Title

2.5-year-olds succeed in identity and location elicited-response false-belief tasks with adequate response practice

Authors

Scott, Rose M
Roby, Erin
Setoh, Peipei

Publication Date

2020-10-01

DOI

10.1016/j.jecp.2020.104890

2.5-year-olds succeed in identity and location elicited-response false-belief tasks with adequate response practice

Rose M. Scott[a,*], Erin Roby[b], & Peipei Setoh[c]

Published in Journal of Experimental Child Psychology. Authors final submitted version; text might differ from final published version.

Word count: 6788

[a]Psychological Sciences, University of California, Merced, 5200 N. Lake Road, Merced, CA 95343, United States of America

[b] Department of Pediatrics, NYU School of Medicine, 550 First Ave., New York, NY 10016, United States of America

[c] Nanyang Technological University, 48 Nanyang Avenue, HSS-04-12, Singapore 639818.

*Corresponding author: rscott@ucmerced.edu

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by grants from the National Science Foundation (1844416), the University of California Merced Committee on Research, and the Hellman Fellows Fund. We thank Jacqueline Pagobo and the staff of the University of California Merced Center for Early Cognition and Language for their assistance with data collection and the parents and children who participated in the research.

Abstract

Researchers have argued that traditional elicited-response false-belief tasks involve considerable processing demands and hence underestimate children's false-belief understanding. Consistent with this claim, Setoh et al. (2016) recently found that when processing demands were sufficiently reduced, children could succeed in an elicited-response task as early as 2.5 years of age. Here we examined whether 2.5-year-olds could also succeed in a low-demand elicited-response task involving false beliefs about identity, which have been argued to provide a critical test of whether children truly represent beliefs, while also clarifying how the practice trials in Setoh et al.'s task facilitated children's elicited-response performance. 2.5-year-olds were tested in a version of Setoh et al.'s elicited-response task in which they heard a location or identity false-belief story. We varied whether the practice trials had the same type of wh-question as the test trial. Children who heard the same type of wh-question on all trials succeeded regardless of which story they heard (location or identity) and performance did not differ across belief type. This replicates Setoh et al.'s positive results and demonstrates that when processing demands are sufficiently reduced, children can succeed in elicited-response tasks involving false beliefs about object location or identity. This suggests that children are capable of attributing genuine false beliefs prior to age 4. However, children performed at chance if the practice trials involved a different type of wh-question than the test trials, suggesting that at this age, practice with the wh-question used in the test trial is essential to children's success.

Keywords: social cognition; theory of mind; psychological reasoning; false-belief understanding; task demands

Adults frequently interpret agents' behavior by considering their underlying mental states. An essential component of this psychological reasoning ability is the recognition that mental states are representations, rather than direct reflections, of reality and hence can be false. Traditionally, the development of this understanding has been investigated with elicited-response false-belief tasks, which require children to answer direct questions about the likely behavior of a mistaken agent (e.g., Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Wimmer & Perner, 1983). In one such task (Baron-Cohen et al., 1985), children hear a story about Sally, who places a marble in one of two containers; in her absence, Anne moves the marble to the other container. Children are then asked where Sally will look for her marble. Beginning around age 4, children correctly indicate that Sally will look in the marble's original location. Younger children incorrectly respond with the marble's true location, suggesting a failure to appreciate Sally's false belief. This widely replicated finding led many to conclude that false-belief understanding did not emerge until at least age 4 (e.g., Wellman, Cross, & Watson, 2001).

However, researchers have long argued that traditional elicited-response tasks underestimate children's false-belief understanding because they impose considerable demands on children's linguistic, pragmatic, inhibitory, attention, and memory skills (e.g., Bloom & German, 2000; Chandler, Fritz, & Hala, 1989; Hansen, 2010; Helming, Strickland, & Jacob, 2016; Leslie & Polizzi, 1998; Rubio-Fernandez & Geurts, 2013; Scott & Baillargeon, 2017; Siegal & Beattie, 1991; Westra & Carruthers, 2017). Thus, children might successfully represent an agent's false belief and nevertheless fail an elicited-response task because they cannot cope with these additional processing demands (Scott, Roby, & Smith, 2017). Consistent with this claim, several studies have identified modifications to elicited-response tasks that enable slightly younger, 3.5-year-old children to succeed (Bartsch, 1996; Bialecka-Pikul, Kosno, Bialek, &

Szpak, 2019; Chandler et al., 1989; Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Psouni et al., 2019; Roth & Leslie, 1998; Rubio-Fernandez & Geurts, 2013; Salter & Breheny, 2019). For instance, 3.5-year-olds respond correctly at above chance levels if asked where Sally will look *first* for the marble, which clarifies the experimenter's intention and increases the salience of the marble's original location (e.g., Siegal & Beattie, 1991; Yazdi, German, Defeyter, & Siegal, 2006). Children also perform better in low-inhibition versions of elicited-response tasks in which the target object is not present in the scene (Bartsch, 1996; Kikuno, Mitchell, & Ziegler, 2007; Wimmer & Perner, 1983), reducing the salience of the object's true location and hence making it easier to inhibit any prepotent tendency to respond based on reality (Scott & Baillargeon, 2017).

Recently, Setoh, Scott, and Baillargeon (2016) investigated whether further reducing processing demands enabled much younger, 2.5-year-old children to succeed at elicited-response tasks. Children heard a false-belief story accompanied by a picture book. In the story, Emma placed an apple in one of two containers and then left. In her absence, her brother Ethan found the apple and took it away. On the final page of the story, children saw the two containers and were asked the standard test question, "Where will Emma look for her apple?" The task included two modifications designed to reduce processing demands. First, Ethan took the apple away to an undisclosed location, thereby reducing inhibitory demands. Second, in order to reduce the response-generation demands evoked by the test question (i.e. children must interpret the test question and generate an appropriate response), the story included two practice trials in which children heard a "where" question that required them to point to one of two pictures. These trials thus provided practice with the type of question and response involved in the test trial. With these two modifications, children performed reliably above chance, correctly pointing to the

location where Emma falsely believed her apple was located (for a replication in German, see Grosso, Schuwerk, Kaltefleiter, & Sodian, 2019). Additional experiments showed that both task modifications were critical: 2.5-year-olds failed if they received fewer practice trials (0 or 1) or if the task was high-inhibition. These results demonstrate that traditional elicited-response tasks impose substantial demands on children. When these demands are sufficiently reduced, children can succeed in elicited-response tasks as early as 2.5 years of age. Together with recent evidence that infants and toddlers succeed in a range of non-elicited-response false-belief tasks (Scott, Roby, & Baillargeon, in press), these findings suggest that false-belief understanding emerges well before age 4.

The present study sought to address two issues raised by Setoh et al.'s findings. First, can 2.5-year-olds succeed in low-demand elicited-response tasks involving false beliefs about object identity? Older children succeed in a variety of elicited-response tasks involving different types of false beliefs, including false beliefs about the location (Baron-Cohen et al., 1985), identity (Gopnik & Astington, 1988), and contents (Hogrefe, Wimmer, & Perner, 1986) of objects. Here we sought to clarify whether 2.5-year-olds' elicited-response performance is similarly flexible and robust. We chose to focus on false beliefs about identity because of their relevance to theoretical accounts of the development of false-belief understanding. Specifically, several researchers have proposed that humans possess two psychological reasoning systems (e.g., Butterfill & Apperly, 2013; Low, Apperly, Butterfill, & Rakoczy, 2016). The late-developing system that emerges around age 4 is capable of representing false beliefs and hence enables success in a range of elicited-response false-belief tasks. In contrast, the early-developing system that is present prior to age 4 tracks belief-like states called registrations. This system enables young children to succeed in tasks involving false beliefs about object location by tracking where

the agent last registered the object. However, because registrations cannot capture *how* an agent represents an object, this system is not sufficient for success in tasks involving false beliefs about object identity.

Advocates of such accounts might argue that 2.5-year-olds succeeded in Setoh et al.'s (2016) low-demand task by tracking registrations rather than representing the agent's belief. If that were the case, then 2.5-year-olds should fail elicited-response tasks involving false beliefs about object identity. Recent evidence would appear to support this prediction: although 4-year-olds succeed in elicited-response tasks involving false beliefs about identity (Oktay-Gür, Schulz, & Rakoczy, 2018; Rakoczy, Bergfeld, Schwarz, & Fiske, 2015), 3-year-olds do not (Oktay-Gür & Rakoczy, 2017). However, the tasks used in these studies involved greater processing demands than Setoh et al.'s (2016) task. In particular, they were all high-inhibition tasks because the target object remained in the scene. It is therefore possible that younger children failed due to difficulties coping with these processing demands rather than an inability to represent false beliefs about object identity. Examining 2.5-year-olds in a low-demand task involving a false belief about object identity therefore provides a critical test of whether their elicited-response performance truly reflects a capacity to represent beliefs.

Second, why do practice trials facilitate 2.5-year-olds' performance in a low-inhibition false-belief task? Setoh et al. (2016) found that children performed at chance if they received fewer than two practice trials, suggesting these trials were critical to children's success. However, these trials were designed to reduce the demands of the response-generation process in several ways: (1) they clarified the pragmatics of the testing situation by making children aware that they would be asked questions that required them to demonstrate their knowledge, (2) they provided practice interpreting the specific type of wh-question used in the test trial (i.e. "where"

questions), and (3) they gave children practice with the required response (i.e. pointing to one of two pictures). Any of these factors, individually or in combination, could have aided children's performance.

Setoh et al. (2016) began to tease apart these factors by examining whether children needed practice with the form of the test response. Two groups of children, 30-month-olds and 33-month-olds, were tested in a version of the task in which the practice trials had only one picture. These trials provided practice with the pragmatics of the task and the type of wh-question used in the test trial, but not the specific response required in the test trial, selecting between two pictures. Results revealed that 33-month-olds succeeded, but 30-month-olds performed at chance. This suggests that at younger ages, children need practice with the form of the test response in order to succeed in a low-inhibition elicited-response task. By 33 months, this type of practice is no longer essential, and thus some other aspect of these trials must be facilitating 33-month-olds' performance.

One possibility is that 33-month-olds simply needed to be familiarized with the pragmatics of the testing situation by being asked and answering questions (Hansen, 2010; Helming et al., 2016; Scott, 2016). If so, then answering any two questions prior to the test question should enable 33-month-olds to succeed in a low-inhibition elicited-response task. Alternatively, the practice trials might have helped children cope with the linguistic demands imposed by the test question. Research with both children and adults suggests that hearing or producing a particular sentence structure facilitates subsequent processing of that sentence structure (e.g., Bock 1986; Branigan & McLean, 2016; Rissman, Legendre, & Landau, 2013; Thothathiri & Snedeker, 2008; Tooley, Swaab, Boudewyn, Zirnstein, & Traxler, 2014). For instance, hearing the sentence "Give the ball to the lion" facilitates 3-year-olds subsequent

interpretation of a sentence with the same structure (i.e. “Give the birdhouse to the sheep.”) but not a sentence with a different structure (i.e. “Give the sheep the birdhouse.”; Thothathiri & Snedeker, 2008). Similarly, in Setoh et al.’s (2016) task, hearing two “where” questions in the practice trials might have facilitated children’s processing of the “where” test question, thereby reducing the linguistic demands (and overall processing demands) in the test trial. This linguistic facilitation would only occur if children received practice with the same type of wh-question used in the test trial.

There are several reasons to suspect that reducing the linguistic demands of the test trial might be critical to younger children’s success in elicited-response tasks. First, children’s performance on traditional elicited-response tasks is positively correlated with their verbal ability (e.g., Milligan, Astington, & Dack, 2007), suggesting children only succeed if they can cope with the task’s linguistic demands (see also Scott & Roby, 2015). The wh-question used in the test trial might be especially challenging: although some ability to comprehend wh-questions emerges by age two (Seidl, Hollich, & Jusczyk, 2003), the ability to interpret and respond appropriately to wh-questions is challenging for young children and continues to improve into the early school years (e.g., Ryder & Leinonen, 2003; Tyack & Ingram, 1977). Second, elicited-response tasks often include control questions, and in some cases these occur before the critical test question, yet there is no evidence that these questions facilitated children’s performance in prior studies (e.g., Bartsch, 1996; Clements & Perner, 1994; Gopnik & Astingon, 1988). For instance, Bartsch (1996) tested 3.5-year-old children in both high- and low-inhibition versions of a change-of-location task and all children answered two “what” questions (i.e. “What’s in the yellow/green mailbox?”) prior to the critical test question (“Where will Wendy look for the letter?”). Children failed the high-inhibition task and performed at chance in the low-inhibition

task, suggesting that the control questions did not facilitate their performance. Together, these findings raise the possibility that only practice with the specific type of wh-question used in the test trial would lower the language demands sufficiently for young children to succeed in elicited-response tasks.

The present study thus had three goals. First, we sought to replicate Setoh et al.'s finding that 2.5-year-olds can succeed in a low-demand elicited-response task. Second, we sought to extend this positive finding to a new type of false belief, object identity. Third, we sought to clarify how practice trials facilitate children's performance in elicited-response tasks by investigating whether 2.5-year-olds need practice with the specific type of wh-question used in the test trial in order to succeed.

To address these goals, 2.5-year-olds were tested in 2x2 between-subjects design using an elicited-response task adapted from Setoh et al. (2016). Half of the children heard the change-of-location false-belief story from Setoh et al., whereas the other half heard a new story in which the protagonist held a false belief about an object's identity. Within each story type, we varied whether the practice trials involved the same type of wh-question as the test trial. Children in the *consistent* condition heard the same type of wh-question on all three trials (either 3 "where" questions or 3 "which" questions). Children in the *inconsistent* condition heard one type of question in the practice trials and the other type in the test trials (2 "where" practice questions and a "which" test question or vice versa). If children simply require practice with hearing and answering questions, then any two practice questions should suffice and children in both conditions should perform above chance in the test trial. If, however, children require practice with the type of wh-question used in the test trial, then only children in the consistent condition should succeed.

Method

Participants

Ninety-six 2.5-year-olds participated (31.01 – 36.30 months, *M* = 33.11 months; 48 male; see Table 1 for demographic information). All participants were native English speakers. An additional 12 children were tested but excluded because they were inattentive (2), failed to complete the experiment (1), failed to respond in the test trial (6), answered both practice trials incorrectly (1), or due to parental interference (2). Equal numbers of males and females were randomly assigned to each combination of belief-type (location, identity) and condition (consistent, inconsistent).

Table 1

Demographic characteristics of the final sample

Race	*N*
White	71
Asian	3
Hawaiian/Pacific Islander	2
African American	3
Other race	8
More than one race	4
NA	5
Ethnicity	
Hispanic/Latinx	35
Non-hispanic/Latinx	55
NA	6
Highest level of education reported by either parent	
High school or less	38

Associate's degree	8
BA/BS	23
MA/MS	13
MD/PhD	11
NA	3

The children's names were obtained from birth records provided by the California Department of Public Health, as well as from a database of parents who had previously expressed interest in developmental research. Parents were reimbursed for transportation expenses and their child received a small gift. Parents gave written informed consent for their child's participation. The Institutional Review Board at [redacted for blind review] approved all procedures.

Materials

Children heard one of two false-belief stories (location, identity) accompanied by a picture book. Each story had six story trials, two practice trials, and one test trial. On story trials, children saw a single picture and heard a line of the story. On practice and test trials, children saw two pictures and were asked a question.

The *location* false-belief story was identical to that used by Setoh et al. (2016) (see Appendix A). The *identity* story mirrored the structure of the location story (see Appendix B): the story began by introducing Ava (story-1) who found a ball in one of two containers (story-2). Ava put the ball in the other container and then left (story-3). In her absence, her friend Lily found the ball and discovered that it transformed into a bunny (story-4; the ball had a hidden button that, when pressed, caused the ball to open into bunny). Ava returned and saw Lily leave with a bunny (story-5). Ava then wanted to play with her ball (story-6). In the final test trial, children saw the two containers and were asked the test question ("Where/which place will Ava

look for her ball?"). Ava was unaware that the ball could transform into a bunny, and thus when she encountered it in its bunny guise (story-5) she should falsely believe it to be a distinct object (i.e. much like one might believe Superman to be a distinct individual from Clark Kent). This false belief about the object's identity should lead her to falsely believe that her ball was still where she left it. If children understood this set of false beliefs, they should point to the container where Ava falsely believed her ball was located (henceforth the false-belief container).

In both stories, children received practice trials after the second and fourth story trial. In these trials, children were shown two pictures of objects and were asked a question that required them to point to one of the pictures. Children were randomly assigned to the *consistent* or the *inconsistent* condition. In the *consistent* condition, the practice trials used the same type of wh-question as the test trial. In the *inconsistent* condition, the practice trials used a different type of wh-question from the test trial. The type of wh-question was counterbalanced within condition and belief-type. In the consistent condition, half the children in each story heard three "where" questions (e.g., "Where is Emma's apple?" "Where is Emma's ball?" "Where will Emma look for her apple?") and half heard three "which" questions (e.g., "Which one is Emma's apple?" "Which one is Emma's ball?" "Which place will Emma look for her apple?"). In the inconsistent condition, half the children in each story heard two "where" practice questions followed by a "which" test question, and half heard two "which" practice questions followed by a "where" test question. This ensured that any difference in performance between the two conditions was due to the consistency of the wh-questions across trials rather than the particular wh-question used in the test trial.

Apparatus and Procedure

Children sat on their parent's lap facing a table. Parents were asked to remain quiet and neutral, and to close their eyes or look down to prevent them from biasing their children's responses.

On the table sat a wooden bookstand (56 × 53 cm; inclined at a 70° angle) that held a picture book. Pages of the book were attached to the top of the bookstand with binder rings. Each page (56 × 28 cm) consisted of a clear plastic photo sheet with white paper backing; one or two color photos (20 × 25 cm) were affixed to the sheet. Single photos were centered, and double photos were 4.5 cm apart. A camera behind the bookstand captured the child's face. A second camera above and behind the child captured the stimuli and children's pointing responses. Video footage was used to verify that all participants viewed the correct pictures and heard the correct story line/question on each trial.

The experimenter stood behind the bookstand across from the child. The pages of the book began face down behind the bookstand. On each story trial, the experimenter turned a page towards the child, recited a line of the story, and then paused briefly, looking naturally between the book and the child. In each practice trial, the experimenter turned a page towards the child, asked the practice question, and then paused for up to 5 seconds. If the child responded correctly (171/192 trials), the experimenter praised the child and continued the story. If the child did not respond (8/192 trials), the experimenter asked again (all children responded by the second prompt). If children responded incorrectly (13/192 trials), the experimenter prompted the child again, ensuring the child responded correctly before proceeding. When children required multiple prompts, prompt wording was varied slightly to avoid unnatural repetition (e.g., "Do you know where Emma's apple is?"). The critical wh-word was held constant across all practice

prompts for a given participant. Averaged across practice trials, children required 1.06 prompts (*SD* = .18).

In the test trial, the experimenter turned the page towards the child, asked the test question, and paused for up to 5 seconds. If the child did not respond, the experimenter asked up to four more times, for a maximum of five prompts. On average, children required 1.44 prompts (*SD* = .95). The number of practice and test prompts required did not differ across belief-type, condition, or the wh-word used, all *F*s < .74, all *p*s > .38 (all *p*-values in this report are 2-tailed).

Throughout the practice and test trials, the experimenter looked continuously at the children to ensure that they (1) would interpret the question as a direct question (He et al., 2012) and (2) could not use the experimenter's gaze as a cue for where to point.

The container where Emma/Ava found the object (story-2), the container where she placed the object (story-3), and the side of the false-belief container in the test trial were counterbalanced within belief-type and condition. Each child received one practice trial with the correct image on the left and one with the correct image on the right. These trials were counterbalanced with the side of the false-belief container, such that equal numbers of children were tested in each of the following side orders: left-right-left, left-right-right, right-left-left, right-left-right.

Coding

For each practice and test trial, a coder verified that the experimenter used the correct wh-word and coded how many prompts the child received and where they pointed. Each test trial was coded independently by a naïve coder who did not know which was the false-belief container. Agreement on prompt number and point direction was 98% and 99%, respectively; a third naïve coder resolved disagreements.

Preliminary analyses of the test data revealed no effect of age, sex, original hiding container, side order, the wh-question used in the practice trials, or the wh-question used in the test trials, all *p*s > .15. The data were collapsed across these factors in subsequent analyses.

Results

Examination of test performance for the entire sample revealed that children performed reliably above chance: 62/96 (65%) children pointed to the false-belief container, $P = .003$ (cumulative binomial probability). However, a chi-square analysis indicated that test performance differed across conditions, $\chi^2(1, N = 96) = 4.55$, $p = .033$. Children in the consistent condition performed reliably above chance: 36/48 (75%) children pointed to the false-belief container, $P < .001$. In contrast, performance in the inconsistent condition did not differ from chance: 26/48 (54%) children pointed to the false-belief container, $P = .33$. Test performance did not differ by belief-type when the sample was analyzed as a whole, $\chi^2(1, N = 96) = .00$, $p = 1.00$, or separately by condition, consistent condition $\chi^2(1, N = 48) = .44$, $p = .51$, inconsistent condition $\chi^2(1, N = 48) = .34$, $p = .56$. In the consistent condition, performance was above chance for both belief types (location: 19/24, $P = .003$; identity: 17/24, $P = .03$). In the inconsistent condition, performance did not differ from chance for either belief type (location: 12/24, $P = .58$; identity: 14/24, $P = .27$).

These results suggest that 2.5-year-olds can succeed in low-demand elicited-response tasks involving false beliefs about object location or identity. However, their performance depends on the language used in the practice trials: they only succeed if the practice trials use the same type of wh-question as the test trial.

In an additional exploratory analysis, we examined whether children's performance varied as a function of their parent's level of education, which is frequently used as a proxy for

socioeconomic status (SES). To avoid expected values less than 5 (which renders chi-square tests suspect), we recoded parental education into three categories: a high-school education or less (N = 38), Associate's or Bachelor's degree (N = 31), or advanced degree (MA/MD/PhD; N = 24) (parent education information was unavailable for 3 participants). A chi-square analysis revealed no effect of parental education on children's test performance for the entire sample, $\chi^2(1, N = 93) = 1.32$, $p = .52$, or in either condition (both ps > .54).

General Discussion

The present study replicated Setoh et al.'s (2016) finding that 2.5-year-olds can succeed in a low-demand elicited-response task. Together with the recent replication by Grosso et al. (2019), our results suggest that this finding is robust: it has now been demonstrated in two languages across three different labs with three different samples of 2.5-year-olds. Moreover, we extended this finding to a new type of false belief: object identity. Children in the consistent condition succeeded regardless of whether they heard the location or identity story, and performance did not differ across belief type. Finally, children only succeeded if they received practice trials with the same type of wh-question as the test question. If they received one type of wh-question in the practice trials and a different type in the test trial, they performed at chance. This suggests that at this age, practice with the wh-question used in the test trial is essential to children's success.

These findings advance our understanding of the nature of early false-belief reasoning in several ways. First, our results are inconsistent with minimalist or two-system accounts, which predict that before age 4 children should fail tasks involving false beliefs about identity (Butterfill & Apperly, 2013; Low et al., 2016). Contrary to this prediction, we found that 2.5-year-olds succeeded at an elicited-response task in which an agent had a false belief about an

object's identity. Is it possible that children succeeded without truly representing a false belief about object identity? Specifically, one might argue that children in the identity condition did not understand that the ball and the bunny were the same object and hence thought that Lily left with a different toy (the bunny) rather than the ball. If children thought the ball and the bunny were distinct objects, then the story would no longer involve a false belief about object identity. We find this possibility unlikely. Such a misunderstanding would in essence transform the identity story into a true-belief task because both the children and Ava would believe that the ball was where Ava left it. This should make the identity condition much easier for children because children readily succeed in a variety of true-belief tasks prior to age 4 (e.g., Fabricius, Boyer, Weimer, & Carroll, 2010; Oktay-Gür & Rakoczy, 2017; Surian & Leslie, 1999; Wellman & Bartsch, 1988) and they do so without any need for task modifications that reduce processing demands (e.g., Oktay-Gür & Rakoczy, 2017; Surian & Leslie, 1999). We would therefore expect children who heard the identity story to perform better than those who heard the location story and to succeed in the test trial regardless of the nature of the response practice that they received. This was not the case: performance did not differ across belief type, and the nature of response practice affected performance for both the location and the identity story. This pattern of results is inconsistent with what we would expect if children thought the ball and the bunny were distinct objects. Our results are therefore more consistent with the conclusion that children represented Ava's false belief about the identity of the toy.

Our study thus provides the first evidence that 2.5-year-olds can succeed in an elicited-response task involving false beliefs about identity when processing demands are sufficiently reduced. This result adds to a growing body of evidence that infants and toddlers can succeed in non-elicited-response tasks involving false beliefs about object identity (Buttelmann & Kovács,

2019; Buttelmann, Suhrke, & Buttelmann, 2015; Forgács et al., 2019; Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015; Song & Baillargeon, 2008; for reviews, see Scott & Baillargeon, 2017; Scott, Roby, & Baillargeon, in press). Together with the present results, these findings suggest that young children are capable of attributing genuine false beliefs prior to age 4.

However, a recent study by Fizke and colleagues would appear to be at odds with this pattern of positive findings (Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017). In this study, 2.5-year-olds were tested in a non-elicited-response task modeled after the helping paradigm devised by Buttelman, Carpenter, and Tomasello (2009). Children were first introduced to a toy that had two different aspects, such as a toy carrot that could be transformed into a bunny. A naïve agent who did not know that the toy could be transformed then entered, found the toy in the form of a bunny, and placed it into a box. Next, in the agent's absence (false-belief condition) or presence (true-belief condition), an experimenter took the toy out of the box, transformed it into a carrot, and placed it back in the box. In both conditions, the agent then watched as the experimenter removed the carrot from the box and placed it on the table. The agent then attempted to open the box, failed, and expressed disappointment. The authors coded whether children attempted to help by approaching/acting on the box or acting on/pointing to the object on the table. Unlike in the original Buttelmann et al. (2009) study, children's helping responses did not differ significantly across the two conditions: the majority of children in both conditions approached/acted on the box. Based on this lack of a difference between conditions, the authors concluded that the children were incapable of representing the agent's false belief about the identity of the toy.

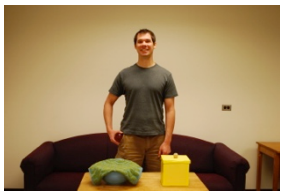
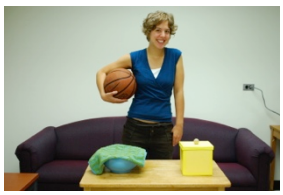
An alternative possibility, however, is that this null result stems from the nature of the required response in the false-belief condition in this task. Suppose a child successfully inferred that the agent falsely believed the bunny and the carrot were different objects and therefore falsely believed the bunny was still in the box and hence she attempted to open the box to get the bunny. How should this child go about helping the agent achieve her goal of getting the bunny? Pointing to or giving her the carrot does not seem particularly helpful because she does not want a carrot – she wants a bunny. The child would instead need (a) to explain to the agent that the carrot was in fact the bunny or (b) to transform the carrot back into the bunny for her. It is unclear whether 2.5-year-olds could easily produce either of these responses. If they could not, perhaps they defaulted to opening the box in order to show the agent that her bunny was not inside. Consistent with this possibility, when tested in a helping task that involves simpler helping responses (i.e. bringing an agent either object-A or object-B), even 18-month-olds demonstrate an understanding of false beliefs about object identity (Buttelmann et al., 2015).

Second, our study provides additional evidence for the claim that young children fail elicited-response tasks because of the substantial demands these tasks impose rather than an inability to represent beliefs. Our results also shed new light on the nature of these demands. The fact that children in the inconsistent condition performed at chance in the test trial indicates that practice with the overall pragmatics of the testing situation (i.e. being asked and answering questions) is not sufficient to facilitate successful performance at this age. Instead, our results suggest that the language demands imposed by the test question are substantial and thus 2.5-year-olds require sufficient practice with the same type of wh-question in order to cope with these demands. This result is broadly consistent with prior evidence that language ability is correlated with preschoolers' performance on elicited-response tasks (e.g., Milligan et al., 2007)

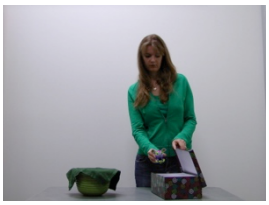
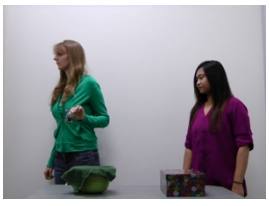
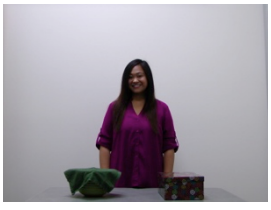
and 2.5-year-olds' performance on high-demand spontaneous-response tasks (Scott & Roby, 2015). Our results complement other recent findings indicating that children's elicited-response performance depends on factors such as attention, practice with the test response, and shared engagement with the experimenter (e.g., Psouni et al., 2019; Rubio-Fernandez & Geurts, 2013; Salter & Breheny, 2019; Setoh et al., 2016). Together, these findings begin to provide a better understanding of a widely used measure and, more generally, the skills that children need to engage in false-belief reasoning successfully in various situations.

Third, 2.5-year-olds' performance in our low-demand elicited-response task did not vary based on their parents' level of education. Although admittedly exploratory, this result is intriguing because it contrasts with the results of a recent meta-analysis (Devine & Hughes, 2018), which found a robust positive relationship between household SES and children's performance on traditional high-demand elicited-response tasks. One possibility is that children from lower SES backgrounds have greater difficulty coping with the processing demands imposed by traditional tasks, rather than difficulty representing beliefs per se. Thus, in our task, where processing demands were greatly reduced, these children performed just as well as children from higher SES backgrounds. However, given the exploratory nature of this analysis and our coarse measure of household SES, further research is needed to test this possibility. We are currently exploring these issues in ongoing research (e.g., Roby et al., 2020; Scott, Roby, & Sullivan, 2019).

Appendix A: Pictures and Script Used in the Location Story

Story-1		"This is a story about a girl named Emma. Look! There's Emma!"
Story-2		"Emma finds an apple in a bowl."
Practice-1	 	"Where/Which one is Emma's apple?"
Story-3		"Emma puts the apple in a box for later."
Story-4		"Then she goes outside to play with a ball."
Practice-2	 	"Where/Which one is Emma's ball?"
Story-5		"When Emma is gone, her brother Ethan finds the apple and takes it away."
Story-6		"Emma is hungry. She comes in to look for her apple."
Test trial	 	"Where/Which place will Emma look for her apple?"

Appendix B: Pictures and Script used in the Identity Story

Story-1		"This is a story about a girl named Ava. Look! There's Ava!"
Story-2		"Ava finds a ball in a bowl."
Practice-1	 	"Where/Which one is the ball?"
Story-3		"Ava puts the ball in a box for later. Then she goes to the bathroom"
Story-4		"While Ava is gone, her friend Lily finds the ball and look! It turns into a bunny!"
Practice-2	 	"Where/Which one is the bunny?"
Story-5		"Ava comes back and sees Lily leave with a toy."
Story-6		"Ava wants to play with the ball."
Test trial	 	"Where/Which place will Ava look for the ball?"

References

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

Bartsch, K. (1996). Between desires and beliefs: Young children's action predictions. *Child Development*, *67*, 1671-1685.

Bialecka-Pikul, M., Kosno, M., Bialek, A., & Szpak, M. (2019). Let's do it together! The role of interaction in false-belief understanding. *Journal of Experimental Child Psychology, 177,* 141-151.

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

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355-387.

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult–child dialogue. *Journal of Memory and Language*, *91*, 141-157.

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

Buttelmann, F. & Kovács, A. M. (2019). 14-month-olds anticipate others' actions based on their belief about an object's identity. *Infancy, 24*, 738-751.

Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, *131*, 94-103.

Butterfill, S., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and*

Language, 28, 606-637.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development*, *60,* 1263-1277.

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

Devine, R. T., & Hughes, C. (2018). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development*, *89*, 971–987.

Fabricius, W. V., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: Do 5-year-olds understand belief?. *Developmental Psychology*, *46*, 1402.

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind?. *Journal of Experimental Child Psychology*, *162*, 209-224.

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen‐month‐old infants track the language comprehension of communicative partners. *Developmental Science*, *22*, e12751.

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.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J., & Sodian, B. (2019). 33-month-old children succeed in a false-belief task with reduced processing demands: A replication of Setoh et al. (2016). *Infant Behavior and Development*, *54*, 151-155.

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

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-

ascription. *Mind & Language, 31*, 438-469.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57,* 567-582.

Kikuno, H., Mitchell, P., & Ziegler, F. (2007). How do young children process beliefs about beliefs?: Evidence from response latency. *Mind & Language*, *22*, 297-316.

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.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*, 184-189.

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*, *78*, 622-646.

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

Oktay-Gür, N., & Rakoczy, H. (2017). Children's difficulty with true belief tasks: Competence deficit or performance problem? *Cognition*, *166*, 28-41.

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60-74.

Psouni, E., Falck, A., Boström, L., Persson, M., Sidén, L., & Wallin, M. (2019). Together I can! Joint attention boosts 3‐to 4‐year‐olds' performance in a verbal false‐belief test. *Child Development*, *90*, 35-50.

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, *86*, 486-502.

Rissman, L., Legendre, G., & Landau, B. (2013). Abstract morphosyntax in two-and three-year-old children: Evidence from priming. *Language Learning and Development*, *9*, 278-292.

Roby, E., Cates, C. B., Canfield, C. F., Miller, E. B., Gutierrez, J., Morris, P., Shaw, D. S., & Mendelsohn, A. L. (2020, May). *Relations between parental mental-state language and early false-belief understanding among toddlers from low-income homes.* Paper presented at the Pediatric Academic Societies Meeting, Philadelphia, PA (conference cancelled).

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

Rubio-Fernández, P. & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*, 27-33.

Ryder, N., & Leinonen, E. (2003). Use of context in question answering by 3-, 4- and 5-year-old children. *Journal of Psycholinguistic Research*, *32*, 397-415.

Salter, G. & Breheny, R. (2019). Removing shared information improves 3- and 4-year-olds' performance on a change-of-location false belief task. *Journal of Experimental Child Psychology, 187,* 104665.

Scott, R.M. (2016). A question of pragmatics, or the pragmatics of questions? Invited

commentary on Helming, Strickland, & Jacob "Solving the puzzle about early belief-ascription" (*Mind & Language*, *31*, 438-469), *Brains*.

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

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*, 237-249.

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, *82*, 32-56.

Scott, R. M., & Roby, E. (2015). Processing demands impact 3-year-olds' performance in a spontaneous-response task: New evidence for the processing-load account of early false-belief understanding. *PloS One*, *10*, e0142405.

Scott, R. M., Roby, E., & Baillargeon, R. (in press). How sophisticated is infants' theory of mind? In O. Houde & G. Borst (Eds.), *The Cambridge Handbook of Cognitive Development*.

Scott, R. M., Roby, E., & Smith, M. (2017). False-belief understanding in the first years of life. In J. Kiverstein (Ed.), *Routledge Handbook of the Philosophy of the Social Mind* (pp. 152-171). New York, NY: Routledge.

Scott, R. M., Roby, E., & Sullivan, J. (2019, October). *The relationship between socioeconomic status and false-belief understanding: New evidence from a low-demand elicited-response task*. Poster presented at the biennial meeting of the Cognitive Development Society, Louisville, KY.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early understanding of subject and object wh-questions. *Infancy*, *4*, 423-436.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, *113*, 13360-13365.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*, 1-12.

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

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.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, *58*, 188-213.

Tooley, K. M., Swaab, T. Y., Boudewyn, M. A., Zirnstein, M., & Traxler, M. J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. *Language, Cognition and Neuroscience*, *29*, 289-311.

Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language, 4*, 211-224.

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.

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165-176.

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 about false belief? *Cognition, 100,* 343-368.