consecutive windows. First, we coded the 5-s window following the anticipatory prompt ('Where will she think they are?'); we refer to this window as the *test window*. Second, as a comparison point, we coded the 5-s window immediately preceding the test window; we refer to this window as the *baseline window*. The baseline window typically began just before or during the 2-s pause that preceded the anticipatory prompt, and it ended when E2 finished uttering the prompt. In the false-belief and knowledge conditions, E2 looked at the ceiling throughout the baseline and test window; in the question-false-belief condition, she looked at the child throughout both windows.

In each window, we coded whether the child was looking at the target container, the non-target container, E2, or away. All children were coded independently by a second coder who was naive about the experiment's hypotheses. The two coders agreed on the children's direction of gaze for 97% of coded video frames. Windows in which agreement was less than 90% (12/114 windows) were resolved through discussion.

Preliminary analyses revealed no significant interactions of condition, window, and location with sex or target container side, both Fs < 1; the data were therefore collapsed across these two factors in subsequent analyses.

---

[2] *Although E2 spoke the same words in the test phases of the false-belief and question-false-belief conditions, her intonation might have varied systematically when she spoke to herself or to the child, providing prosodic cues about whether the prompt should be interpreted as a self-addressed utterance or as a direct question. To ascertain whether E2's intonation differed reliably in the two conditions, a naive coder listened (using audio alone) to the test phases in the two conditions, randomly mixed. The coder rated E2's speech on a scale of 1 (clearly self-addressed) to 5 (clearly interrogative); one child's audio could not be coded. Ratings in the false-belief ($M = 3.28$, $SD = 1.56$) and question-false-belief ($M = 3.18$, $SD = 1.55$) conditions did not differ reliably, $t(33) = .19$, $p = .8505$, suggesting that E2 used a similar (and fairly neutral) intonation in both conditions.*

Table 1: Mean looking times (and standard deviations) at E2, the target container, and the non-target container during the baseline and test windows in the false-belief, knowledge, and question-false-belief conditions

	False-belief condition		Knowledge condition		Question-false-belief condition	
	Baseline	Test	Baseline	Test	Baseline	Test
E2	3.45 (1.44)	2.43 (1.35)	3.77 (1.33)	3.49 (1.70)	3.67 (1.50)	3.17 (1.41)
Target Container	0.22 (0.44)	0.98 (0.76)	0.37 (0.55)	0.31 (0.52)	0.32 (0.70)	0.42 (0.68)
Non-target Container	0.17 (0.56)	0.32 (0.94)	0.00 (0.00)	0.25 (0.91)	0.15 (0.15)	0.22 (0.61)

Results

Our predictions were that following the anticipatory prompt, (1) children in the knowledge and question-false-belief conditions would respond similarly, (2) children in the false-belief condition would look reliably longer at the target container than those in the knowledge and question-false-belief conditions, and (3) children in the knowledge and question-false-belief conditions would look reliably longer at E2 than those in the false-belief condition. To address the first prediction, we compared the looking times of children in the knowledge and question-false-belief conditions by means of a $2 \times 2 \times 2$ analysis of variance (ANOVA) with condition (knowledge, question-false-belief) as a between-subjects factor and window (baseline, test) and location (target container, E2) as within-subject factors (see Table 1). As predicted, the main effect of condition was not significant, or was any interaction involving this factor, all $Fs(1, 36) < 1$. The only significant result was the main effect of location, $F(1, 36) = 148.38$, $p < .0001$: overall, children looked reliably longer at E2 than at the target container.

To address the second and third predictions, we conducted a $3 \times 2 \times 2$ ANOVA similar to that above except that it also included the data from the false-belief condition. The main effect of location was again significant, indicating that children looked reliably longer overall at E2 than at the target container, $F(1, 54) = 208.80$, $p < .0001$. The ANOVA also yielded a significant window $\times$ location interaction, $F(1, 54) = 12.74$, $p = .0008$, and a significant condition $\times$ window $\times$ location interaction, $F(1, 54) = 3.70$, $p = .0312$. To examine this three-way interaction, we carried out a planned contrast, for each location, comparing the responses of children in the false-belief condition to those of children in the knowledge and question-false-belief conditions; these contrasts are described next.

Responses to the target container

As predicted, during the test window, children in the false-belief condition looked reliably longer at the target container than did children in the knowledge and question-false-belief conditions, $F(1, 54) = 5.54$, $p = .0223$; no such difference was found in the baseline window, $F(1, 54) < 1$. Examination of children's individual responses revealed similar results. In the false-belief condition, 4/19 children looked at the target container during the baseline window, and 14/19 did so in the test window (Fisher's exact test,

$p = .0029$); the corresponding numbers were 8/19 and 6/19 in the knowledge condition ($p = .7374$), and 4/19 and 6/19 in the question-false-belief condition ($p = .7140$). The false-belief condition differed reliably from the other two conditions in the test window ($p = .0044$), but not in the baseline window ($p = .5368$).

Responses to E2

As predicted, during the test window, children in the knowledge and question-false-belief conditions looked reliably longer at E2 than did children in the false-belief condition, $F(1, 54) = 12.07$, $p = .0010$; no such difference was found in the baseline window, $F(1, 54) = 1.09$, $p = .3011$. Because most children looked at E2 during both windows (only one child in the false-belief condition failed to look at E2 during the test window), comparisons of how many children in each condition looked at E2 were uninformative.

Responses to the target and non-target containers

In additional analyses, we compared children's looking times at the target and non-target containers during the test window, to confirm that children in the false-belief condition correctly identified which container E1 would search for the scissors. Because only three or four children in each condition looked at the non-target container during the test window, the data violated ANOVA assumptions. Accordingly, we computed difference scores by subtracting children's looking time at the non-target container from their looking time at the target container. Two-tailed one-sample *t*-tests against chance (0.00) revealed that children looked preferentially at the target container in the false-belief condition ($M = 0.66$, $SD = 1.20$), $t(18) = 2.40$, $p = .0274$, but not in the knowledge ($M = 0.06$, $SD = 1.11$) or the question-false-belief ($M = 0.20$, $SD = 0.79$) condition, both $ts(18) < 1$.

Overt responses

In addition to coding children's looking behaviour during the test window, coders recorded pointing gestures and verbal comments. We first examined overt responses suggesting that E1 would think the scissors were in the *target container*. In the false-belief condition, 7/19 children produced overt responses to the target container (four pointed to it and made a comment such as 'In there!', and three simply pointed to it); in the other two conditions, only 3/38 children (two in the knowledge condition and one in the question-false-belief condition) produced overt responses to the target container (all three pointed and made comments) (Fisher's exact test, $p = .0112$). We next examined overt responses suggesting that E1 would think the scissors were in *E2's possession*. In the false-belief condition, no child (0/19) had an overt response directed at E2; in the other two conditions, 11/38 children (four in the knowledge condition and seven in the question-false-belief condition) produced overt responses to E2 (five made a comment such as 'In the pocket!', two pointed, and four did both) (Fisher's exact test, $p = .0103$).

Discussion

Following the anticipatory prompt, children in the false-belief condition looked reliably longer at the target container – the scissors' original location – than did children in the knowledge and question-false-belief conditions; in contrast, children in the knowledge and question-false-belief conditions looked reliably longer at E2 – the scissors' current location – than did than children in the false-belief condition. Moreover, although relatively few children (21/57 or 37%) produced overt responses following the anticipatory prompt, these responses also suggested that children in the false-belief condition expected E1 to falsely think the scissors were still in the target container, whereas children in the other two conditions expected E1 to know the scissors were now in E2's pocket.

Together, these results suggest two main conclusions. First, the results of the false-belief condition indicate that 2.5-year-old toddlers can demonstrate false-belief understanding in a verbal anticipatory-looking task. These results (1) extend the positive findings of Garnham and Ruffman (2001) to a sample composed exclusively of toddlers under age 3, (2) support the positive findings that have been obtained with toddlers in other non-elicited-response tasks (e.g., Buttelmann *et al.*, 2009; He *et al.*, 2011; Scott *et al.*, 2011; Southgate *et al.*, 2007), and (3) provide additional evidence of robust continuity in early false-belief understanding as assessed by non-elicited-response tasks.

Second, the contrasting results of the false-belief and question-false-belief conditions support the suggestion that toddlers succeed at a verbal anticipatory-looking task when they interpret the anticipatory prompt as a self-addressed utterance (making the task a non-elicited-response task), but not when they interpret the prompt as a direct question (making the task an elicited-response task). Even though the same prompt was used in the false-belief and question-false-belief conditions, children gave evidence of false-belief understanding if E2 looked at the ceiling as she delivered the prompt, but not if she looked at the child. These results may thus help explain the discrepant findings of Clements and Perner (1994) and Garnham and Ruffman (2001): recall that the experimenter looked at the child in the former experiment, but looked away from the child in the latter.

More generally, the contrasting results of the false-belief and question-false-belief conditions reinforce the distinction between non-elicited- and elicited-response false-belief tasks. Why are elicited-response false-belief tasks so challenging for young children? Below, we discuss three different answers that have been offered to this question. The first account (our own) assumes that non-elicited- and elicited-response tasks tap the *same* form of false-belief understanding and looks elsewhere for an explanation of young children's difficulties with elicited-response tasks. The other two accounts, by Low (2010) and Perner (2010), assume that non-elicited- and elicited-response tasks tap *different* forms of false-belief understanding, albeit in very different ways.

Processing-load account

We have argued that when young children watch a mistaken agent in a non-elicited-response false-belief task, their psychological-reasoning system represents what the agent wants in the scene, what the agent knows and does not know about the scene, and what the agent falsely believes about the scene; attributing to the agent this interlocking set of motivational, epistemic, and fictional mental states allows children to correctly predict and interpret the agent's actions (e.g., Baillargeon, Scott, & He, 2010; Scott & Baillargeon, 2009).

If young children can represent false beliefs in non-elicited-response tasks, why do they fail to do so in elicited-response tasks? In a recent *processing-load* account (e.g., Scott, Baillargeon, Song, & Leslie, 2010; Scott *et al.*, 2011), we have proposed that elicited-response tasks generally overwhelm young children's information-processing resources, because they not only require false-belief representation but also involve at least two executive-function processes. One is an *inhibition* process: when children are asked the test question (and thus shift from merely observing the test scene to engaging in a verbal interaction 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., Bloom & German, 2000; Carlson & Moses, 2001; Hala, Hug, & Henderson, 2003; Roth & Leslie, 1998; Russell, Mauthner, Sharpe, & Tidswell, 1991). The other process is a *response-selection* process: children must select a response to the test question (e.g., Scott & Baillargeon, 2009; Setoh, Scott, & Baillargeon, 2011; see also Mueller, Brass, Waszak, & Prinz, 2007; Saxe, Schulz, & Jiang, 2006).

According to the processing-load account, toddlers in the present research thus failed when the anticipatory prompt was delivered as a direct question because the task then engaged the false-belief-representation, inhibition, and response-selection processes; children were overwhelmed, could not inhibit their own perspective, and mistakenly attributed to E1 their knowledge that the scissors were now in E2's pocket. Conversely, toddlers succeeded when the anticipatory prompt was delivered as a self-addressed utterance because the task then engaged only the false-belief-representation process; since children observed E2's actions as bystanders, their own perspective was less salient, leaving them free to demonstrate their understanding of E1's false belief in their spontaneous responses.

The tacit–explicit account

Like the processing-load account, Low's (2010) account assumes that infants and toddlers are capable of attributing various mental states, including false beliefs, to others. However, false-belief understanding is thought to undergo a profound representational change during early childhood, from a *tacit* to an *explicit* form of understanding.[3] Tacit understanding is sufficient for success in non-elicited-response false-belief tasks, but explicit understanding is required for success in elicited-response false-belief tasks.

In a series of experiments, Low (2010) administered a battery of tasks to 3- and 4-year-olds and obtained three key findings. First, performance in a non-elicited-response false-belief task – a verbal anticipatory-looking task modelled after that of Garnham and Ruffman (2001) – was related to performance in various elicited-response false-belief tasks. Second, performance in a language task and an executive-control task was related to performance in the elicited- but not the non-elicited-response false-belief tasks. Finally, performance in the language and executive-control tasks was correlated, giving rise to the possibility that the two tasks measured overlapping competences. Low concluded that, with the development of language and executive control, tacit false-belief understanding gradually metamorphoses 'into a higher order format that supports conscious and verbally correct false-belief judgments' (p. 612).

How might this representational change occur? Low (2010) suggested that, early in life, infants occasionally encounter situations where an agent acts in an unexpected way (e.g., searches for an object in the wrong place), and they correctly infer that the agent is acting on a false belief (they may then reveal this tacit understanding in spontaneous anticipatory glances, helping actions, and so on). However, because such

[3] *Although Low (2010) used the word 'implicit' to describe early false-belief understanding, here we use the word 'tacit' (with Low's permission) to distinguish it from the very different implicit understanding described by Perner (2010).*

inferences are relatively uncommon – agents do not routinely act on false beliefs – this early false-belief understanding remains for several years local and piecemeal. Over time, these isolated inferences continue to accumulate and are eventually integrated into a general and explicit representational format that supports 'conscious and verbally correct' responses in elicited-response false-belief tasks.

According to the tacit–explicit account, toddlers in the present research thus failed when the anticipatory prompt was delivered as a direct question because they still lacked explicit false-belief understanding; conversely, toddlers succeeded when the anticipatory prompt was delivered as a self-addressed utterance because their tacit false-belief understanding then allowed them to represent E1's false belief.

Although we agree that children eventually develop a conscious and verbally accessible folk theory of mind, just as they develop folk theories of astronomy, biology, and other domains (e.g., Carey, 1985; Leslie, 2000; Vosniadou & Brewer, 1992), it is unclear to us why a folk theory of mind would be *necessary* for success at elicited-response false-belief tasks. In his research, Low (2010) found that executive-function scores predicted children's performance in the elicited-response false-belief tasks (consistent with our processing-load account). If limited executive-function skills already explain children's inability to suppress their own perspective when asked a direct question about a mistaken agent's likely actions, why *also* assume that children are failing because they cannot consciously shift from their own perspective to that of the agent? Why posit two separate explanations when one suffices?

The implicit–explicit account

Like Low (2010), Perner (2010) assumes that false-belief understanding undergoes a profound representational change during early childhood. However, this change, from an *implicit* to an *explicit* form of understanding, is very different from that proposed by Low: according to Perner, young children are *incapable* of attributing mental states, including false beliefs, to others (see also Perner & Ruffman, 2005; Ruffman & Perner, 2005). In non-elicited-response tasks, young children use *situation-action rules* to predict agents' actions; these rules 'provide an implicit theory of mind [in that] they capture the workings of the mind without mentioning the mind' (p. 259). As they observe agents in everyday life, young children detect statistical regularities that capture how agents act in specific situations (e.g., 'agents who are searching for an object typically search for it where they saw it last'); when appropriate situational conditions are met, these regularities are used to predict agents' actions. Later in childhood, through learning processes that remain to be elucidated, children begin to explicitly understand agents' actions in terms of underlying mental states, as evidenced by their correct responses in elicited-response false-belief tasks and by their correct justifications for these responses.

According to the implicit–explicit account, toddlers in the present research thus succeeded when the anticipatory prompt was delivered as a self-addressed utterance because the situational conditions led them to retrieve the appropriate situation-action rule ('agents who are searching for an object typically search for it where they saw it last'). Conversely, toddlers failed when the anticipatory prompt was delivered as a direct question because they lacked the explicit theory of mind necessary to correctly answer the question.

Perner (2010) argues that, for any non-elicited-response task where young children successfully predict an agent's action, a *behavioural* account that invokes a situation-action rule will be simpler and more parsimonious than a *mentalistic* account that grants the ability to attribute mental states. In the behavioural account, the child only has to retrieve the situation-action rule that links the observed situational conditions to a predicted action; in the mentalistic account; however, the child must first use the situational conditions to infer the agent's mental states, and then use these mental states to predict the agent's action. On either account, the child will need to acquire 'a myriad of rules' (p. 243) to respond correctly in different situations; in each situation, however, the behavioural rule will be simpler than the mentalistic rule, because it will include no intervening mental states.

We take exception to the implicit–explicit account on several grounds but limit ourselves here to one general point. In a mentalistic account such as our processing-load account, the child does not in fact need 'a myriad of rules' for predicting actions in different situations: because children understand mental states, mentalistic rules can be stated in these terms and hence can be very general, applying across a wide range of situations. For example, consider the mentalistic rule, 'agents who are pursuing a goal will act on the information available to them, whether this information is correct or incorrect'. By establishing what *goal* an agent is pursuing, and what *information* is available to the agent, children can predict the agent's likely actions in situations where the agent is knowledgeable, ignorant, or mistaken about aspects of the scene (e.g., He *et al.*, 2011; Scott & Baillargeon, 2009; Scott *et al.*, 2010). True, a mentalistic rule will (by some measures) be more complex than a situation-action rule, but far fewer rules will be needed across situations.

As a rough analogy, consider a child who is presented with the problem '7 + 2'. A child with no understanding of number would have to use a shallow local rule such as 'when you see 7 and 2, say 9'. On the other hand, a child with some understanding of number could use a more meaningful rule such as 'when you are given an addition problem with two numbers, find the larger number, count up the other number, and then state the resulting number'. This general rule is more complex than the local rule, but it is also more powerful, as becomes immediately apparent when the child is presented with new problems such as '6 + 3', '5 + 4', and '8 + 1'. A separate local rule would have to be learned for each new problem, but the general rule would work for all the problems. Of course, children might have difficulty applying the general rule correctly as the numbers get larger (e.g., '11 + 9') and their information-processing resources become overwhelmed; but such failures would not mean that they lack any understanding of number and are limited to shallow local rules.

Conclusion

The present research indicates that 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task when the experimenter's behaviour during and after the anticipatory prompt helps them interpret it as a self-addressed utterance. These data provide additional evidence of robust continuity in early false-belief understanding and reinforce the distinction between elicited- and non-elicited-response false-belief tasks.

Acknowledgements

The research was supported by grants from NICHD (HD-021104) to Renée Baillargeon and by the 985-3 Research Program of Sun Yat-sen University (16110-3281302 and 90026-3284000). We thank Andrei Cimpian, Cindy Fisher, and Rose Scott for helpful comments and suggestions; the University of Illinois Infant Cognition Laboratory for their help with the data collection; and the parents and children who participated in the research.

References

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.

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). 18-month-old infants show false-belief understanding in an active helping paradigm. *Cognition*, *112*, 337–342.

Carey, S. (1985). *Conceptual development in childhood*. Cambridge, MA: MIT Press,

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032–1053.

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.

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 really related to understanding of belief? *Developmental Science*, *4*, 94–100.

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.

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Journal of Cognition and Development*, *4*, 275–298.

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.

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. (2000). How to acquire a 'representational theory of mind'. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 197–223). Oxford: Oxford University Press.

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

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.

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* (pp. 241–261). Hove, UK: Psychology Press.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*, 214–216.

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 implicit 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. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*, 366–395.

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2011). False-belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental Science*. Advance online publication. doi:10.1111/j.1467-7687.2011.01103.x

Setoh, P., Scott, R., & Baillargeon, R. (2011, April). *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.

Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive Psychology*, *10*, 271–301.

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 by 13-month-old infants. *Psychological Science*, *18*, 580–586.

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

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

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.

Received 26 May 2011; revised version received 7 October 2011