This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.

Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands

Setoh, Peipei; Scott, Rose M.; Baillargeon, Renée

2016

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 of the United States of America, 113(47), 13360-13365.

https://hdl.handle.net/10356/84102

https://doi.org/10.1073/pnas.1609203113

© 2016 The Author(s) (Published by National Academy of Sciences). This is the author created version of a work that has been peer reviewed and accepted for publication by Proceedings of the National Academy of Sciences of the United States of America, The Author(s) (Published by National Academy of Sciences). It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1073/pnas.1609203113].

Downloaded on 16 Feb 2025 23:01:30 SGT

Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands

Peipei Setoh[a,1,2], Rose M. Scott[b,1,2], and Renée Baillargeon[c,2]

[a]Division of Psychology, Nanyang Technological University, Singapore 637332; [b]Psychological Sciences, University of California, Merced, CA 95343; and [c]Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820

Contributed by Renée Baillargeon, October 3, 2016 (sent for review June 7, 2016; reviewed by Peter Carruthers and Alan M. Leslie)

When tested with traditional false-belief tasks, which require answering a standard question about the likely behavior of an agent with a false belief, children perform below chance until age 4 y or later. When tested without such questions, however, children give evidence of false-belief understanding much earlier. Are traditional tasks difficult because they tap a more advanced form of false-belief understanding (fundamental-change view) or because they impose greater processing demands (processing-demands view)? Evidence that young children succeed at traditional false-belief tasks when processing demands are reduced would support the latter view. In prior research, reductions in inhibitory-control demands led to improvements in young children's performance, but often only to chance (instead of below-chance) levels. Here we examined whether further reductions in processing demands might lead to success. We speculated that: (*i*) young children could respond randomly in a traditional low-inhibition task because their limited information-processing resources are overwhelmed by the total concurrent processing demands in the task; and (*ii*) these demands include those from the response-generation process activated by the standard question. This analysis suggested that 2.5-y-old toddlers might succeed at a traditional low-inhibition task if response-generation demands were also reduced via practice trials. As predicted, toddlers performed above chance following two response-generation practice trials; toddlers failed when these trials either were rendered less effective or were used in a high-inhibition task. These results support the processing-demands view: Even toddlers succeed at a traditional false-belief task when overall processing demands are reduced.

false-belief understanding | inhibitory control | information-processing resources

PNAS proof
Embargoed

Adults routinely interpret others' actions by inferring the mental states that underlie these actions, and social scientists have long been interested in understanding the development of this ability. An enduring controversy within this broad field of research centers on the attribution of false beliefs and other counterfactual mental states, because different tasks suggest very different conclusions (for reviews, see refs. 1 and 2).

In traditional false-belief tasks, children must answer a standard question about the likely behavior of an agent with a false belief (3–6). In a typical task (7), children listen to a story enacted with props: Sally hides a marble in one of two containers and then leaves; in her absence, Anne moves the marble to the other container; Sally then returns, and children are asked the standard question, "Where will Sally look for her marble?" Beginning around age 4 y, children answer correctly and point to the marble's original location; in contrast, younger children point to the marble's current location, as though they fail to understand that Sally holds a false belief about the marble's location. This developmental pattern (from below-chance to above-chance performance) has been observed in cultures around the world, although its timing varies somewhat across cultures (8, 9).

However, when tested with nontraditional tasks, which do not involve answering a standard question, toddlers (ages 2–3 y) and even infants (under age 2 y) give evidence of false-belief understanding. Nontraditional tasks can be divided into spontaneous-response and elicited-intervention tasks (1). In spontaneous-response tasks, children watch a scene in which an agent comes to hold a false belief, and their false-belief understanding is assessed via their spontaneous responses to the unfolding scene. Spontaneous-response tasks can use behavioral methods, such as the violation-of-expectation (10), preferential-looking (11), anticipatory-looking (12), and anticipatory-pointing (13) methods, or they can use neuroscientific methods, such as the electroencephalographic measurement of sensorimotor alpha-band suppression (a neural correlate of action prediction; 14) or temporal gamma-band activation (a neural correlate of sustained object representation during occlusion; 15). Spontaneous-response tasks with infants are typically nonverbal, whereas those with toddlers can be either nonverbal or verbal. Some verbal spontaneous-response tasks make linguistic demands comparable to those of traditional tasks, and some even incorporate the standard question: Instead of directing this question at the child, however, the experimenter either directs it at a third party (11) or utters it in a self-addressed manner, as though thinking out loud (16). In elicited-intervention tasks, children watch a false-belief scene and then are prompted to perform some action for the agent (e.g., "Go on, help him!"); for children to succeed, their actions must be guided by an understanding of the agent's false belief. For example, children may be prompted to help the agent retrieve an object (17), to select one of two objects for the agent (18), to open one of two doors for the returning agent (19), or to move the agent to the location she wants to search (20). Nontraditional tasks have yielded similar findings in Western and traditional non-Western cultures (21).

How can we explain the marked discrepancy between the findings of traditional and nontraditional false-belief tasks? According to the fundamental-change view, traditional tasks tap a more advanced form of false-belief understanding. In this view,

Significance

Among social scientists interested in the development of children's ability to infer mental states, an enduring controversy concerns false-belief understanding. When tested with traditional tasks, which require answering questions about the likely actions of agents with false beliefs, children do not succeed until age 4 y or later. When given nontraditional tasks without such questions, however, children succeed much earlier. Are traditional tasks more difficult because they tap an advanced form of false-belief understanding or because they impose greater processing demands? Our experiments support the latter possibility: 2.5-y-old toddlers succeeded at a traditional task when response-generation and inhibitory-control demands were both reduced. Traditional tasks thus assess the same form of false-belief understanding as nontraditional tasks but impose additional processing demands.

Author contributions: P.S., R.M.S., and R.B. designed research; P.S. and R.M.S. performed research; P.S., R.M.S., and R.B. analyzed data; and P.S., R.M.S., and R.B. wrote the paper.

Reviewers: P.C., University of Maryland; and A.M.L., Rutgers University.

The authors declare no conflict of interest.

[1]P.S. and R.M.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: psetoh@ntu.edu.sg, rscott@ucmerced.edu, or rbaillar@illinois.edu.

a major transition takes place at about age 4 y in children's false-belief understanding (3–6, 8, 9), which allows them to correctly answer standard questions such as "Where will Sally look for her marble?" According to Perner and Roessler (22), for example, answering such questions correctly "requires an intentional switch of perspectives not possible before 4 years of age." For some proponents of the fundamental-change view, the evidence from nontraditional tasks is open to alternative, low-level interpretations and reveals no genuine false-belief understanding (23, 24); for other proponents, this evidence reveals only a minimal form of false-belief understanding (25, 26). Either way, a significant shift is thought to occur around age 4 y—as a result of conceptual, executive-function, and linguistic advances—that makes possible correct responses in traditional false-belief tasks.

According to the processing-demands view, in contrast, there is substantial continuity in false-belief understanding from infancy to childhood, and early difficulties with traditional false-belief tasks are due primarily to these tasks' heavy processing demands (2, 27–29). Two separate lines of research have highlighted inhibitory-control demands, in particular. First, in their information-processing account of traditional false-belief tasks, Leslie and Polizzi (30) and Leslie et al. (31) proposed that an inhibitory process plays a key role in allowing children to express their false-belief understanding. To illustrate, consider once again the Sally–Anne task. When children are asked the standard question, "Where will Sally look for her marble?" an inappropriate prepotent response focused on the marble's actual location is triggered (exactly why this is so is widely debated; 32–37). This prepotent response must then be inhibited for children to select an alternative response consistent with their representation of Sally's false belief. Because young children's inhibitory control is immature (38), however, they cannot effectively suppress this prepotent response and thus mistakenly point to the marble's current location. Second, many correlational studies with 3- to 6-y-olds have reported a significant association between performance in traditional false-belief tasks and performance in tasks that measure conflict inhibitory control, the ability to suppress a prepotent response while activating a conflicting response (e.g., saying "day" when shown a picture of the moon and saying "night" when shown a picture of the sun; 38–42). Although this association is generally taken to indicate that inhibitory-control advances are necessary for the emergence of false-belief understanding, it is also consistent with the possibility that inhibitory-control advances contribute only to the expression of this understanding, as Leslie and Polizzi (30) and Leslie et al. (31) suggested.

A key prediction from the processing-demands view is that young children should succeed at traditional false-belief tasks when processing demands are reduced. When tested with traditional tasks in which inhibitory-control demands are lowered by various means, 3.5- to 4-y-olds often succeed, but younger children perform at chance (5, 43–48). In one low-inhibition version of the Sally–Anne task, for example, Anne takes the marble away to an undisclosed location, leaving both containers empty. Because children do not know the marble's location, the incorrect prepotent response triggered by the standard question should be weaker, and hence less inhibitory control should be needed to suppress it. As might be expected, the finding that reducing inhibitory-control demands in traditional tasks "does not increase young children's probability of passing these tasks to above-chance levels" (49) is generally perceived as an important challenge for the processing-demands view (5, 49). Here we took up this challenge: We asked whether it might be possible to raise young children's performance in a traditional low-inhibition task to above-chance levels.

We first developed an expanded processing-demands (EPD) account of children's performance in false-belief tasks, which has two main assumptions. First, reconciling findings from different false-belief tasks requires considering the full range of processing demands associated with each task. Second, children may fail to express their false-belief understanding in a task for either one of two reasons: because they lack sufficient skill at one of the processes involved (e.g., inhibitory control) or because the total amount of concurrent processing demands in the task exceeds their limited information-processing resources.

The low-inhibition Sally–Anne task described above involves at least three processes. The first is false-belief representation: As the story unfolds, children must build and maintain a representation of Sally's false belief. The second is response generation: When asked the standard question, children must interpret the question, hold it in mind, and generate a response. [We formerly referred to this as the response-selection process (29). To avoid confusion with alternative uses of the term response selection in the adult literature on executive functions, however, we now use the term response-generation instead.] The third is inhibitory control: Children must inhibit the weak incorrect prepotent response triggered by the standard question to tap their representation of Sally's false belief and generate the correct response. According to the EPD account, young children could fail at the task for at least two reasons. One is that the degree of inhibitory control required, although reduced relative to that in traditional high-inhibition tasks, is still beyond the executive skills of a substantial percentage of children; thus, some fail whereas others succeed, resulting in an overall chance performance. The other possibility is that young children generally possess sufficient inhibitory control to suppress the weak prepotent response triggered by the standard question but cannot cope with the total amount of concurrent processing demands in the task; their limited information-processing resources are overwhelmed, resulting in confused or random responding. [Attempts to explain children's failure at a cognitive task by focusing on the total amount of concurrent processing demands in the task are by no means new. A well-known case involves young infants' failure to search for hidden objects (for a review, see ref. 50). After years of debate, infancy researchers eventually agreed upon a processing-demands account of this failure: Although young infants can represent hidden objects (as shown in violation-of-expectation tasks with hidden objects) and can plan means–end actions (as shown in retrieval tasks with visible objects), they are unable to carry out both of these activities at once, due to limited information-processing resources, and they therefore fail at tasks that require performing means–end actions to retrieve hidden objects.]

This second possibility led to a prediction: Young children might succeed at a traditional low-inhibition false-belief task if the response-generation demands of the task were also reduced via practice trials. With both the response-generation and inhibitory-control demands reduced, the total concurrent processing demands of the task might no longer overwhelm young children's limited information-processing resources. As a result, children might be able to suppress the weak prepotent response triggered by the standard question, access their representation of the agent's false belief, and generate the correct response to the standard question. Exp. 1 tested this prediction: Young children received a traditional low-inhibition false-belief task that included practice trials designed to lower the response-generation demands—and hence the overall processing demands—of the test trial. Exps. 2 and 3 explored which features of the practice trials were critical for success. Finally, Exp. 4 tested another prediction from the EPD account: Young children should fail at a traditional high-inhibition false-belief task even if the response-generation demands of the task were reduced via practice trials. Due to their poor inhibitory control, children should be unable to suppress the strong prepotent response triggered by the standard question, resulting in the below-chance performance typically found in these tasks. Because 2.5-y-old toddlers have been shown to succeed at highly verbal spontaneous-response false-belief tasks (11, 16, 21), and the present tasks were also highly verbal, Exps. 1–4 focused on this age group.

Exp. 1 examined whether 2.5-y-old toddlers (n = 32) would succeed at a traditional low-inhibition false-belief task that included two response-generation practice trials. Children listened to a story accompanied by a large picture book (for other

PNAS proof
Embargoed

picture-book false-belief tasks, see refs. 11, 21, and 51). Each child sat on a parent's lap at a large table, facing the picture book; parents were asked to remain silent and close their eyes (Fig. 1*A*). The picture book rested on an inclined frame, with its pages bound at the top of the frame; the experimenter stood behind the frame, across from the child. In each of six story trials the experimenter turned a page toward the child, so that the picture on the page became visible, and then she recited a line of the story (Fig. 1*B*). The story introduced Emma (trial 1), who found an apple in one of two containers resting on a table: a bowl covered with a towel on the left and a lidded box on the right (which container held the apple was counterbalanced; trial 2). Emma then moved her apple to the other container (trial 3; this transfer served to draw children's attention to both containers). Next, Emma went outside to play with her ball (trial 4). In her absence, her brother Ethan found the apple and took it away (trial 5). Emma then returned to look for her apple (trial 6). In the test trial, children saw pictures of the bowl and box (with sides counterbalanced) and were asked, "Where will Emma look for her apple?" To succeed, children had to point to the container Emma falsely believed held her apple (henceforth the false-belief container).

To reduce the response-generation demands of the test trial, two practice trials were interspersed among the story trials. In each practice trial children were shown two pictures and were asked a "where" question, as in the test trial. The first practice trial occurred after story trial 2. Children saw an apple on the left and a banana on the right, and they were asked, "Where is Emma's apple?" The second practice trial occurred after story trial 4. Children saw a frisbee on the left and a ball on the right, and they were asked, "Where is Emma's ball?" In each trial, children were required to point to the matching picture; across trials, they pointed to both a left (apple) and a right (ball) picture.

The practice trials were not intended to support the false-belief-representation process: Pointing to the pictures of Emma's apple and ball was unlikely to help toddlers understand Emma's false belief about the apple's location. Rather, the practice trials were intended to support the response-generation process, and they included at least three features that could help reduce response-generation demands in the test trial: (*i*) they gave children practice interpreting "where" questions, (*ii*) they gave children practice choosing between two pictures and pointing to the selected one, and (*iii*) they provided children with a reliable clue as to when they would be asked a question: Because each practice trial was accompanied by two pictures (story trials had only one picture), children could know in the test trial that a question was imminent as soon as the experimenter turned the page to reveal the two pictures of the containers.

We reasoned that if the practice trials adequately reduced the response-generation demands—and hence the overall processing demands—of the test trial, then when asked where Emma would look for her apple children should reliably point to the false-belief container. Children's responses were videotaped and subsequently coded. Preliminary analyses of the test data in all four experiments revealed no effect of child's sex, false-belief container, or side of false-belief container in the test trial; the data were therefore collapsed across these factors.

Children performed reliably above chance in the test trial (Fig. 2): 25 of 32 (78%) children pointed to the false-belief container, $P = 0.001$ (cumulative binomial probability). To examine potential age effects, toddlers were divided via a median split into an older group of 33-mo-olds ($n = 16$) and a younger group of 30-mo-olds ($n = 16$). The same positive pattern was observed in the 33-mo-olds (12 of 16, $P = 0.038$) and the 30-mo-olds (13 of 16, $P = 0.011$); the two age groups did not differ reliably, $P > 0.950$ (Fisher's exact test). Thus, as predicted by the EPD account, reducing response-generation demands in a traditional low-inhibition false-belief task led toddlers to succeed at the task. After receiving two practice trials that required them to point to one of two pictures ("Where is Emma's apple?" or "Where is Emma's ball?") toddlers correctly pointed to the false-belief container in the test trial ("Where will Emma look for her apple?").

Which features of the practice trials were critical for toddlers' success? In Exp. 2, we began to explore this question by removing the two-picture feature of the practice trials in Exp. 1. Children now saw a single picture in each practice trial: the apple in the first practice trial and the ball in the second practice trial (children were required to point to this picture when asked the practice "where" question). Children thus saw a page with two pictures for the first time in the test trial. We reasoned that if the two-picture feature was critical for toddlers' success (by giving them practice at choosing between pictures and/or by allowing them to anticipate questions), then performance in Exp. 2 should revert to the chance level typically found at this age in traditional low-inhibition false-belief tasks. According to the EPD account, any changes that rendered the practice trials less effective at decreasing the response-generation demands—and hence the overall processing demands—of the test trial should adversely affect performance.

Participants were additional 2.5-y-old toddlers ($n = 32$), who were once again divided into an older group of 33-mo-olds ($n = 16$) and a younger group of 30-mo-olds ($n = 16$). Children performed at chance overall (Fig. 2): Only 19 of 32 (59%) toddlers pointed to the false-belief container in the test trial, $P = 0.189$ (cumulative binomial probability). However, different patterns emerged in the two age groups: The 33-mo-olds reliably selected the false-belief container

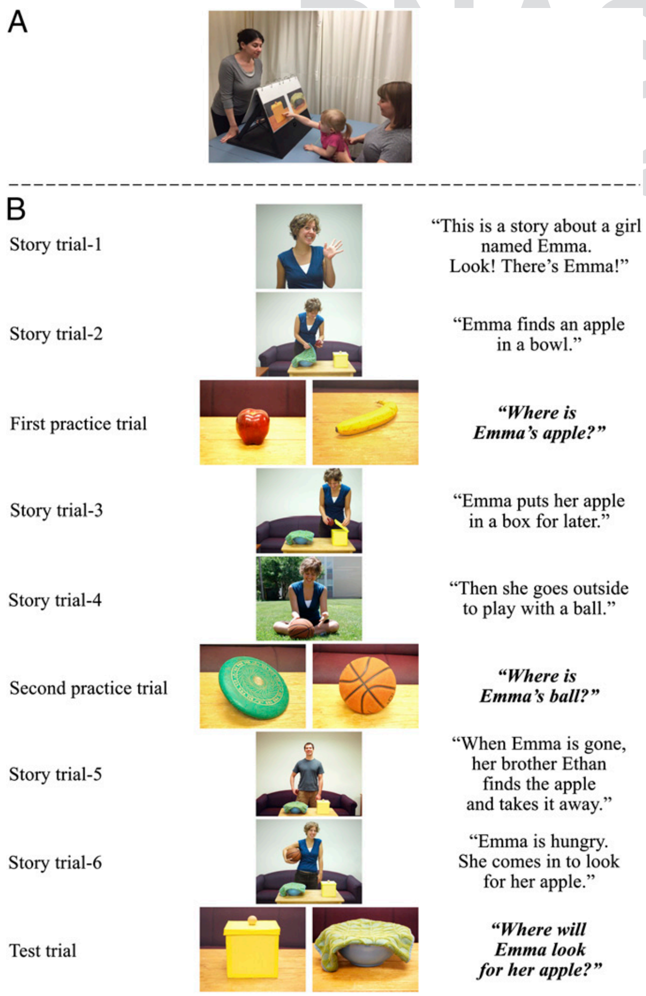

Fig. 1. Apparatus (*A*) and script and pictures used in Exp. 1 (*B*).

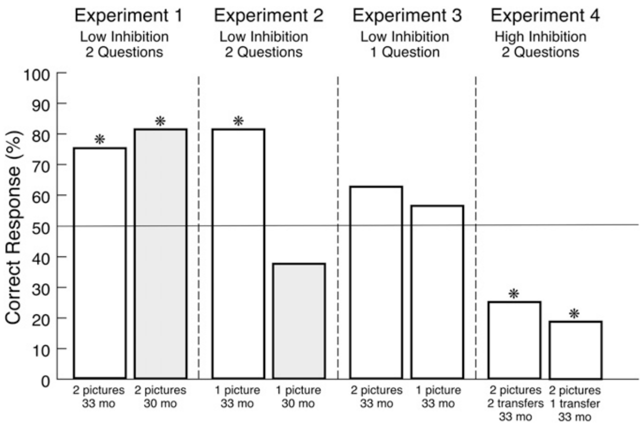

Fig. 2. Results of Exps. 1–4. An asterisk denotes that performance differed reliably from chance ($P < 0.05$ or better).

(13 of 16, $P = 0.011$), whereas the 30-mo-olds did not (6 of 16, $P = 0.895$). The responses of the two age groups differed reliably, $P = 0.029$ (Fisher's exact test).

Exp. 2 supported three conclusions. First, the positive result obtained with the 33-mo-olds confirmed that reducing response-generation demands in a traditional low-inhibition false-belief task can enable children under age 3 y to succeed at the task. Second, the negative result obtained with the 30-mo-olds replicated previous findings of chance performance in these tasks. Finally, the contrast between the results of Exps. 1 and 2 indicated that for the 30-mo-olds closer alignment of the practice and test trials was needed to adequately reduce response-generation demands. The younger toddlers succeeded when each practice trial involved two pictures (Exp. 1), like the test trial, but they failed when each practice trial involved only one picture (Exp. 2), so that they could no longer practice choosing between pictures, anticipate questions, or both.

In Exp. 3, we focused on the older toddlers and continued to explore which features of the practice trials were critical for their success. In the preceding experiments, 33-mo-olds correctly pointed to the false-belief container in the test trial after receiving two practice trials in which they were asked a "where" question that required them to point to one of two pictures (Exp. 1) or a single picture (Exp. 2). Were both of these "where" questions necessary to reduce the overall processing demands of the test trial to a manageable level? Would 33-mo-olds still succeed if asked only one "where" question? Exp. 3 examined this issue.

Participants were additional 33-mo-olds ($n = 32$). Half the children saw two pictures in each practice trial, as in Exp. 1 (two-picture condition), and half saw one picture, as in Exp. 2 (one-picture condition). In each condition, half the children heard the "where" question in the first (apple) practice trial, and half heard the "where" question in the second (ball) practice trial. On the practice trial without a "where" question, the experimenter simply said, "There is Emma's apple/ball!" We reasoned that if both of the "where" questions used in Exps. 1 and 2 were necessary to significantly reduce the response-generation demands of the test trial for 33-mo-olds, then the single "where" question used in Exp. 3 should result in the chance-level performance typically observed in traditional low-inhibition false-belief tasks.

Children performed at chance overall (Fig. 2): Only 19 of 32 (59%) children pointed to the false-belief container in the test trial, $P = 0.189$ (cumulative binomial probability). This pattern was found in the two-picture condition (10 of 16, $P = 0.227$) and in the one-picture condition (9 of 16, $P = 0.402$); the two conditions did not differ reliably, $P > 0.950$ (Fisher's exact test). In a follow-up condition, additional 2.5-y-olds ($n = 16$) were tested as in the one-picture condition of Exp. 3 except that they did not hear a "where" question in either practice trial. In line with the results of Exp. 3, only 9 of 16 (56%) children pointed to the false-belief container, $P = 0.402$ (cumulative binomial probability).

The contrast between the positive results obtained with the 33-mo-olds in Exps. 1 and 2 and the negative results obtained with the 33-mo-olds in Exp. 3 indicated that hearing two "where" questions before the test trial, one in each practice trial, was critical for success at this age. The toddlers who received only one practice question performed at chance, as did those who received no practice questions at all.

Together, the results of Exps. 1–3 indicated that: (*i*) 30- and 33-mo-olds succeeded at our task when the response-generation practice trials reduced the overall processing demands of the test trial sufficiently that these no longer exceeded their information-processing resources and (*ii*) the type of practice needed to achieve this reduction varied with age. To adults, the response-generation demands of our task no doubt seem negligible; our results make clear, however, that these demands are far from trivial for toddlers. Even at 33 mo, toddlers needed to hear two practice questions ("Where is Emma's apple?" and "Where is Emma's ball?") to correctly answer the standard question, "Where will Emma look for her apple?"

This finding may raise questions for readers familiar with the procedural details of traditional high-inhibition false-belief tasks. After all, these tasks often include control questions to ensure that children have understood key aspects of the false-belief story. In one task (48), for example, 3-y-olds performed below chance even though they received two "where" control questions ("Where did Sally put the ball in the beginning?" and "Where is the ball now?") before the standard question ("Where does Sally think the ball is?"). As discussed earlier, however, below-chance results are exactly those predicted by the EPD account for such tasks. Even though response-selection demands are reduced, the standard question still triggers a strong prepotent response focused on the toy's current location, which young children cannot suppress due to their poor inhibitory control.

To confirm this prediction from the EPD account, in Exp. 4 we tested additional 33-mo-olds ($n = 32$) in a high-inhibition version of our task. All toddlers received the same two practice trials that had effectively reduced response-generation demands in the low-inhibition version of our task in Exp. 1. The story used was identical to that in Exp. 1, with two exceptions. First, instead of taking the apple away, Emma's brother moved it from the container where he had found it to the other container, and then he left (this produced a high-inhibition task, because toddlers now knew the apple's actual location). Second, at the start of the story, we counterbalanced whether Emma found and hid the apple in the same container or in different containers. For half the children, Emma found the apple in one container and hid it in the other container, as in our previous experiments, so that the apple was transferred twice between the two containers, once by Emma and once by her brother (two-transfer condition). For the other children, Emma found and hid the apple in the same container, as in most high-inhibition tasks, so that the apple was transferred to the other container only once, by her brother (one-transfer condition).

Children performed reliably below chance overall (Fig. 2): Only 7 of 32 (22%) children pointed to the false-belief container in the test trial, $P = 0.001$ (cumulative binomial probability). This pattern was found in the two-transfer condition (4 of 16, $P = 0.038$) and in the one-transfer condition (3 of 16, $P = 0.011$); the two conditions did not differ reliably, $P > 0.950$ (Fisher's exact test). Thus, as predicted by the EPD account, when asked where Emma would look for her apple, the 33-mo-olds in Exp. 4 pointed to the container that currently held the apple, rather than to the container that Emma falsely believed held her apple. Although the task included the same two practice trials as in Exp. 1, the standard question triggered a strong prepotent response focused on the apple's actual location, and toddlers lacked sufficient inhibitory control to suppress this response.

PNAS proof
Embargoed

In the present research, we sought to address a challenge to the processing-demands view of early difficulties in traditional false-belief tasks: Reducing inhibitory-control demands only improves performance to chance levels in children age 3 y and younger (this finding was confirmed by the negative results obtained with the 30-mo-olds in Exp. 2 and the 33-mo-olds in Exp. 3). We developed an EPD account, which suggested that (*i*) young children could respond randomly in a traditional low-inhibition task because their limited information-processing resources are overwhelmed by the total concurrent processing demands in the task and (*ii*) these demands include those from the response-generation process activated by the standard question. This suggestion led to a prediction: Young children might succeed at a traditional false-belief task when both response-generation and response-inhibition demands were sufficiently reduced. Exp. 1 supported this prediction: 30- and 33-mo-olds succeeded at a traditional false-belief task in which: (*i*) response-generation demands were reduced via two practice questions that required pointing to one of two pictures and (*ii*) response-inhibition demands were reduced (as in prior research) by modifying the false-belief story so that the mistaken agent's goal object was removed to an undisclosed location. With both of these changes toddlers had sufficient information-processing resources to handle the concurrent demands of the false-belief-representation, response-generation, and inhibitory-control processes in the test trial, resulting in an above-chance performance.

Additional results supported this conclusion and the EPD account more generally. First, the positive result obtained with the 33-mo-olds in Exp. 2 confirmed that children under age 3 y could succeed at a traditional false-belief task when response-generation and inhibitory-control demands were adequately reduced. Second, the negative results obtained with the 30-mo-olds in Exp. 2 and the 33-mo-olds in Exp. 3 indicated that toddlers succeeded only when the type and amount of practice provided were sufficient, at each age, to reduce the response-generation demands of the test trial to a manageable level. Finally, the negative results of Exp. 4 confirmed that reducing the response-generation but not the inhibitory-control demands of the test trial led to below-chance responding, because children's poor inhibitory control left them unable to suppress the strong incorrect prepotent response triggered by the standard question.

The present results provide strong support for the view that early difficulties with traditional false-belief tasks stem from these tasks' processing demands. By the same token, our results also cast doubt on the view that a fundamental change in children's false-belief understanding takes place around age 4 y and makes possible success at these tasks. When processing demands are sufficiently reduced even 2.5-y-olds succeed at a traditional task, suggesting that a single psychological-reasoning system, capable of attributing counterfactual states as well as motivational and epistemic states, exists from infancy onward (1, 31, 52, 53).

Our results also make clear how developmental changes in many facets of lower- and higher-order cognition—including processing speed, working memory, inhibitory control, and language ability—could contribute to performance in false-belief tasks (38–41, 48, 54). In the present research, 33-mo-olds failed if they received fewer than two practice trials, and 30-mo-olds failed even with two practice trials if these used a different number of pictures than the test trial. In ongoing research (32), 33-mo-olds failed if the practice questions used a different question word ("Which one is Emma's apple/ball?") than the standard question ("Where will Emma look for her apple?"). False-belief understanding, although essential, is only one of the components required for success at false-belief tasks; whether children succeed or fail at any particular task (including spontaneous-response tasks; refs. 36 and 51, and see ref. 55 for similar findings with adults) will depend on the full range of processing demands in the task.

Finally, the present results suggest a possible answer to what has been described as another important challenge to the processing-demands view of early false-belief understanding (49): Groups of young children with enhanced inhibitory-control skills, such as crib bilinguals (56) and Chinese preschoolers (57), do not perform above chance in traditional high-inhibition tasks. As our research makes clear, however, inhibitory-control demands are not the only processing demands in these tasks, and reducing only these demands may not be sufficient to raise performance above chance level. Consistent with this suggestion, Duh et al. (58) recently found that in a large sample of Chinese preschoolers individual differences in working memory, but not conflict inhibition, predicted performance in traditional high-inhibition false-belief tasks.

In the present research, slight changes in the processing demands of a traditional false-belief task led 2.5-y-old toddlers to perform above chance, at chance, or below chance across experiments. These results provide strong support for the EPD account and for the claim that early failures at traditional false-belief tasks stem from limitations in young children's ability to cope with these tasks' processing demands, rather than limitations in their ability to understand false beliefs.

Methods

Participants. Participants were 144 English-speaking toddlers (72 males, range 28;17 to 36;21). Mean ages were 31;27 (Exp. 1), 31;9 (Exp. 2), 33;13 (Exp. 3), 31;13 (Exp. 3 follow-up condition), and 33;25 (Exp. 4). Children in Exps. 1 and 2 were divided via a median split into an older group (Exp. 1, mean age 33;25, range 31;13 to 36;15; Exp. 2, mean age 32;16; range 31;12 to 34;27) and a younger group (Exp. 1, mean age 29;29, range 28;17 to 31;1; Exp. 2, mean age 30;3, range 28;22 to 31;10). Another 29 toddlers were tested but excluded because the parent interfered (1 subject) or because they were distracted (3 subjects), refused to continue (1 subject), or failed to point in the practice trials (10 subjects) or test trial (14 subjects). Because our experiments examined the effects of response-generation practice on test performance, only children who responded to the practice and test questions were retained in the analyses. Written informed consent was obtained from the parent before the test session, and all protocols were approved by the University of Illinois Institutional Review Board.

Apparatus and Stimuli. The picture book was mounted on a black wooden frame (46.5-cm tall × 56-cm wide × 52.5-cm deep) inclined at a 70° angle. The nine book pages (for the six story trials, two practice trials, and one test trial) were bound at the top of the frame with six binder rings. The pages consisted of clear plastic sheet protectors (32 cm × 56 cm) containing white paper to which color photos (20 cm × 25 cm) were affixed. Single photos were centered at the bottom of the page, and double photos were placed 4.5 cm apart at the bottom of the page.

Coding. One camera (located behind a 5-cm hole in the frame, below the book) captured children's responses, an overhead camera captured both the book and children's responses, and a third camera captured the experimenter. For each practice and test question we coded where children pointed and how many prompts they received (number of prompts was coded not as a proxy for response latency but as a measure of our practice manipulation: The more prompts children received, then arguably the more practice they had in processing a question). Each test trial was coded independently by a naive coder who did not know which was the false-belief container. In each experiment, the two coders agreed on points and prompts on all trials.

Procedure. In Exp. 1, in each story trial the experimenter turned a page toward the child, recited a line of the story, and then paused for 2 s, looking naturally between the book and the child. Spontaneous pointing in the story trials was relatively rare, occurring on 11% of the trials across experiments. In each practice trial, the experimenter turned a page toward the child, asked the practice question, and then paused for 5 s. If the child responded correctly, the experimenter praised the child and went on with the story. If the child did not respond, the experimenter repeated the question, with slight variations (e.g., "Can you show me where Emma's apple is?"), for a maximum of four prompts, each with a 5-s pause. Children responded readily and received, on average, 1.13 prompts per question (SD = 0.38). In the test trial, the experimenter followed the same procedure except that children received up to five prompts, to maximize the probability of a response. If the child did not respond to the initial test question ("Where will Emma look for her apple?),

PNAS proof
Embargoed

the experimenter repeated the question with slight variations ("Can you show me where Emma will look for her apple?"). Children responded after 1.72 prompts (SD = 0.99) on average. Throughout the practice and test trials the experimenter looked continuously at the children to ensure that they: (*i*) would interpret the question as a direct question (16) and (*ii*) could not use the experimenter's gaze as a cue for where to point. In Exps. 2–4, the procedure was identical to that in Exp. 1 except as indicated in the main body of the article. In practice trials without a "where" question the experimenter said, "There is Emma's apple/ball!", paused for 5 s, and then went on with the story (the experimenter did not point to a picture in this or any other trial, as this could have provided indirect response-generation practice). In Exp. 2, on average, children responded to each practice question after 1.36 prompts (SD = 0.63) and to the test question after 1.59 prompts (SD = 1.01). The corresponding numbers of prompts for the other experiments were as follows: Exp. 3, practice question 1.31 (0.54), test question 1.41 (0.84); follow-up condition without practice questions, test question 2.06 (1.06); and Exp. 4, practice questions 1.21 (0.41), test question 1.22 (0.79).

ACKNOWLEDGMENTS. We thank the two reviewers and Francesco Antilici, Valerie Camos, and Dan Hyde for helpful suggestions or discussions, and the families who participated in the research. This study was supported National Institute of Child Health and Human Development Grant HD021104 (to R.B.).

1. Baillargeon R, Scott RM, Bian L (2016) Psychological reasoning in infancy. *Annu Rev Psychol* 67:159–186.
2. Carruthers P (2013) Mindreading in infancy. *Mind Lang* 28:141–172.
3. Gopnik A, Astington JW (1988) Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Dev* 59(1):26–37.
4. Perner J, Leekam SR, Wimmer H (1987) Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *Br J Dev Psychol* 5:125–137.
5. Wellman HM, Cross D, Watson J (2001) Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev* 72(3):655–684.
6. Wimmer H, Perner J (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1):103–128.
7. Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind"? *Cognition* 21(1):37–46.
8. Callaghan T, et al. (2005) Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychol Sci* 16(5):378–384.
9. Liu D, Wellman HM, Tardif T, Sabbagh MA (2008) Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Dev Psychol* 44(2):523–531.
10. Onishi KH, Baillargeon R (2005) Do 15-month-old infants understand false beliefs? *Science* 308(5719):255–258.
11. Scott RM, He Z, Baillargeon R, Cummins D (2012) False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Dev Sci* 15(2): 181–193.
12. Southgate V, Senju A, Csibra G (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci* 18(7):587–592.
13. Knudsen B, Liszkowski U (2012) Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Dev Sci* 15(1):113–122.
14. Southgate V, Vernetti A (2014) Belief-based action prediction in preverbal infants. *Cognition* 130(1):1–10.
15. Kampis D, Parise E, Csibra G, Kovács AM (2015) Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proc Biol Sci* 282(1819):20151683.
16. He Z, Bolz M, Baillargeon R (2012) 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *Br J Dev Psychol* 30(Pt 1):14–29.
17. Buttelmann D, Carpenter M, Tomasello M (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112(2):337–342.
18. Buttelmann D, Over H, Carpenter M, Tomasello M (2014) Eighteen-month-olds understand false beliefs in an unexpected-contents task. *J Exp Child Psychol* 119: 120–126.
19. Rhodes M, Brandone AC (2014) Three-year-olds' theories of mind in actions and words. *Front Psychol* 5:263.
20. Rubio-Fernández P, Geurts B (2013) How to pass the false-belief task before your fourth birthday. *Psychol Sci* 24(1):27–33.
21. Barrett HC, et al. (2013) Early false-belief understanding in traditional non-Western societies. *Proc Biol Sci* 280(1755):20122654.
22. Perner J, Roessler J (2012) From infants' to children's appreciation of belief. *Trends Cogn Sci* 16(10):519–525.
23. Heyes C (2014) False belief in infancy: A fresh look. *Dev Sci* 17(5):647–659.
24. Ruffman T (2014) To belief or not belief: Children's theory of mind. *Dev Rev* 34: 265–293.
25. Butterfill S, Apperly IA (2013) How to construct a minimal theory of mind. *Mind Lang* 28:606–637.
26. Low J, Watts J (2013) Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychol Sci* 24(3):305–311.
27. Bloom P, German TP (2000) Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77(1):B25–B31.
28. Roth D, Leslie AM (1998) Solving belief problems: Toward a task analysis. *Cognition* 66(1):1–31.
29. Scott RM, Baillargeon R (2009) Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Dev* 80(4):1172–1196.
30. Leslie AM, Polizzi P (1998) Inhibitory processing in the false belief task: Two conjectures. *Dev Sci* 1:247–253.
31. Leslie AM, Friedman O, German TP (2004) Core mechanisms in "theory of mind". *Trends Cogn Sci* 8(12):528–533.
32. Baillargeon R, et al. (2015) Psychological and sociomoral reasoning in infancy. *APA Handbook of Personality and Social Psychology*, eds Mikulincer M, Shaver PR, Borgida E, Bargh JA (American Psychological Association, Washington, DC), Vol 1, pp 79–150. Q:15
33. Helming KA, Strickland B, Jacob P (2014) Making sense of early false-belief understanding. *Trends Cogn Sci* 18(4):167–170.
34. Rubio-Fernández P, Geurts B (2015) Don't mention the marble! The role of attentional processes in false-belief tasks. *Rev Phil Psychol*, in press. Q:16
35. Scott RM (2016) The developmental origins of false-belief understanding. *Curr Dir Psychol Sci*, in press.
36. Wang L, Leslie AM (2016) Is implicit theory of mind the "real deal"? The own-belief/true-belief default in adults and young preschoolers. *Mind Lang* 31:147–176.
37. Westra E, Carruthers P (2016) The child as pragmatist: A new look at the theory-of-mind scale. Q:17
38. Diamond A (2013) Executive functions. *Annu Rev Psychol* 64:135–168.
39. Carlson SM, Moses LJ (2001) Individual differences in inhibitory control and children's theory of mind. *Child Dev* 72(4):1032–1053.
40. Chasiotis A, Kiessling F, Hofer J, Campos D (2006) Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica, and Cameroon. *Int J Behav Dev* 30:249–260.
41. Hala S, Hug S, Henderson A (2003) Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *J Cogn Dev* 4:275–298.
42. Müller U, Liebermann-Finestone DP, Carpendale JIM, Hammond SI, Bibok MB (2012) Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *J Exp Child Psychol* 111(2): 331–348.
43. Bartsch K (1996) Between desires and beliefs: Young children's action predictions. *Child Dev* 67(4):1671–1685.
44. Kikuno H, Mitchell P, Ziegler F (2007) How do young children process beliefs about beliefs?: Evidence from response latency. *Mind Lang* 22:297–316.
45. Mitchell P, Lacohée H (1991) Children's early understanding of false belief. *Cognition* 39(2):107–127.
46. Siegal M, Beattie K (1991) Where to look first for children's knowledge of false beliefs. *Cognition* 38(1):1–12.
47. Surian L, Leslie AM (1999) Competence and performance in false belief understanding: A comparison of autistic and normal 3-yr-old children. *Br J Dev Psychol* 17:141–155.
48. Yazdi AA, German TP, Defeyter MA, 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(2):343–368.
49. Devine RT, Hughes C (2014) Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Dev* 85(5):1777–1794.
50. Hauf P, Paulus M, Baillargeon R (2012) Infants use compression information to infer objects' weights: Examining cognition, exploration, and prospective action in a preferential-reaching task. *Child Dev* 83(6):1978–1995.
51. Scott RM, 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(11):e0142405.
52. Carruthers P (2016) Two systems for mindreading? *Rev Phil Psychol* 7:141–162.
53. Scott RM, Richman JC, Baillargeon R (2015) Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognit Psychol* 82:32–56.
54. Milligan K, Astington JW, Dack LA (2007) Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Dev* 78(2):622–646.
55. Schneider D, Lam R, Bayliss AP, Dux PE (2012) Cognitive load disrupts implicit theory-of-mind processing. *Psychol Sci* 23(8):842–847.
56. Kovács ÁM (2009) Early bilingualism enhances mechanisms of false-belief reasoning. *Dev Sci* 12(1):48–54.
57. Sabbagh MA, Xu F, Carlson SM, Moses LJ, Lee K (2006) The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychol Sci* 17(1):74–81.
58. Duh S, et al. (2016) Theory of mind and executive function in Chinese preschool children. *Dev Psychol* 52(4):582–591.

PNAS proof
Embargoed