Journal of Experimental Child Psychology 247 (2024) 106039

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Continuity in false belief understanding from 33 to 52 months of age

Beate Sodian *, Larissa J. Kaltefleiter, Tobias Schuwerk, Daniela Kloo

Department of Psychology, Ludwig Maximilian University of Munich, D-80802 Munich, Germany

ARTICLE INFO

Article history:
Received 30 January 2024
Revised 11 July 2024
Available online 17 August 2024

Keywords:
Theory of mind
False belief understanding
Low-demands false belief task
Conceptual continuity
Interrelation of false belief understanding and executive function
Toddlerhood to early childhood

ABSTRACT

Conceptual continuity in children's false belief understanding from toddlerhood to childhood was investigated in a longitudinal study of 75 children. Performance in a low-demands false belief task at 33 months of age was significantly correlated with performance in a content false belief task at 52 months independent of language ability and executive function. In contrast, there was no correlation with performance in a location false belief task, which differed from the "Sally–Anne" format of the low-demands task and was high in executive demands. These findings support the view that explicit false belief understanding may be continuous from toddlerhood to childhood and that developmental change may be characterized in terms of enrichment and increasing stability of core conceptual understanding rather than in terms of fundamental change.

© 2024 The Author(s). Published by Elsevier Inc.

Introduction

The fundamental human ability to distinguish beliefs from reality and to predict actions based on an agent's belief has been shown to develop in children around 4 years of age, when verbal tasks were used (for reviews, see Rakoczy, 2022; Wellman, 2018). Since the 1980s, a large body of research on false belief (FB) understanding in children has consistently found steep developmental progress between 3.5 and 6 years of age. When the processing demands of the tasks were reduced, 3-year-olds sometimes succeeded, but there was still marked developmental change from 2.5 to 4 years

* Corresponding author.
E-mail address: sodian@psy.lmu.de (B. Sodian).

https://doi.org/10.1016/j.jecp.2024.106039
0022-0965/© 2024 The Author(s). Published by Elsevier Inc.

(Wellman et al., 2001). These findings support the view that there is conceptual change in young children's emerging understanding of the mental domain (Bartsch & Wellman, 1995; Perner, 1991).

This standard view has been challenged over the last 20 years, based on new empirical findings on false belief understanding (FBU) in infants and toddlers. Baillargeon et al. (2010, 2016) proposed that infants and toddlers possess an understanding of belief that is conceptually continuous with later explicit verbal FBU. In the current article, we briefly review the experimental evidence for this high-level conceptual continuity view. We further argue that longitudinal data on the predictive relation of early and later FBU are an important source of evidence for theoretical accounts of the development of FBU. We focus on the claim that toddlers, before their third birthday, possess an explicit understanding of FB, which is revealed in tasks with low inhibitory control and response generation demands (Scott et al., 2020; Setoh et al., 2016) and report the first longitudinal study of FBU from 33 to 52 months of age.

Whereas traditional theory of mind (ToM) research has used predominantly verbal elicited response tasks to assess children's FBU, spontaneous response tasks (looking-time, anticipatory-looking, and prompted action paradigms) to study FBU in infancy were only developed in the last 20 years. By now, more than 30 studies have shown that infants, when tested with spontaneous response tasks, can take agents' FBs into account in forming action expectations (for reviews, see Baillargeon et al., 2010; Scott & Baillargeon, 2017). In violation-of-expectation tasks, 13- to 18-month-old infants were repeatedly shown to look significantly longer when an agent acted in a way that was inconsistent, rather than consistent, with his or her belief (e.g., Onishi & Baillargeon, 2005; Scott & Baillargeon, 2009; Song et al., 2008; Surian et al., 2007). Some looking-time studies (Kovács et al., 2010; Luo, 2011) even found evidence for FBU in the first year of life. In anticipatory-looking tasks, 18- to 25-month-old infants were shown to anticipate an agent to approach the location where he or she falsely believed a hidden object to be rather than the location where it really was (Senju et al., 2011; Southgate et al., 2007; Surian & Geraci, 2012; Thoermer et al., 2012). Finally, prompted action paradigms, in which children were typically supposed to help an experimenter find an object, yielded additional support for the view that 18-month-olds take others' FBs into account when performing a goal-directed action such as retrieving a toy from its new hiding place when the experimenter had been absent at the transfer of the toy from the old location to the new location (Buttelmann et al., 2009).

Subsequent replication attempts of all three types of infant FB tasks have not always been successful (e.g., Crivello & Poulin-Dubois, 2018; Dörrenberg et al., 2018; Powell et al., 2018; Schuwerk et al., 2018; see Kulke & Rakoczy, 2018, for an overview). In some cases, the replication failures may have been due to apparently minor, but significant, deviations from the original procedure (Baillargeon et al., 2018). In other cases, the original findings were gathered from a small sample (Southgate et al., 2007), and a replication attempt in a larger sample by the original authors showed that the task failed to reliably elicit action prediction even when no belief attribution was required (Kampis et al., 2021). A *meta*-analysis (Barone et al., 2019) of 56 FB conditions of different types of spontaneous response tasks indicated that overall correct performance of infants on spontaneous FB tasks was more likely than incorrect performance, which is consistent with the interpretation that these tasks tap a real phenomenon. However, heterogeneity was high across studies, there was evidence for effects of task type and effects of year of publication, and there was evidence for publication bias, suggesting that the phenomena under study are not yet well-understood. An international consortium (ManyBabies 2) is currently testing for the replicability and reliability of findings from spontaneous ToM tasks (Schuwerk et al., 2024).

The findings of spontaneous FB tasks have sparked a heated debate over the interpretation of ToM development in early childhood. Whereas high-level conceptual continuity views (Baillargeon et al., 2010, 2016) maintain that infants possess an understanding of belief that is conceptually continuous with later explicit verbal FBU, two-systems views assume a preconceptual form of mental state representation in infancy that is restricted by signature limits and is relatively independent of the later conceptual system of ToM (Apperly & Butterfill, 2009). In contrast, radically minimalist accounts assume that infants' performance on spontaneous FB tasks is not the result of mind-reading at all but rather of low-level perception-based submentalizing (Heyes, 2014).

Longitudinal studies of ToM development from infancy to early childhood are an important source of evidence for theoretical accounts of ToM in infancy. Whereas high-level conceptual continuity theory predicts correlations of infant performance on spontaneous FB tasks and preschoolers' explicit verbal FBU, two-systems theories and minimalist accounts do not entail such predictions. To date, one long-term longitudinal study of ToM development from infancy to 6 years of age yielded evidence for conceptual continuity in FBU (see Sodian et al., 2020, for a review). Implicit FBU was assessed with an anticipatory-looking task at 18 months of age, and various explicit measures of ToM were employed at 4, 5, and 6 years of age. Anticipatory looking at 18 months toward the location where a person would expect a target object based on his or her FB predicted explicit FBU at 4 to 6 years independent of language and executive function (Kloo et al., 2020; Thoermer et al., 2012). Similarly, implicit FBU predicted later belief-based intention understanding in a morally relevant context (Sodian et al., 2016). Moreover, there was evidence for conceptual coherence among measures of goal-encoding, FBU, and Level-1 perspective taking in infancy (Sodian et al., 2016; Thoermer et al., 2012). In contrast, two other longitudinal studies of ToM in infancy and at preschool age (Poulin-Dubois et al., 2020, 2023), using violation-of-expectation and interactive tasks as measures of infant FBU, did not find evidence for longitudinal relations. However, these authors could not rule out chance responding to FB tasks in infancy, and the studies were plagued with FB task comprehension problems at preschool age (Sodian, 2023).

High-level conceptual continuity in FBU is inconsistent with 2- and 3-year-olds' pervasive failure in elicited response FB tasks as well as the drastic age-related improvement in traditional FBU tasks at 3 to 5 years of age. It has long been argued that traditional FB tasks may underestimate 2- and 3-year-olds' FBU due to task demands. However, reductions in task demands often led only to performance at chance levels in toddlers (Wellman et al., 2001). The nature and effects of specific task demands were only recently revealed in a study of 2.5-year-old toddlers' performance in a traditional elicited response FB about location task (Setoh et al., 2016): An agent (A) placed a target in one of two locations and left the scene. Another agent (B) took it away in her absence. Participants were required to point to the location where A would look for the target upon her return. The authors analyzed task demands in terms of inhibitory control and response generation demands. Inhibitory control demands were reduced by removing the target object from the scene, a manipulation that has previously been shown to be effective (Wimmer & Perner, 1983) but not effective enough to achieve above-chance performance in very young children. Only when children were also provided with two practice trials in answering factual "where" questions that required them to point to one of two objects were responses to the test question reliably above chance, with 78% of the toddlers pointing to the container that the agent falsely believed held the target object. Neither one of these manipulations alone was sufficient to achieve above-chance performance, nor was one practice trial sufficient to do so. These findings support the view that young children's failures at elicited response FB tasks are due to their limited information processing abilities, in particular when required to inhibit a prepotent response and when asked to generate a verbal response to a "where" question without previous familiarization to the question format. In particular, the findings showed that toddlers may perform below chance, at chance, or above chance in the same task when task demands are not lowered, only partly lowered, or appropriately lowered. Above-chance performance in Setoh et al.'s (2016) task was independently replicated in 33-month-old German children by Grosso et al. (2019). Furthermore, Scott et al. (2020) replicated the findings and extended them to FB about identity.

Setoh et al.'s (2016) claim that toddlers possess a genuine understanding of FB was challenged by lower-level interpretations. Rubio-Fernández et al. (2017) argued that children's training on factual "where" questions might prompt them to point to the last location where the target was. In the task by Setoh et al. (2016), this location coincides with the last location where the agent saw it, and thus searching there corresponds to the behavior rule to "search for an object where one last saw it disappear." In their response, Scott et al. (2017) pointed out that the two preceding "where" practice questions, which offered two alternatives, each made this interpretation unlikely. Fenici and Garofoli (2020) proposed a low-level associationist account, arguing that correct responses could be due to the strength of the association among the object (apple), the actor, and the location where the actor put her apple. In their reply, Setoh et al. (2020) argued that it is implausible that children's failure to point to the right location when given only one practice trial is due to a failure to form a

three-way association among actor, object, and location under higher processing demands. Rather, children would be expected to resort to a lower-level associationist processing heuristic under increased processing demands. Furthermore, the findings generalized to a different task and question type in the FB about identity task by Scott et al. (2020), which does not support the low-level accounts specified for the FB about location task.

In the current study, we aimed to evaluate the high-level conceptual continuity account proposed by Baillargeon et al. (2010, 2016) by investigating longitudinal relations between the low-demands FB task at 33 months of age and standard FB tasks at 52 months. The low-demands task by Setoh et al. (2016) yielded a competence level of more than 70% in 30- to 33-month-old children. A similar competence level is commonly reached in traditional FB tasks only at 4 and 5 years of age. Thus, the general age trend reflects continuity rather than developmental change. To determine whether there is continuity on the individual level between FBU in toddlerhood and at preschool age, we conducted a longitudinal study with two measurement points: Time 1 at 33 months and Time 2 at 52 months.

Participants' FBU was assessed with the low-demands FB task at Time 1 (33 months of age) and with two standard tasks from the ToM scale, a content FB task and a location FB task (Wellman & Liu, 2004) at Time 2 (52 months). We expected to find a unique source of variance between the low-demands task at 33 months and either one of the standard FB tasks at 52 months, that is, significant correlations independent of more general cognitive functioning. The low-demands FB task by Setoh et al. (2016) is a location FB task in which the protagonist, who was not present at the transfer event holds a false belief about the location of a target object. Similarly, the location FB task ("Paul and the backpack") from the ToM scale requires children to understand that a protagonist will search for a target object where he or she falsely believes it to be, not where it really is. In contrast, in the content FB task from the ToM scale, the protagonist holds a false belief about the contents of a candy tube that is widely known to typically contain Smarties but that really contains a piglet, a fact that was revealed to the child but not to the protagonist. If the low-demands location FB task at Time 1 predicts performance on the typical location FB task, but not on the content FB task at Time 2, it might be argued that the relationship is due to common superficial task features. In contrast, a high-level conceptual continuity account would be supported by task-independent longitudinal relations of the low-demands FB task with both standard FB tasks.

Above-chance performance on the low-demands FB task in toddlers was achieved by reducing executive function and language demands to a minimum. With respect to the long-standing theoretical debate about the relation between FBU and the development of executive functions (for reviews, see Miller & Marcovitch, 2012; Perner & Lang, 1999), the findings by Setoh et al. (2016) clearly support an expression account, positing that executive demands may hinder young children from expressing their conceptual understanding of the mind (e.g., Hala & Russell, 2001; Moses, 2001). In contrast, emergence accounts claim that executive control, developing rapidly around 4 years of age, is necessary for FB understanding to emerge (e.g., Benson et al., 2013; Devine & Hughes, 2014). A predictive longitudinal relation of performance on the low-demands FB task in toddlerhood and on high-demands FB tasks at 4 years of age, when controlling for executive functions and language, would strongly support an expression account.

In sum, the current study investigated the predictive relations between a low-demands FB task at 33 months of age and two standard FB tasks at 52 months. To assess the correlations of each of the FB tasks with linguistic and inhibitory skills, and to control for these general cognitive abilities, we used a standard language measure and a measure of executive inhibition.

Method

Participants

A total of 75 children (29 boys and 46 girls) participated in the current longitudinal study at 33 months of age (range = 32.5–34.0 months, *M* = 33.1 months = Time 1) and 52 months of age

(range = 51.4–55.4 months, *M* = 52 months = Time 2). A sample of 67 is considered sufficient to detect a medium-sized correlation at a power level of 80% (alpha = .05, one-tailed).[1]

There were missing data at Time 2. One child missed the language testing, and 6 children did not receive the inhibitory control task. Three children missed the content FB task, and 1 child needed to be excluded due to experimenter error. Two children missed the location FB task, and another 4 children needed to be excluded due to experimenter error (n = 3) or parent intervention (n = 1).

The current study was part of a larger longitudinal study on early ToM development. The data reported by Grosso et al. (2019) on FB understanding in 33-month-old children were gathered from a subsample of the current sample. All children were, at the time of the assessments, typically developing. The large majority of the children were monolingual German; about 25% grew up bilingually but with German as the primary language.

Data collection took place in a child-friendly laboratory at the university. Due to the outbreak of the COVID-19 pandemic, 10 of the 33-month-olds and 55 of the 52-month-olds received parts of the assessments via a videoconferencing tool. There were no significant differences for any of the tasks under study between videoconferencing and lab testing.

Low-demands FB task at 33 months of age

We created a picture book with nine pages composed of clear plastic sheet protectors holding white paper backgrounds to which 11 picture photos were attached. A black solid paperboard stand kept the pages in place. The stand allowed the screen to be positioned at a 70° angle. All photos were centered at the bottom of the page, with double photos placed 4.5 cm apart one from the other (see Fig. 1).

Children and the experimenter sat next to each other, and the stand was placed in the middle of the table in front of the children. An overhead camera captured this setting in top view. A second camera, positioned behind the children, captured the experimenter, the book, and childreńs responses. A second experimenter coded children's behavior during the test session. All responses were independently coded from video-recordings by another two raters, who reached very high inter-rater reliability for all questions (Cohen's kappa = 1).

Six story events, two practice trials, and one test trial were presented to the children. After flipping each page to reveal the respective picture(s), the experimenter recited the accompanying line of that event. The story first introduced Lily (Event 1), who found an apple in a bucket covered with a towel (Event 2). This was followed by the first practice trial in which a picture of the apple and a picture of a banana were shown, and children were asked, "Where is Lily's apple?" Then, in the third and fourth story events, Lily moved the apple into a basket, covered it with a plate, and went outside to play with a ball. In the second practice trial, the experimenter presented a picture of a rattle and a picture of the ball and asked, "Where is Lily's ball?" The story continued with the arrival of Lily's brother Peter, who took away the apple from the scene (Event 5). Lily then returned to look for her apple (Event 6). In the test trial, the experimenter revealed a picture of the basket and a picture of the bucket and asked, "Where will Lily look for her apple?" A pointing gesture or verbal referral to the container where Lily falsely believed that the apple was located was coded as a correct response. Children were excluded from the task if they failed to correctly answer one of the two practice trials.

FB understanding at 52 months of age

Children's ToM was assessed using two FB tasks from the German version of the ToM scale (Kristen et al., 2006). In the content FB task, children were asked to judge another person's FB about the content of a perceptually misleading container (a Smarties "box") when the children knew what was actually in it. First, children were shown a Smarties tube and were asked what they thought was inside the "box". Then, the tube was opened, revealing a piglet inside. The tube was closed again, and the

[1] Based on a G*Power analysis, we originally projected a sample size of 150. A total of 142 children participated at Time 1, but only 75 returned at Time 2. The main reason for participant attrition was the outbreak of the Covid-19 pandemic, which led to the loss of participants at Time 2 who could not be tested at the right age or whose parents withdrew their permission.

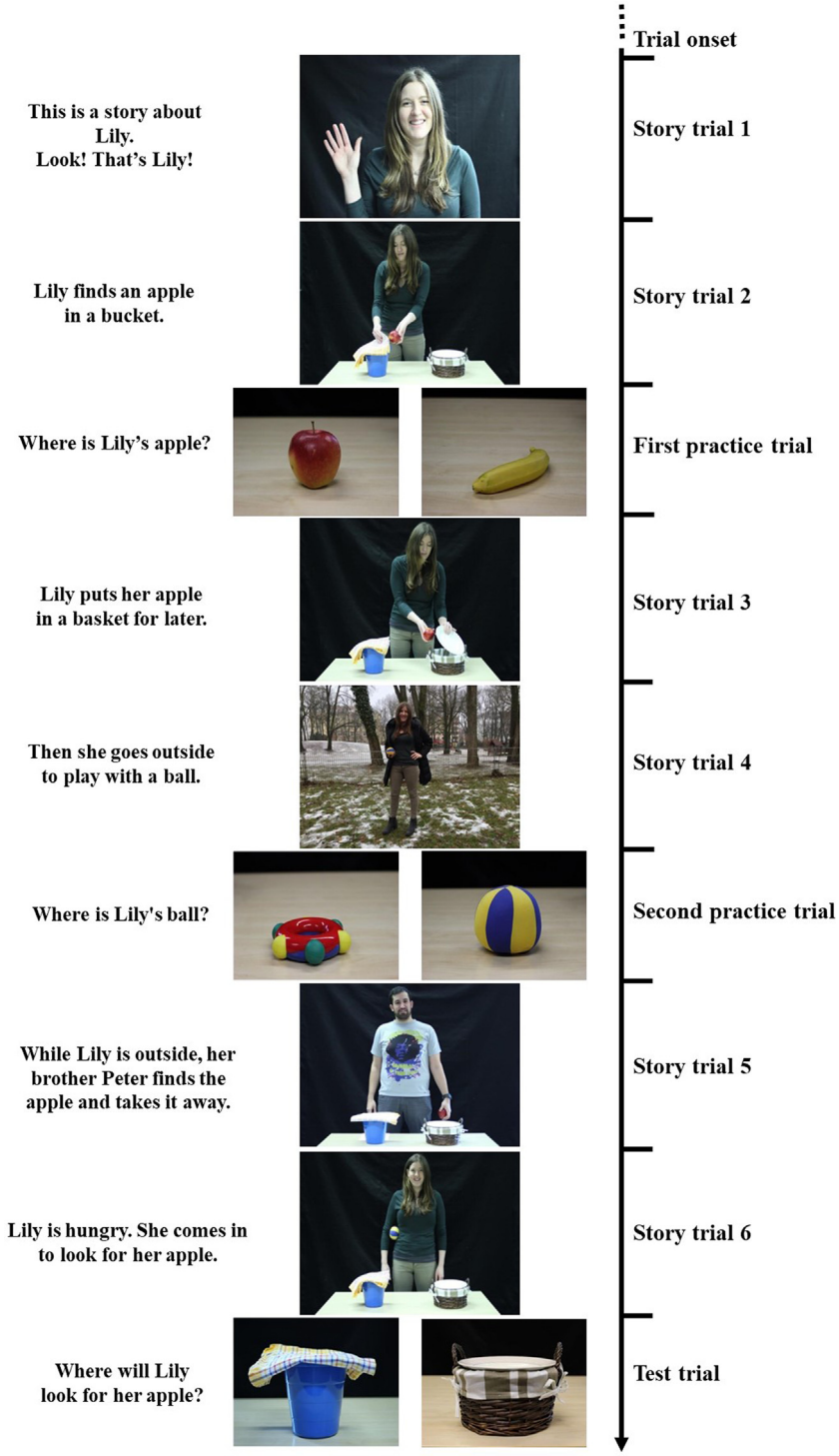

Fig. 1. Sequence of events in the low-demands FB task, adapted from Setoh et al., 2016.

children were asked a first control question (“What is in the box?”). Then, a Playmobil figure was introduced. Children were told that Lukas had never seen what was inside this box. Finally, the test question (“So, what does Lukas think is inside the box?”) and the second control question (“Has Lukas ever looked inside this box?”) were asked. Children were credited with 1 point for this task if they correctly answered the test question and the second control question. Children were excluded from the task if they failed to form the correct assumption about the content of the Smarties tube (i.e., Smarties candy)[2] or if they were unable to correctly answer the first control question.

The second task was a location (explicit) FB task. Participants needed to predict where a protagonist would search for an object based on his FB. Children were told a story about a Playmobil figure called Paul, who was looking for his mittens; his mittens might be either in the closet or in his backpack. These two possible locations were presented as colored drawings on a sheet of paper.[3] Then, children were told, “Really, Paul’s mittens are in his backpack. But Paul thinks that his mittens are in the closet.” This was followed by the test question (“Where will Paul look for his mittens?”) and the memory control question (“Where are Paul’s mittens really?”). Children were credited with 1 point for this task if they correctly answered the test and control questions.

SETK3-5 at 52 months of age

A German standardized language development test for 3- to 5-year-old children (SETK3-5; Grimm, 2015) was conducted with subtests assessing language comprehension, language production, and language memory. Children’s raw values in each of the tasks were transformed into standardized *t* values according to an age-specific norm table.

Day–night Stroop at 52 months

Children’s executive functions were assessed using a Stroop-like task based on Gerstadt et al. (1994). Small cards depicting either a yellow moon and stars against a black background or a yellow sun against a white background were used. Children were instructed to say “day” when shown a card with moon and stars and to say “night” when shown a card with a sun. Then, up to 12 training trials were conducted to ensure that children understood the task. The experimenter presented the cards one after the other, and children received corrective feedback on their answers. After four consecutive correct trials or after a maximum of 12 training trials, testing began. The test phase consisted of 16 cards but was terminated after 12 cards if the training phase had been terminated after four correct responses in a row (and these responses were used for Cards 13 to 16 from the test phase). No feedback was given in the test phase. Children received a score from 0 to 16 based on the number of correct responses in the test phase.

Results

The majority of the children (72%) passed the low-demands FB task at 33 months of age. The success rates in the standard FB tasks at 52 months of age were 66% and 59% for the content and explicit FB tasks, respectively (see Table 1). There was no significant difference in difficulty between the low-demands FB task and the content FB task, whereas the location FB task was marginally significantly more difficult for the 52-month-olds than the low-demands task was for the 33-month-olds (McNemar’s test, $p = .08$; $N = 69$). Cross-sectionally, the location FB and content FB tasks were correlated (*phi* = .27, $p = .03$; $N = 67$). A nonlinear logistic regression analysis on location FB showed that content

[2] Two children appeared to assume that the experimenter used the box to store toys and guessed (e.g., “a lion”). However, when asked subsequently what Lucas thought was in the box, they promptly replied “Smarties.” These children were scored as correct.

[3] After presenting the two locations, one experimenter mistakenly prompted 4 children to utter their own guess about where the mittens were. All these children believed they were in the backpack. After this, the task was continued in the standard way. All 4 children correctly answered the control question, and 3 of the 4 answered the test question correctly. There was no change in any of the findings reported below if these 4 children’s data were removed.

FB was the only significant predictor (p = .02; N = 62) when accounting for language and executive function. See bivariate correlations in Table 2.

The low-demands FB task was longitudinally correlated with the content FB task (*phi* = .28, p = .02, two-tailed; N = 71). A nonlinear logistic regression analysis on content FB performance with performance on the low-demands FB task, language ability, and executive function as predictors yielded a significant effect of FB understanding on the low-demands task at 33 months of age when accounting for language and executive function (Table 3).

In contrast, the association between the low-demands FB task at 33 months of age and the location FB task at 52 months did not reach significance (see Table 2). A nonlinear logistic regression analysis on location FB performance with performance on the low-demands FB task, language ability, and executive function as predictors yielded a significant effect of executive function on location FB performance (p = .028) when one-tailed testing was used (Table 4). No other predictors were significant.

The correlation of the low-demands FB task at 33 months of age and the sum score of the location and content FB tasks at 52 months was significant (r = .25, p = .04, two-tailed; N = 68). The partial correlation remained significant when both language ability and executive function were partialled out (r = .21, p = .048, one-tailed; df = 60).

Discussion

Infants show an understanding of FB in spontaneous response tasks, and children as young as 30 months master traditional elicited response FB tasks when inhibition and response generation demands are lowered (Scott et al., 2020; Setoh et al., 2016). One interpretation of these findings is that FBU develops in infancy and is conceptually continuous across the life span. If there is conceptual continuity in FBU, then there should be longitudinal relations among FB assessments from infancy to childhood and beyond. To date, there is evidence for predictive relations from infant implicit FBU to explicit verbal FBU at 4 to 6 years of age (Sodian et al., 2020), but the relation of explicit FBU in toddlerhood and later explicit FBU is a missing link.

The current longitudinal study found evidence for a long-term predictive relation of FBU from toddlerhood to preschool age that was independent of language ability and executive function. The low-demands FB task (Setoh et al., 2016), which was administered at 33 months of age, was significantly correlated with the content FB task 19 months later. This correlation cannot be attributed to superficial features of the two tasks given that the low-demands task was a location task, which was correlated with a task assessing children's understanding of a false belief about the content of a typical container. Therefore, the unique source of variance cannot be accounted for by the application of the same behavior rule, such as "People will search for an object where they last saw it disappear," or by the formation of an actor–object–location association. The low-demands task was also correlated with the sum score of the content and location FB tasks at 52 months of age. Thus, the current findings support a high-level conceptual interpretation of the relation between early and later FBU.

On the level of individual tasks, the current study did not, however, find a significant relation between the low-demands task at 33 months of age and the location FB task at 52 months. As the correlational pattern shows, the location FB task was significantly correlated with executive function (inhibitory control). Other than in the low-demands task at 33 months, children in the location FB task at 52 months needed to inhibit their knowledge of the real location of the target object to predict the protagonist's mistaken search. Still, it appears puzzling that success on the location FB task was not significantly predicted by performance on the low-demands task, when controlling for executive function, given that the two tasks shared similarities in task format. Importantly, however, the location FB task that was used in the current study differed from the standard "Sally–Anne" location FB task format that was used in the low-demands task. Whereas the Sally–Anne format provides children with information on the agent's perceptual access to the story events (or lack thereof) from which they can infer the agent's FB, the current location FB task provides participants solely with an explicit statement of the agent's belief without any information on the sources of this belief. To solve this task, children need to possess a firm understanding of the causal impact of beliefs on action, from which they need to derive their answer irrespective of conflicting information on the state of reality. Thus, the current

Table 1
Descriptive statistics of performance on all measures.

Measure	*M* (*SD*)	Range	*N*
FB task 33 months	.72 (.45)	0–1	75
Content FB task 52 months	.66 (.48)	0–1	71
Explicit FB task 52 months	.59 (.50)	0–1	69
Language test (SETK3-5) 52 months	44.0 (5.12)	35–58	74
Executive function 52 months	12.0 (4.49)	0–16	69

Note. FB, false belief.

Table 2
Bivariate correlations among study variables.

Task	1	2	3	4	5
1. Low-demands FB		.28*	.18	.03	.19
2. Content FB			.27*	.23~	−.00
3. Location FB				.02	.32**
4. Language					.03
5. Executive function					

Note. Shown are two-tailed contingency (phi) coefficients among the three false belief (FB) tasks; Spearman rank correlations of language ability, executive function, and FB tasks; and Pearson correlations of language ability and executive function.
* $p < .05$.
** $p < .01$.
~ $p < .10$.

Table 3
Logistic regression analysis on content FB performance.

	B	*SE*	*Wald*	*df*	Significance	Exp(*B*)
Predictor						
Low-demands FB	1.30	0.62	4.45	1	.035	3.67
Language	.09	.06	2.31	1	.13	1.10
Executive function	−.001	.07	.00	1	.99	.99
Constant	−4.17	2.85	2.14	1	.14	.02

Note. $N = 66$. FB, false belief.

Table 4
Logistic regression analysis on location FB performance.

	B	*SE*	*Wald*	*df*	Significance	Exp(*B*)
Predictor						
Low-demands FB	.42	.61	.47	1	.49	1.52
Language	−.13	.05	.06	1	.80	.99
Executive function	.12	.06	3.65	1	.056	1.13
Constant	−0.77	2.48	.09	1	.76	.47

Note. $N = 64$. FB, false belief.

location FB task specifically focused on the link between belief and action rather than presenting the whole framework of causes and consequences of a false belief. This specificity is also reflected in the moderate correlation of the location and content FB tasks at 52 months of age. Informational causation and the causal impact of beliefs on action have long been recognized as two facets of belief understanding (Perner, 1991). However, there is little research on the mastery of these individual components of belief understanding in children. In typical FB paradigms such as the Sally–Anne task, children can succeed by inferring mistaken action from an agent's incomplete access to information

that generated a false belief. Thus, understandings of the causal impact of information on belief and of belief on action are not tested in isolation. When tested in isolation, understanding the causal impact of belief on action may still be challenging for 4-year-olds, as the current findings show. This difficulty was mainly attributable to the executive demands of the task, and these demands interact with the conceptual content of the task. Children need to inhibit their conflicting information on the real location of the hidden object in order to predict that the agent's search will be determined by his or her (false) belief. Thus, emerging conflict inhibition skills may foster children's understanding of the causal impact of beliefs on action.

In sum, the current study provides evidence for both continuity and change in the development of FBU from toddlerhood to preschool age. Conceptual continuity was supported by a longitudinal predictive relation of a low-demands FB task and a standard content FB task 19 months later. This finding also confirms an expression account of the relation of executive function and FBU. In contrast, the finding that performance on a specialized location FB task that focused on the causal impact of beliefs on action was only predicted by executive function (conflict inhibition) suggests that developmental progress in executive functions may be essential for a firm understanding of the causal link of beliefs and action to develop. This interpretation is consistent with a moderate version of an emergence account of executive function and FBU, with the developmental change being incremental rather than fundamental as was previously assumed. Toddlers' FBU appears to be centered around a conceptual core consisting of the interrelations of perceptual access, knowledge or belief, and action. To master tasks assessing individual elements of this causal framework in isolation, children's initial representations need to get both strengthened and enriched in the course of development (see Carey, 2009).

The current study is the first to investigate longitudinal stability of explicit FBU from toddlerhood to 4 years of age. The findings support a high-level conceptual continuity interpretation of the development of FBU, thereby contributing to an ongoing theoretical debate about the origins of ToM. The study needs to be followed up by further longitudinal investigations that include larger batteries of both FB and executive function tasks. Moreover, the developmental relation of implicit FBU in infancy and explicit FBU, assessed by the low-demands task in toddlerhood, is still a missing link.

CRediT authorship contribution statement

Beate Sodian: Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Data curation, Conceptualization. **Larissa J. Kaltefleiter:** Writing – review & editing, Project administration, Methodology, Investigation, Data curation. **Tobias Schuwerk:** Writing – review & editing, Supervision, Software, Methodology, Data curation. **Daniela Kloo:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization.

Data availability

The authors do not have permission to share data.

Acknowledgment

This research was funded by a grant (SO 213/33-1,2) from Deutsche Forschungsgemeinschaft.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970. https://doi.org/10.1037/a0016923.

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations [invited commentary]. *Cognitive Development, 46*, 112–124. https://doi.org/10.1016/j.cogdev.2018.06.001.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159–186. https://doi.org/10.1146/annurev-psych-010213-115033.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006.

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior and Development, 57*. https://doi.org/10.1016/j.infbeh.2019.101350 101350.
Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford University Press.
Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*(9), 1615–1627. https://doi.org/10.1037/a0031056.
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. https://doi.org/10.1016/j.cognition.2009.05.006.
Carey, S. (2009). *The origin of concepts*. Oxford University Press.
Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development, 46*, 51–57. https://doi.org/10.1016/j.cogdev.2017.10.003.
Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*(5), 1777–1794. https://doi.org/10.1111/cdev.12237.
Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development, 46*, 12–30. https://doi.org/10.1016/j.cogdev.2018.01.001.
Fenici, M., & Garofoli, D. (2020). An associationist bias explains different processing demands for toddlers in different traditional false-belief tasks. *Human Development, 64*, 4–6. https://doi.org/10.1159/000505208.
Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition, 53*, 129–153. https://doi.org/10.1016/0010-0277(94)90068-x.
Grimm, H. (2015). Sprachentwicklungstest für drei bis fünfjährige Kinder (SETK 3 5): Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen (3.Aufl.). Hogrefe.
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. https://doi.org/10.1016/j.infbeh.2018.09.012.
Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology, 80*(2), 112–141. https://doi.org/10.1006/jecp.2000.2627.
Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science, 17*(5), 647–659. https://doi.org/10.1111/desc.12148.
Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society Open Science, 8*(8). https://doi.org/10.1098/rsos.210190 210190.
Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development, 44*(2), 107–115. https://doi.org/10.1177/0165025419850901.
Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792.
Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von "theory of mind" Aufgaben. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(4), 186–195. https://doi.org/10.1026/0049-8637.38.4.186.
Kulke, L., & Rakoczy, H. (2018). Implicit theory of mind—An overview of current replications and non-replications. *Data in Brief, 16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016.
Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition, 121*(3), 289–298. https://doi.org/10.1016/j.cognition.2011.07.011.
Miller, S. E., & Marcovitch, S. (2012). How theory of mind and executive function co-develop. *Review of Philosophy and Psychology, 3*, 597–625. https://doi.org/10.1007/s13164-012-0117-0.
Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development, 72*(3), 688–690. https://doi.org/10.1111/1467-8624.00306.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. https://doi.org/10.1126/sceince.1107621.
Perner, J. (1991). *Understanding the representational mind*. MIT Press.
Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337–344. https://doi.org/10.1016/s1364-6613(99)01362-5.
Poulin-Dubois, D., Azar, N., Elkaim, B., & Burnside, K. (2020). Testing the stability of theory of mind: A longitudinal approach. *PLoS One, 15*(11). https://doi.org/10.1371/journal.pone.0241721 e241721.
Poulin-Dubois, D., Goldman, E. J., Meltzer, A., & Psaradellis, E. (2023). Discontinuity from implicit to explicit theory of mind from infancy to preschool age. *Cognitive Development, 65*. https://doi.org/10.1016/j.cogdev.2022.101273 101273.
Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development, 46*, 40–50. https://doi.org/10.1016/j.cogdev.2017.10.004.
Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology, 1*(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z.
Rubio-Fernández, P., Jara-Ettinger, J., & Gibson, E. (2017). Can processing demands explain toddlers' performance in false-belief tasks? *Proceedings of the National Academy of Sciences of the United States of America, 114*(19). https://doi.org/10.1073/pnas.1701286114 E3750.
Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., Dörrenberg, S., Fisher, C., Franchin, L., Fulcher, T., Garbisch, I., Geraci, A., Grosse Wiesmann, C., Hamlin, K., Haun, D., Hepach, R., Hunnius, S., Hyde, D. C., Karman, P., … Rakoczy, H. (2024). Action anticipation based on an agent's epistemic state in toddlers and adults. *PsyArXiv*. https://doi.org/10.31234/osf.io/x4jbm.
Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science, 5*(5). https://doi.org/10.1098/rsos.172273 172273.
Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012.

Scott, R. M., Roby, E., & Setoh, P. (2020). 2.5-year-olds succeed in identity and location elicited-response false-belief tasks with adequate response practice. *Journal of Experimental Child Psychology, 198*. https://doi.org/10.1016/j.jecp.2020.104890 104890.

Scott, R. M., Setoh, P., & Baillargeon, R. (2017). Reply to Rubio-Fernández et al.: Different traditional false-belief tasks impose different processing demands for toddlers. *Proceedings of the National Academy of Sciences of the United States of America, 114* (19), E3751–E3752. https://doi.org/10.1073/pnas.1703665114.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science, 22*(7), 878–880. https://doi.org/10.1177/0956797611411584.

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

Setoh, P., Scott, R. M., & Baillargeon, R. (2020). Reply to Fenici and Garofoli: Why would toddlers act on low-level associations only when processing demands are reduced? *Human Development, 64*, 7–9. https://doi.org/10.1159/00050680.

Sodian, B. (2023). Reply to Poulin-Dubois et al. (2023): Replication problems concerning both implicit and explicit false belief reasoning greatly reduced the chance of finding longitudinal correlations. *Cognitive Development, 65*, 1–3. https://doi.org/10.1016/j.cogdev.2022.101294.

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives, 14*(3), 171–177. https://doi.org/10.1111/cdep.12381.

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development, 87*(4), 1221–1232. https://doi.org/10.1111/cdev.12533.

Song, H. J., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition, 109*(3), 295–315. https://doi.org/10.1016/j.cognition.2008.08.008.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x.

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30*(1), 172–187. https://doi.org/10.1111/j.2044-835X.2011.02067.x.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology, 15*(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x.

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. https://doi.org/10.1016/0010-0277(83)90004-5.