AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

2021, Vol. 57, No. 8, 1210–1227
https://doi.org/10.1037/dev0001012

Do Complement Clauses Really Support False-Belief Reasoning? A Longitudinal Study With English-Speaking 2- to 3-Year-Olds

Ditte Boeg Thomsen[1], Anna Theakston[2], Birsu Kandemirci[3], and Silke Brandt[4]

[1] Department of Cross-Cultural and Regional Studies, University of Copenhagen
[2] Division of Human Communication, Development and Hearing, University of Manchester
[3] Department of Psychology, Kingston University
[4] Department of Linguistics and English Language, Lancaster University

To examine whether children's acquisition of perspective-marking language supports development in their ability to reason about mental states, we conducted a longitudinal study testing whether proficiency with complement clauses around age 3 explained variance in false-belief reasoning 6 months later. Forty-five English-speaking 2- and 3-year-olds (23 female, Time 1 age range = 33–41 months) from middle-class families in the North-West of England took part in the study, which addresses a series of uncertainties in previous studies. We avoided the confound of using complement clauses in the false-belief tests, assessed complement-clause proficiency with a new comprehensive test designed to capture gradual development, and controlled for individual differences in executive functioning that could affect both linguistic and sociocognitive performance. Further, we aimed to disentangle the influence of two aspects of complement-clause acquisition: proficiency with the perspective-marking syntactic structure itself and understanding of the specific mental verbs used in this syntactic structure. To investigate direction of causality, we also tested whether early false-belief reasoning predicted later complement-clause proficiency. The results provide strong support for the hypothesis that complement-clause acquisition promotes development in false-belief reasoning. Proficiency with the general structure of complement-clause constructions and understanding of the specific mental verbs "think" and "know" in third-person complements at Time 1 both contributed uniquely to predicting false-belief performance at Time 2. However, false-belief performance at Time 1 also contributed uniquely to predicting complement-clause proficiency at Time 2. Together, these results indicate a bidirectional relationship between linguistic and sociocognitive development.

Keywords: complement clauses, false belief, language acquisition, mental verbs, social cognition

Supplemental materials: https://doi.org/10.1037/dev0001012.supp

Being able to see situations from others' perspectives, being aware of one's own perspective, and being able to reason about mental phenomena, such as mistakes and surprises, are central sociocognitive skills. During the preschool years, children develop abilities to represent and to reason flexibly about their own and others' false beliefs, that is, beliefs that clash with their own current understanding of reality. A body of evidence suggests that linguistic experience promotes this development (e.g., Milligan et al., 2007; Schick et al., 2007; Tomasello, 2018). Whereas different linguistic skills appear to play a role, and the choice of linguistic tool can be affected by cross-linguistic differences in availability and usage patterns (e.g., Cheung et al., 2009; Tomasello & Rakoczy, 2003), it has been argued that for languages like English and German, children's acquisition of finite complement clauses is especially advantageous for facilitating false-belief reasoning (e.g., de Villiers & Pyers, 2002; Lohmann & Tomasello, 2003).

The complement-clause construction is a crosslinguistically widespread type of perspective-marking grammar that allows speakers to communicate flexibly and explicitly about the relationships between persons and propositions (Boeg Thomsen, 2016; Verhagen, 2005). In this syntactic construction, a complement clause expressing a proposition, for example, "It's his ball," is embedded in another clause, for example, "He says . . ." or "I hope . . .," which presents a specific viewpoint on that proposition, as in "He says [it's his ball]" or "I hope [it's his ball]." Complement-clause constructions present

We thank Victoria Southgate for sharing her video materials for the implicit False Belief test. Warm thanks also to all the children and parents who participated in the study and to Katharina Kaduk and the Lancaster University Babylab for helping us with recruitment and letting us use their lab space for testing. Thanks to Liam Blything, James Brand, Claire Nance, and Tony Trotter for lending their voices to various puppets. Finally, we thank Laura Stiefenhöfer for instructing us on how to use the eye-tracker.

Data sets for all tasks and R scripts are available on the Open Science Framework: https://osf.io/qkjn6/ (Boeg Thomsen, Theakston, et al., 2021).

Correspondence concerning this article should be addressed to Silke Brandt, Department of Linguistics and English Language, Lancaster University, County South, Lancaster, LA1 4YL, United Kingdom. Email: s.brandt@lancaster.ac.uk

the unique trait of suspending speaker commitment, providing a means for talking about propositions that the speaker does not believe in, as in "He says [it's his ball], but actually it's not." De Villiers and de Villiers (2000) therefore proposed that complement clauses support children's false-belief understanding by providing a representational format for reasoning about false beliefs.

In a longitudinal study, de Villiers and Pyers (2002) found empirical support for this hypothesis: children's comprehension of complement clauses such as "She thought [it was a toy bird], but it was really a funny hat. What did she think she bought?" at around age 3 years, 8 months predicted their ability to pass false-belief tests 4 months later. Examining causality more directly, training studies by Lohmann and Tomasello (2003), Hale and Tager-Flusberg (2003) and Mo et al. (2014) found improved false-belief performance in German-, English- and Mandarin-speaking children trained with complement clauses. In atypical development, similar training effects have been reported for Danish-speaking children with autism (Boeg Thomsen, 2016) and French-speaking children with autism, developmental language disorder (DLD), and (as a control) typical development (Durrleman et al., 2019). Although not addressing causality, other studies with typically developing children (Brandt et al., 2016; Burnel et al., 2020; Durrleman et al., 2016, 2017; Durrleman & Franck, 2015; Grosse Wiesmann et al., 2017; Low, 2010; Moore et al., 1990), children with DLD (Durrleman et al., 2017; Miller, 2004), deaf children (Schick et al., 2007), and children with autism (Durrleman et al., 2016, 2017; Durrleman & Franck, 2015; Lind & Bowler, 2009) have found concurrent correlations between complement-clause proficiency and false-belief reasoning.

However, there are critical uncertainties pertaining to previous conclusions about a tight relationship between complement-clause proficiency and false-belief reasoning. First, most studies use complement clauses in their tests to assess false-belief ability, thus confounding the two skills to be compared. Second, it is often unclear what exactly the tests of complement-clause mastery are measuring. Third, few studies include measures of executive-function skills, making it impossible to exclude the possibility that complement comprehension and false-belief ability correlate because success on both tasks depends on the same underlying executive-functioning skills.

The present longitudinal study was designed to test whether we can replicate de Villiers and Pyers' (2002) finding that complement-clause mastery predicts later false-belief reasoning when we (a) use false-belief tests without complement clauses, (b) use a more precise measure of complement-clause mastery, and (c) control for a range of executive-functioning and other background measures. In addition, we address two further issues: First, previous research has focused on language and false-belief understanding around 4 years of age, but children make important advances in both false-belief reasoning (e.g., Hansen, 2010) and complement-clause mastery at younger ages (e.g., Boeg Thomsen, 2016; Brandt et al., 2010), suggesting that the crucial phase for an interplay may start earlier. Second, it is unclear which aspects of complement-clause acquisition might play a critical role in children's false-belief reasoning. It could be the syntactic embedding structure (de Villiers & Pyers, 2002) or specific mental verbs frequently used in the construction (e.g., Brandt et al., 2016). In contrast with previous studies, we use separate tests to examine different degrees of abstraction in children's proficiency with complement clauses in general on the one hand and their understanding of specific high-frequency mental verbs (*think* and *know*) in complement-clause constructions on the other, to discern whether both play an independent role in supporting false-belief reasoning. To further ascertain the specificity of the potential influence from mental verbs in complement-clause constructions, we compare them with a different type of certainty-marking verbs in simple clauses: epistemic modals (cf. Moore et al., 1990). We now explore each of the identified issues with previous research in more detail before outlining the current study.

Confounds and Underspecifications in Previous Studies

Linguistic Confounds in False-Belief Tests

One persistent problem in studies comparing false-belief reasoning and complement-clause proficiency is that the vast majority use complement clauses in some or all of their false-belief tests, thus making complement-clause mastery a prerequisite for passing these sociocognitive tests. For example, if we compare questions in the complement-clauses test and the unexpected contents false-belief test in de Villiers and Pyers (2002), we see substantial lexical and structural similarities between them:

"What did she think she bought?"
(Complement clauses)

"What did you think was in the box?"
(False belief: Unexpected contents)

"What will Sarah think is in the box?"
(False belief: Unexpected contents)

Similar lexical and structural confounds are evident in most other longitudinal, training, and correlational studies (e.g., Brandt et al., 2016; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). Thus, when studies report correlations or predictive relations between false-belief and complement-clause ability or effects of complement-clause training, it is impossible to gauge whether they emerge because complement-clause acquisition has a deep conceptual influence on mental-state reasoning or because it allows children to understand the false-belief test questions.

Promisingly, a number of studies targeting atypical development (e.g., deafness, DLD, and autism) avoided this confound by using false-belief tests without complement clauses and still found correlations between complement-clause comprehension and false-belief understanding (Durrleman et al., 2016, 2017; Schick et al., 2007) and effects of complement-clause training (Boeg Thomsen, 2016; Durrleman et al., 2019). In the current study, we also avoid the use of complement clauses in the false-belief tests while targeting typically developing children and using a longitudinal design.

Measures of Complement-Clause Mastery

A second reason for caution when interpreting previous studies on complement clauses and false belief is that it is not entirely clear what kind of proficiency the complements tests are measuring due to three specific properties of the tests: (a) treating mastery as a binary measure (present/absent), (b) the use of a nonprototypical construction, (c) high executive-functioning demands.

Treating Mastery as a Binary Measure. The most widely used complements test, Memory for Complements, was developed in the late 1990s (de Villiers & Pyers, 1997). The conception of complement-clause proficiency underlying this test and later adaptions of it is that there is an abstract syntactic complement construction that the child either has or has not acquired. Children are presented with four to 12 items of the same structure with one or two different recurring complement-taking verbs. If the child performs well with these verbs (typically *say*, *think* or *tell*), they are taken to have acquired the construction, whereas poor performance is taken to indicate that the child has not yet acquired the syntax of complementation. Over the last 20 years, however, corpus studies and production experiments have indicated that acquisition of complement syntax proceeds in a gradual manner, and that verb frequency plays a central role in this process (Brandt et al., 2010; Diessel & Tomasello, 2001; Kidd et al., 2010). The evidence suggests that children slowly build up abstract and flexible schemas, starting with lexically specified chunks (e.g., "I think . . .") and low-level schemas based around verbs that are frequently used in complement-clause constructions in their input (e.g., [NP] SAY [S], [NP] KNOW [S]). Complements tests that use only one or two verbs cannot capture such gradual development, and it is unclear why performance on only one or two high-frequency verbs should be the level of abstraction that is relevant for false-belief reasoning.

Use of a Nonprototypical Construction. A related problem is posed by those complements tests that take children's performance with the long-distance dependency construction (e.g., "What did the girl say she was cutting?") as representative of children's proficiency (e.g., de Villiers & Pyers, 2002; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). First, Dąbrowska (2004) showed that long-distance dependency constructions are very infrequent and that these constructions emerge late in child speech, mostly around age 4, relative to declarative complement-clause constructions such as "He said he has something to play with for me" at 2 years, 10 months (Diessel & Tomasello, 2001). Moreover, compared with declarative complement-clause constructions, these long-distance dependency constructions show high levels of lexical specificity (Dąbrowska, 2004). Verhagen (2005) suggested that the properties of this stereotypic lexically specific construction "do not follow from general properties of complementation, and thus also should not be accounted for in such general terms" (p. 126). Thus, it is unclear how long-distance dependency questions can be representative of children's complement mastery, or why acquisition of this specific construction should be related to false-belief development.

High Executive-Functioning Demands. Many of the previously used complements tests place high demands on executive functioning. For example, the memory for complements test (e.g., de Villiers & Pyers, 2002) places high demands on inhibitory control by requiring children to ignore salient visual information in order to pass. The visual stimuli accompanying the test questions always show a close-up of the distractor referent, but never the target referent (e.g., in the item "The woman said the girl had a bug in her hair. But it was only a leaf! What did the woman say the girl had in her hair?", the child sees a close-up of the leaf, but no picture of the bug). This test also poses high demands on working memory by always presenting the distractor clause ("But it was only a leaf!") after the complement clause ("The woman said the girl had a bug in her hair"), requiring the child to keep information from the complement clause in mind while processing the distractor clause. Because previous studies typically do not provide independent measures of memory and response inhibition, it is unclear to what degree complement-clause performance reflects proficiency with complements or executive function.

This is problematic because performance on false-belief tests has repeatedly been found to correlate with executive functioning (e.g., Carlson et al., 2004; Hughes & Ensor, 2007), entailing the risk that correlations between performance with complements and false belief could depend on underinvestigated shared executive-function requirements. Among the few studies controlling executive-function measures, de Villiers and de Villiers (2012) still found a unique contribution of complement-clause comprehension to explaining concurrent false-belief performance, as did Durrleman and Franck (2015) and Burnel et al. (2020); though promising, these concurrent correlations are not ideal for evaluating causality.

How Early in Development Can We Discern a Complements-False Belief Relationship?

Traditionally, studies comparing complements and false belief have focused on the transition from 3 to 4 years, following Wellman et al.'s (2001) meta-analysis of false-belief studies where above-chance performance was found from 48 months. However, the meta-analysis also found above-chance performance from 40 months when the most supportive task modifications were used. In addition, Rubio-Fernández and Geurts (2013) argued that young children may fail explicit false-belief tasks because removing the protagonist from the scene disrupts their ability to track the protagonist's perspective. They demonstrated that 3-year-old children pass unseen location change tasks if the protagonist stays in the child's visual field throughout the story, reducing memory demands (see also Rubio-Fernández and Geurts [2016] for attention constraints). Finally, Hansen and colleagues (Hansen, 2010; Hansen & Markman, 2005) have argued that young children may fail false-belief tests because they interpret the test questions as requests for indirect information about the real state of affairs and try to be cooperative by informing the experimenter about reality. They found that 3-year-olds were able to pass location change, appearance-reality, and unexpected contents tests, when the experimenter stressed shared knowledge about the real state of affairs (see also Helming et al., 2014).

It appears that 3-year-olds already exhibit false-belief understanding in pragmatically transparent tasks with low memory demands. Thus, to argue that complements support development in false-belief reasoning, it is necessary to demonstrate a relationship between complement-clause acquisition and false belief at a younger age than the transition from 3 to 4 years. Because 2- and 3-year-olds have typically already begun to acquire complement clauses (Boeg Thomsen, 2016; Brandt et al., 2010), it is not inconceivable that they will show an influence from complement-clause acquisition on false-belief reasoning. We therefore targeted 2-year-olds (from 2 years, 9 months) and young 3-year-olds (up to 3 years, 5 months) at our Time 1 (T1) testing.

Mental Verbs in Complement-Clause Constructions

Finally, it is important to clarify which aspects of complementation may support sociocognitive development. One possibility is that children build up an abstract and general complement-clause construction which allows them to represent different mental perspectives on propositions (Boeg Thomsen, 2016; Verhagen, 2005). On the other hand, experience with specific frequent mental verbs in the construction (e.g., *think* and *know*) may help children consolidate their understanding of different mental states and degrees of certainty, supporting false-belief reasoning (e.g., Brandt et al., 2016; Moore et al., 1990).[1]

If knowledge of high-frequency mental verbs in complement-clause constructions supports mental-state reasoning, seemingly small differences in their linguistic contexts may make a difference, too. Diessel and Tomasello (2001) suggested that children treat the high-frequency string "I think" as an adverb-like epistemic marker without explicit mental-state reference. Further, Howard et al. (2008) showed that mothers often use "I think" in certainty contexts (e.g., "I think we better tidy up"). Thus, encountering "think" in first-person clauses (i.e., with "I") might not provide children with the same support for false-belief understanding as third-person clauses (e.g., "He thinks . . ."; see also Lewis et al., 2017). Indeed, using the hidden-objects task, Brandt et al. (2016) compared children's comprehension of *think* and *know* in first- and third-person complements with their false-belief performance and found that only performance with third-person complements correlated with 4-year-olds' success on questions requiring them to remember their own previous false belief. This suggests a weaker relationship between first-person complements with these high-frequency verbs and false belief (for similar findings, see Gola, 2012).

In addition, it is unclear whether complement clauses perform a privileged role in supporting false-belief understanding, or whether other types of verbal epistemic markers can perform the same function. On the one hand, epistemic modals such as *might* and *must* indicate degree of certainty toward the proposition and could thus support children in building up categories for certainty and uncertainty. Indeed, Moore et al. (1990) found correlations between 4-year-olds' performance with epistemic modals and false belief. On the other hand, unlike mental verbs in complement-clause constructions, epistemic modals do not explicitly attribute the mental state of (un)certainty to a person (consider "I think it's a flower" vs. "It might be a flower"), and thus may not afford the same opportunities to draw an explicit link between a mental state and a specific person.

The Current Study

The main question in the current longitudinal study is whether children's acquisition of perspective-marking language supports developments in their ability to reason about their own and others' mental states. Specifically, we ask whether we can replicate de Villiers and Pyers' (2002) finding that proficiency with complement clauses explains later variance in false-belief reasoning when we (a) use pragmatically adjusted false-belief tests without complement clauses, (b) test complements proficiency with a fine-grained test suitable for capturing degrees of constructional flexibility, and (c) control for a series of background variables, including inhibitory control, rule-switching flexibility, working memory, short-term memory, vocabulary, and grammar. To examine direction of causality in the hypothesized relationship between linguistic and sociocognitive development, we also test the reverse longitudinal relation, that is, whether early false-belief reasoning predicts later complement-clause proficiency.

Following studies demonstrating skills with both false belief and complements in three-year-olds, we focus on children who are around 3 years of age (range = 2 years, 9 months–3 years, 5 months) at Time 1 (T1) and around 3 years, 7 months (range = 3 years, 3 months–3 years, 11 months) at Time 2 (T2), asking whether we can discern the roots of a relationship at an earlier age than in previous studies.

Finally, we address the question of which aspects of complementation—if any—help children recognize and reason about mental states. Is it the flexible embedding structure, the specific type of mental verbs used to indicate degrees of certainty, or do both contribute to making children's belief reasoning more stable and flexible? To answer this last question, we use tests of structural flexibility and understanding of the mental verbs *think* and *know* in complement-clause constructions, presented with first- versus third-person subjects, and compare their role with that of another class of certainty-marking verbs, epistemic modals (*might* and *must*).

Method

Participants

Participants were recruited from Lancaster University Babylab's database. The vast majority of participants in the database come from a white middle-class background and more than 70% of the caregivers attended higher education, with about one third holding a postgraduate degree. Fifty-one 2- and 3-year-olds took part. Three were excluded due to not engaging with multiple tasks, leaving 48 children (24 female; age range = 33–41 months, M = 36.85, SD = 2.34) included at T1.[2] We recruited eight boys and eight girls in each of three 3-month windows (range = 2 years, 9 months–2 years, 11 months; 3 years–3 years, 2 months; 3 years, 3 months–3 years, 5 months). All participants were monolingual English speakers with no known history of language or hearing impairment. Forty-four percent had no siblings, 37.5% were the youngest sibling, 14.5% were middle children, and 4% were the oldest child in their family.[3] The retention rate for T2 was 93.75% (n = 45) with three participants dropping out between T1 and T2.

[1] We focus here on mental verbs, but there are also intriguing findings on the relationships between false belief and complements of perception vs. communication verbs (see Durrleman et al. [2016, 2017] for French-speaking children and Perner et al. [2003] for German-speaking children).

[2] Sample size was determined by following the standard recommendations for regression modeling of having at least 10 observations per variable examined by the model (Harrell, 2001). Aiming to examine 22 variables (one was later dropped [see Footnote 5]), we needed at least 220 observations, and as we could not present children with more than five false-belief questions (the tasks build on surprise, and responses get unreliable if children get used to their structure), we needed a minimum of 44 participants returning at T2, which with a retention rate at around 90% would require 48 participants at T1. Since we designed the study, sophisticated and precise ways of calculating power for mixed-effects models, using simulation, have become available (Brysbaert & Stevens, 2018). In our regression models (see Tables 5 and 6), we therefore report post hoc power based on 1,000 simulations per predictor (Brysbaert & Stevens, 2018), using the *simR* package (Green & MacLeod, 2016).

[3] We included sibling status in our analyses because younger siblings have been found to have an advantage in false-belief understanding (Lewis et al., 1996; Ruffman et al., 1998).

Table 1
Overview of Tests Administered at Time 1 (T1) and Time (T2)

Perspective-Marking Language	Social Cognition	Individual Differences
Mental Verbs: 1st person (T1: predictor, T2: outcome)	False Belief (T1: predictor, T2: outcome)	Short-Term Memory (T1: control)
Mental Verbs: 3rd person (T1: predictor, T2: outcome)	Theory of Mind Precursors (T1: control)	Working Memory (T1: control)
Epistemic Modals (T1: predictor, T2: outcome)	Discarded task:	Verbal Inhibitory Control (T1: control)
Complement-Clause Proficiency (T1: predictor, T2: outcome)	Implicit False Belief (T1: control)	Motoric Inhibitory Control (T1: control)
		Cognitive Flexibility (T1: control)
		Vocabulary (T1: control)
		Grammar (T1: control)

The ages of the children at T2 ranged from 39 to 47 months (M = 42.93, SD = 2.46). The study, "Language and social cognition in preschoolers," received approval from the Faculty of Arts and Social Sciences and Lancaster Management School Research Ethics Committee at Lancaster University (University Research Ethics Committee No. FL16283).

Test Materials

We collected measures of children's perspective-marking language (four tests), social cognition (three tests) and individual differences in executive functioning, memory and general language (seven tests). See Table 1 for an overview.

Perspective-Marking Language

We assessed aspects of proficiency with perspective-marking language with two tasks. The first was designed to capture different levels of abstractness in children's schemas for embedding propositions in complement-clause constructions. The second measured understanding of the certainty distinctions marked by the two mental verbs *think* and *know* (with first- vs. third-person complements) and by the two modal verbs *might* and *must* (in simple clauses).

Complement-Clause Proficiency

The test consists of two parts, complement repetition and complement comprehension (see the Appendix for a list of experimental items and Part 1 of the online supplemental material for details of test design and a full list of training items, experimental items and fillers). Complement repetition uses the elicited-imitation paradigm (Lust et al., 1996). The task for the child is to repeat eight complement-clause constructions, half of which present high-frequency complement-taking verbs (*think*, *know*) in high-frequency strings (e.g., "I think the monkey likes the fruit"), and half of which present high-frequency complement-taking verbs (*say*, *see*, *hope*, *pretend*) in lower-frequency clauses in past tense with proper-noun subjects (e.g., "Jill said that she was very thirsty").

Complement comprehension uses the basic paradigm developed by de Villiers and Pyers (1997) in their Memory for Complements test. The child hears a simple main clause (distractor) and a complement-clause construction and is asked a comprehension question targeting the information in the complement clause and requiring the child to ignore information from the distractor clause, as in the following examples:

> "Pam saw that her dad picked flowers.
> (Complement-clause construction)
> Then she found a vase for them. (Distractor clause)
> What did Pam see?"

> "Nick found a rope, (Distractor clause)
> but he shouted that he found a snake.
> (Complement-clause construction)
> What did Nick shout?"

In contrast with the Memory for Complements test, our task does not reuse the same few complement-taking verbs in all items but presents children with eight different verbs: four that are frequent in complement-clause constructions (*say*, *see*, *hope*, *pretend*) and four that are highly infrequent in complement-clause constructions in child-directed speech (*notice*, *hear*, *shout*, *read*; see the Appendix). Our task also poses lower demands on inhibition by avoiding the use of visual stimuli depicting the wrong answer, and it lowers the demands on working memory by only letting the distractor clause intervene between the complement clause and the question in half of the items, whereas the other half present the question directly after the complement clause.[4]

For the repetition task, the stimuli sentences were prerecorded and played from a laptop. In the comprehension task, the stimuli sentences and questions were read aloud by the experimenter (see Part 1 of the online supplemental material for details on test administration and training procedure). For the test questions, responses were coded following a detailed coding manual, with distinct criteria for the two parts of the test. To capture fine-grained differences in children's mastery, assessment was graded: responses to each of the 16 items could be awarded zero, one or two points (see Part 1 of the online supplemental material for coding details and examples).

Mental Verbs and Epistemic Modals

To test children's understanding of the epistemic uses of mental verbs in complement-clause constructions and modal verbs in simple clauses, we used an adjusted version of the Hidden Objects task (Brandt et al., 2016; Moore et al., 1990). The task for the child is to find a hidden sticker based on the advice from two puppet helpers (a pig and a cow) who offer contrasting statements about its location, as in this example:

> **Mental Verbs: First person**
> Pig: "I **know** that the sticker is in the blue box."
> Cow: "I **think** that the sticker is in the red box."

[4] Children performed significantly better when the distractor sentence came first and did not intervene (p = .02).

On each trial, one statement is marked with a verb indicating relatively higher certainty ("know"), the other with a verb indicating relatively lower certainty ("think"), and success on the forced-choice task requires the child to recognize the different degrees of epistemic strength conveyed by the verbs and trust the more certain speaker. Children received six trials with first-person subjects and six trials with third-person subjects, as in this example:

Mental Verbs: Third person

Boy: "The pig **knows** that the sticker is in the blue box."

Boy: "The cow **thinks** that the sticker is in the red box."

Half of the children received the first-person trials on Day 1 and the third-person trials on Day 2, and vice versa. The third-person statements were delivered by a boy puppet to avoid a difference in speaker authority from the first-person condition where the speakers were puppets. All sentences were prerecorded and played from a laptop. We counterbalanced how often each puppet/color box was correct, and whether the more certain statement came first or last. We followed the standard practice of not letting the children open the chosen boxes before having completed all trials in a set, so that they would not learn from feedback during the trials. To reward all children equally, all boxes contained stickers.

We adapted the task in two ways. First, to avoid the pragmatic confound that the second speaker could introduce doubt over the first speaker's credibility in sequences where the less certain statement follows the more certain statement, the puppets entered the stage one at a time, making sure they did not overhear each other's statements and thus did not explicitly contradict each other. Second, we used boxes of different sizes and shapes (differing between but never within pairs) to enhance speaker credibility by making it more plausible that the helpers could remember the hiding places.

We also gave children six trials contrasting the epistemic modal verbs *might* and *must*, as in the following examples. Both modals are frequently used epistemically in children's input, as confirmed by an analysis of the Manchester corpus (Theakston et al., 2001).

Epistemic Modals

Pig: "The sticker **must** be in the blue box."

Cow: "The sticker **might** be in the red box."

If children showed color, speaker, or order biases (such as always choosing their favorite color, the cow's suggestion, or the last box mentioned), by chance they would get three out of six correct without paying attention to the verbs. To make sure that the task measured genuine understanding of the certainty contrast and that the children were not awarded any points without attending to the verbs, any raw score from zero to three on each of the three measures using the Hidden Objects task (Mental Verbs: First person, Mental Verbs: Third person, Epistemic Modals) was given zero points. Raw scores of four correct out of six were given one point for nascent attention to verb differences, whereas raw scores of five or six correct out of six were awarded two points for stable accuracy of interpretation of the certainty distinction marked by the verbs.

Social Cognition

False belief at T2 was our target dependent variable, assessed with a suite of three tests requiring verbal or pointing answers. At T1, we used the same explicit tests with different contents and also included a measure of Theory of Mind (ToM) precursors, that is, sociocognitive skills expected to represent prerequisite steps in the development of false-belief understanding: the understanding of diverse desires, diverse beliefs, and knowledge access (Wellman & Liu, 2004).[5]

False Belief

Five false belief questions that required children to remember their own previous false belief (two "self" questions) or to predict another's false belief (three "other" questions) across three different tasks were used: Unseen Location Change (Other), Unexpected Contents (Self, Other) and Unexpected Identity (Self, Other). For each task, we developed two versions (A and B) with identical structure but different content. Half of the children received Version A at T1 and Version B at T2 and vice versa. The tasks were adapted from standard false-belief tests, with a series of adjustments to avoid known risks of over- or underestimating children's false-belief understanding.

Language. The most crucial adjustment was to keep measures of false belief and complement clauses distinct by not using complement clauses (e.g., "What will Sarah think is in the box?") in the false-belief tests (see Part 3 of the online supplemental material).

Control Questions Establishing Shared Knowledge About Reality. In all three tasks, control questions were used to check whether the child remembered the real state of affairs (reality control) and, in Unseen Location Change, the original state of affairs (memory control). These control questions were always asked before the false-belief test question to make it clear that the child and the experimenter shared knowledge about the real state of affairs. This reduces the risk that the child will treat the false-belief test question as a genuine request for indirect information about reality (cf. Hansen, 2010).[6] If the child did not respond correctly to a control question, they were shown the real contents or identity again or had the location-change story told again and were asked the question again. If the child could not pass the control question(s) for a task, their response on the target false-belief question was not included in the analysis.

Unseen Location Change. This task tests the child's ability to attribute a false belief about a translocated object's location to another character who has not witnessed its transfer. The test has a narrative format where the child sees and hears the experimenter acting out a short story with dolls and props, and follows the basic structure of the standard location change tasks developed by Wimmer and Perner (1983; the Maxi task) and Baron-Cohen et al. (1985; the Sally-Anne task), with the following small adjustments:

Explicit motive for the transfer: We present the location change in a plausible everyday story where the child does not have to infer the

[5] We also aimed to include a measure of implicit false belief (using eye-tracking to measure anticipatory looking with the stimuli from Southgate et al., 2007), but we did not obtain enough valid data points on this test to justify including it in the longitudinal analysis and therefore discarded this measure. See Part 2 of the online supplemental material for further detail.

[6] We acknowledge that asking the control question first could affect performance in Unseen Location Change negatively by drawing attention to the object's actual location (Rubio-Fernández & Geurts, 2016). Future research will have to investigate the trade-off between establishing common ground and drawing attention to the actual location by asking the control question first.

second agent's motive for the transfer, which is practical and common (A: putting a teddy in a pram to prepare the pram for the child's nap; B: putting a carrot in a pot to cook dinner), and it is explicated verbally.

Test question with *first*: Following Siegal and Beattie (1991), we asked where the protagonist will look "first" for the moved object. This makes it clear to the child that s/he is not asked where the protagonist will eventually find the object.

Keeping the protagonist in sight: We facilitated perspective tracking by keeping the protagonist in the child's visual field throughout the story, following Rubio-Fernández and Geurts (2013). The protagonist remained on the table with their back to the transfer scene on the other side of a wall barrier, and we verbally made sure that the child was aware that the protagonist could not see the central scene.

Unexpected Contents. Building on Hogrefe et al. (1986) and Perner et al. (1987), this task tests the child's ability to predict another's false belief and remember their own previous false belief about a familiar container's contents after having obtained new knowledge about its real unexpected contents (Version A: ball in raisin box; Version B: spoon in crayon box).

To avoid the risk that children should use a superficial "memory for own utterance strategy" for answering the self question (cf. Williams & Happé, 2009), rather than beginning the test by asking children to verbalize their beliefs about the contents of the box, we checked their expectations in an indirect way by asking the child to either help put some raisins into a bowl or find a crayon to draw with.

Unexpected Identity. Building on Gopnik and Astington (1988), this task tests the child's ability to predict another's false belief and remember their own previous false belief about the identity of a deceptive object (Version A: flower-shaped pen; Version B: apple-shaped candle), after having discovered its real identity.

ToM Precursors

Studies using the ToM scales indicate that children's false-belief understanding builds on a more basic understanding of mental states (Wellman et al., 2011; Wellman & Liu, 2004). To assess whether differences in prerequisite ToM skills would explain variability in False Belief performance at T2, we gave children three tasks from the scales at T1, Diverse Desires, Diverse Beliefs, and Knowledge Access (further information about false belief and ToM measures can be found in Parts 3 and 4 of the online supplemental material).

Individual Differences

To test the uniqueness of a potential relationship between complement clauses and false belief, we measured a range of executive-functioning, memory, and general-language control variables at T1.

Cognitive Flexibility

Cognitive flexibility was tested with the Dimensional Change Card Sort (Zelazo, 2006), which assesses children's ability to switch flexibly between conflicting rules for sorting.

Inhibitory Control

This was tested with two conflict tasks, one motoric and one verbal. We tested ability to selectively suppress *motoric* responses with the Bear/Dragon (Sheep/Crocodile) task (Carlson et al., 2004; Kochanska et al., 1996), where the child hears commands for simple physical actions such as "Touch your ears" from two puppets and has to ignore commands from one puppet while following commands from the other puppet. Ability to suppress a dominant *verbal* response and initiate a conflicting response was tested with the Black/White Stroop (Vendetti et al., 2015), which requires the child to say "black" in response to white cards and "white" in response to black cards in 21 intermixed trials.

Short-Term Memory

This ability was tested with a digit-span task, Recall of Digits Forward from the *British Ability Scales II* (Elliott, 1996), which requires the child to repeat strings of ordered digits with items of increasing length.

Working Memory

This was tested with the Missing Scan Task (Roman et al., 2014), which measures the ability to retain and manipulate information in young preschoolers. The child sees a set of animal figurines, who are then hidden and presented again with one animal missing. The task for the child is to report the missing animal, which requires scanning and retrieving the contents of immediate memory, and the sets increase in size.

Vocabulary

Receptive vocabulary was tested with the *British Picture Vocabulary Scale III* (Dunn et al., 2009), a picture-selection task requiring the child to identify the picture depicting the word read by the experimenter on a page with three distractor pictures.

Grammar

Receptive grammar was tested with the subtest Sentence Structure from the *Clinical Evaluation of Language Fundamentals Preschool 2* (Wiig et al., 2006), a picture-selection task requiring the child to identify the picture representing the sentence read by the experimenter on a page with three distractor pictures.

Procedure and Principles for Task Ordering

The study was conducted in three separate sessions: two sessions at T1 and one session at T2. Each session lasted up to an hour, including warm-up play. On average, the second T1 session fell 9 days (SD = 3.9) after the first, and the T2 session occurred, on average, 6 months (181 days; SD = 13.7) after the second T1 session. The children were tested individually in a laboratory room with limited visual and auditory distraction by either the first author or the third author (half each). Caregivers accompanied their children during testing but were informed not to help their child in any way. The tasks were administered in a fixed order (see Table 2).

Interrater Reliability

We calculated interrater reliability for the False Belief tests, the Hidden Objects task, and the Complement-Clause Proficiency measure. Responses from 10 of the 48 children (21%) were coded (and transcribed) by a second rater. There was no disagreement about responses in the False Belief tests or in the Hidden Objects task.

Complement-Clause Proficiency had both a transcription and a coding component. For transcription reliability, the two transcribers agreed on whether the word had been produced in 95% of the cases

Table 2
Overview of Task Ordering Within and Across Sessions at Time 1 (T1) and Time 2 (T2)

Tasks	Set A	Set B
SESSION 1 (T1)		
Implicit False Belief	FB1	FB2
Working Memory	Missing Scan Task	Missing Scan Task
False Belief	Unexpected Contents: Ball in raisin box	Unexpected Identity: Apple-candle
False Belief		Unseen Location Change: Carrot story
Short-Term Memory	Forward Digit Span	Forward Digit Span
Complement-Clause Proficiency	Complement Repetition: A	Complement Repetition: B
Mental Verbs	Hidden Objects: 1st Person	Hidden Objects: 3rd Person
Epistemic Modals	Hidden Objects: Epistemic Modals	
Verbal Inhibitory Control	Black/White	Black/White
Receptive Vocabulary	British Picture Vocabulary Scale	British Picture Vocabulary Scale
SESSION 2 (T1)		
Implicit False Belief	FB2	FB1
False Belief	Unexpected Identity: Flower-pen	Unexpected Contents: Spoon in crayon box
False Belief	Unseen Location Change: Teddy story	
Complement-Clause Proficiency	Complement Comprehension: A	Complement Comprehension: B
Mental Verbs	Hidden Objects: 3rd Person	Hidden Objects: 1st Person
Epistemic Modals		Hidden Objects: Epistemic Modals
Cognitive Flexibility	Dimensional Change Card Sort	Dimensional Change Card Sort
ToM Precursors	Diverse Desires	Diverse Desires
ToM Precursors	Diverse Beliefs	Diverse Beliefs
ToM Precursors	Knowledge Access	Knowledge Access
Receptive Grammar	CELF: Sentence Structure	CELF: Sentence Structure
Motoric Inhibitory Control	Sheep/Crocodile	Sheep/Crocodile
SESSION 3 (T2)		
False Belief	Unexpected Contents: Spoon in crayon box	Unexpected Identity: Flower-pen
False Belief	Unseen Location Change: Carrot story	Unseen Location Change: Teddy story
Complement-Clause Proficiency	Complement Repetition: B	Complement Repetition: A
False Belief	Unexpected Identity: Apple-candle	Unexpected Contents: Ball in raisin box
Mental Verbs	Hidden Objects: 1st Person	Hidden Objects: 3rd Person
Epistemic Modals	Hidden Objects: Epistemic Modals	
Complement-Clause Proficiency	Complement Comprehension: B	Complement Comprehension: A
Mental Verbs	Hidden Objects: 3rd Person	Hidden Objects: 1st Person
Epistemic Modals		Hidden Objects: Epistemic Modals

Note. Half of the children received Set A, the other half Set B. CELF = Clinical Evaluation of Language Fundamentals Preschool–2; FB1 = Implicit False Belief video 1; FB2 = Implicit False Belief video 2; ToM = Theory of Mind.

(750 words out of 793), and for only 10 responses, word-transcription disagreement would affect overall response coding: for nine responses, resulting in a one-point difference; for one response, resulting in a two-point difference (Cohen's $\kappa = .927$; weights: squared; responses: 160; raters: 2; $z = 11.7$). All responses with transcription disagreement were transcribed by a third rater and the final decision discussed and made collectively. For coding reliability, a weighted kappa statistic considering the ordered character of the ratings showed that agreement was near perfect ($\kappa = .988$; weights: squared; utterances: 160; raters: 2; $z = 12.5$). The few responses with coding disagreement were coded by a third rater, and the final decision discussed and made collectively.

Results

Our central aim was to test whether T1 performance with complement-clause constructions and mental verbs would predict variance in T2 performance with False Belief 6 months later when T1 performance with False Belief as well as individual differences in ToM precursors, executive functioning and general language were controlled. We address this focal question by means of a generalized linear mixed-effects model analysis. We also examine whether T1 performance with False Belief predicts variance in T2 performance with complement-clause constructions.

As a background for the mixed-effects analyses, we first give an overview of performance and correlations between measures at T1 and of overall group-level progression on our central False Belief measure from T1 to T2. All analyses were run on the focal sample of 45 children who returned at T2, that is, excluding the three children that participated only in T1 testing.

Preliminary Analyses: T1 Descriptive Statistics and Correlations

Table 3 summarizes T1 performance on our test battery, and below, we discuss a few issues with missing data as well as the extent to which the tasks captured individual variation, to bear in mind when we evaluate the role of these T1 predictors in explaining variance in false-belief reasoning at T2.

First, for the Sheep/Crocodile task, measuring Motoric Inhibitory Control, we had to discard data from 24% of the children (11 out of 45) because they did not meet the inclusion criteria. Either they did not pass the practice trials (i.e., they never

Table 3
Means, Standard Deviations, Ranges, and Missing Data Points for Time 1 (T1) Predictors

T1 measures	*M*	*SD*	Range	Maximum score	Missing data points
Social cognition					
ToM Precursors (including Knowledge Access)	2.52	0.79	0–3	3	22 of 45 (49%)
ToM Precursors (without Knowledge Access)	1.56	0.62	0–2	2	0
False Belief	2.00	1.49	0–4	5	17 of 225 (8%)
Executive functioning and memory					
Short-Term Memory	7.47	3.42	0–15	36	0
Working Memory	3.27	1.32	1–6	6	0
Cognitive Flexibility	1.88	2.57	0–6	6	3 of 45 (7%)
Motoric Inhibitory Control	4.00	1.60	0–5	5	11 of 45 (24%)
Verbal Inhibitory Control	8.77	5.64	0–21	21	1 of 45 (2%)
General language					
Receptive Vocabulary	42.76	14.12	17–77	120	0
Receptive Grammar	9.58	4.48	1–18	22	0
Perspective-marking language					
Complement-Clause Proficiency	9.45	6.60	1–26	32	1 of 45 (2%)
Mental Verbs: 1st person	0.47	0.66	0–2	2	0
Mental Verbs: 3rd person	0.47	0.73	0–2	2	0
Epistemic Modals	0.40	0.58	0–2	2	0

Note. ToM = Theory of Mind.

learned the rule of ignoring the crocodile) or the sheep-command control trials (i.e., they ignored both the crocodile and the sheep, instead of selectively suppressing response to the crocodile). For the 34 children included there was a ceiling effect (M = 4.0, SD = 1.6; maximum score = 5), indicating a reduced capacity for capturing individual variation.

A similar problem with missing data points pertained to our ToM Precursors measure. Half of the sample (22 of 45 children) failed the control question for Knowledge Access, indicating that they did not understand the task and invalidating their test responses for this subtask. We therefore excluded scores from Knowledge Access for all children and only included scores from Diverse Desires and Diverse Beliefs in our aggregate ToM score (range = 0–2) which allowed us to examine the role of ToM Precursors in the full sample since there were no problems with failed control questions on these subtasks. However, the children generally performed very well on this measure (M = 1.56, SD = .62; maximum score = 2), indicating that the ToM Precursors tasks had limited capacity for capturing individual variation in social cognition at this age.

For Cognitive Flexibility (measured with the Dimensional Change Card Sort), children generally performed poorly (M = 1.88, SD = 2.57; maximum score = 6), in line with previous findings from children below 4 years of age (Zelazo, 2006).

For each of the three verb-knowledge measures derived from the Hidden Objects task - Mental Verbs: First person (M = .47, SD = .66; maximum score: 2); Mental Verbs: Third person (M = .47, SD = .73; maximum score: 2); Epistemic Modals (M = .40, SD = .58; maximum score: 2) - there was also limited variability. At T1, the majority of children showed no signs of attending to the certainty distinctions marked by the verb pairs, but consistently based their choices on either a specific color, a specific puppet helper or the first or last box mentioned, thus getting a score of 0.

Table 4 presents an overview over concurrent correlations between our T1 predictors. For False Belief, the only significant correlation is with Short-Term Memory (rs = .40, p = .003). The absence of other correlations is somewhat surprising since other studies have repeatedly found concurrent correlations between false-belief reasoning, language proficiency and executive functioning, but can likely be explained by the younger age group we were targeting at T1 (age range = 2 years, 9 months–3 years, 5 months).

Our central predictor variable, Complement-Clause Proficiency, was significantly associated with many of our control measures and had strong positive correlations with Age (rs = .55, p = .0001), Receptive Vocabulary (rs = .60, $p < .0001$), Verbal Inhibitory Control (rs = .51, p = .0005), and intermediate-strength positive correlations with Receptive Grammar (rs = .33, p = .03) and Short-Term Memory (rs = .44, p = .003). These correlations with both general language, memory and inhibitory control confirm the relevance of including such control measures when examining the developmental relationship between complement clauses and false belief.

For our three measures for understanding the certainty-marking function of mental verbs and epistemic modals, we found no significant associations with any other measures, likely due to the generally poor performance on the Hidden Objects test, yielding little variation in scores.

False Belief Progression From T1 to T2

At T1 children on average got two out of five false-belief questions correct (M = 2.00, SD = 1.49; range = 0–4). Their performance approached falling significantly below chance (one-sample Wilcoxon signed-ranks test: V = 229, p = .061), suggesting that their responses were not random, but biased toward the real state of affairs rather than the belief state targeted by the false-belief question.

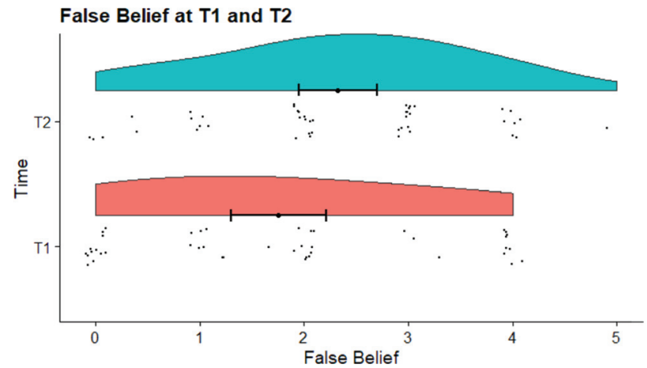
At T2, children on average responded correctly to half of the five questions (M = 2.6, SD = 1.03; range = 1–5). As a group, their performance was not above chance (one-sample Wilcoxon

Table 4
Spearman's Correlations Between T1 Predictors

T1 measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. False Belief														
2. ToM (incl. Knowledge Access)	.20													
3. ToM	.05	.99***												
4. Age	.29†	.12	.04											
5. Short-Term Memory	.40*	.23	.20	.34*										
6. Working Memory	−.22	.13	.20	.07	.04									
7. Cognitive Flexibility	.20	−.03	.16	.26†	.23	.05								
8. Motoric Inhibitory Control	.08	.10	.07	.18	.39*	−.11	.18							
9. Verbal Inhibitory Control	.16	.07	.12	.50***	.47**	−.01	.23	.51**						
10. Receptive Vocabulary	.25	.23	.16	.47**	.47**	−.16	.23	.56***	.54***					
11. Receptive Grammar	.18	.08	.11	.31*	.48***	−.21	−.03	.52**	.30*	.46**				
12. Complement-Clause Proficiency	.27	.17	−.02	.55***	.44**	−.02	.23	.33†	.51***	.60***	.33*			
13. Mental Verbs: 1st person	.19	−.08	.13	.21	.25	−.22	.12	−.01	.04	−.06	.23	.05		
14. Mental Verbs: 3rd person	.15	.15	−.08	.10	.11	.08	−.05	−.01	.16	.00	.11	.18	.11	
15. Epistemic Modals	−.27	−.21	−.20	−.03	−.08	.17	−.16	.28	−.02	.24	.04	.17	.02	−.05

Note. ToM = Theory of Mind.
† $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 1
Raincloud Plot Showing Correct Responses on the Five False-Belief Questions



Note. Raw scores plotted (jittered) under the probability densities with means and 95% confidence intervals. Plot produced in R following Allen et al. (2019). Time 1 = T1; Time 2 = T2. See the online article for the color version of this figure.

signed-ranks test: $V = 446.5$, $p = .615$), but it was significantly better than 6 months earlier (Wilcoxon paired signed-rank test: $V = 73$, $p = .046$; see Figure 1).

Predictors of False Belief at T2

Our main question was whether performance with complement-clause constructions and with mental verbs in complement-clause constructions at T1 (*M* age = 3;1) would explain variance in performance with false belief at T2 (6 months later) while controlling for initial levels of false-belief reasoning and ToM precursors as well as individual differences in executive functioning and general language. To evaluate this question, we fitted a generalized linear mixed effects model to the dependent variable False Belief at T2 in the statistical environment R (Version 3.6.0; R Core Team, 2019), using the package *lme4* (Bates et al., 2015), including random effects of participant, item (i.e., false-belief question), experimenter, set version and set order and starting from the full model including all our control and explanatory variables, which apart from our T1 test measures included age (in months), gender, and sibling status (first child vs. younger sibling). We then fitted the model following the principle of backward selection and a significance-based approach, testing step by step (comparing models by analysis of variance) whether discarding the least significant predictor from the model would significantly decrease the goodness of fit of the model (Gries, 2013). Table 5 summarizes the final model, and Figure 2 represents this model visually, plotting the marginal effects of the predictors on False Belief at T2.

The mixed-effects regression analysis confirmed that children's proficiency with complement clauses at age 2 years, 9 months to 3 years, 5 months was a positive predictor of their likelihood of passing false-belief tests 6 months later. In addition, the regression analysis also demonstrated a unique contribution from children's understanding of the certainty differences marked by the mental verbs most frequently

Table 5

Summary of Generalized Linear Mixed Effects Model Fitted to the Dependent Variable T2 False Belief

Random effects

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	<0.0001	0.0003
Item	(Intercept)	0.2367	0.4865
Experimenter	(Intercept)	<0.0001	<0.0001
Set	(Intercept)	<0.0001	<0.0001
Order	(Intercept)	<0.0001	<0.0001

Fixed effects

	Estimate	*SE*	*z* value	Pr(>|*z*|)	Power [95% CI]
(Intercept)	−1.00	0.37	−2.698	.0070**	
False Belief	0.71	0.32	2.213	.0269*	.62 [.58, .65]
Complement-Clause Proficiency	0.05	0.02	2.144	.0321*	.59 [.56, .62]
Mental Verbs: 3rd Person	0.45	0.22	2.067	.0388*	.53 [.50, .56]

Note. Number of observations = 200; number of items = 5; number of experimenters = 2; number of sets = 2; number of orders = 2. The number of participants in the final model is 44 because one participant did not respond to the Complement Clause Proficiency task at T1.
* $p < .05$. ** $p < .01$.

used in complement-clause constructions: *think* and *know*.[7] However, only children's performance with mental verbs in third-person complements (e.g., "He knows") explained variance in later false-belief reasoning, while no relationship emerged for mental verbs in first-person complements (e.g., "I know").

Predictors of Complement-Clause Proficiency at T2

Having found proficiency with complement-clause constructions to predict false-belief reasoning 6 months later, we wanted to examine whether this predictive relationship was bidirectional. If so, we would expect false-belief performance at T1 to predict variation in complement-clause proficiency at T2, and to evaluate this question, we fitted a new series of linear mixed-effects models in R to the dependent variable T2 Complement-Clause Proficiency. Using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), we followed the same strategy of stepwise backward selection as we did for the model predicting False Belief at T2. Table 6 summarizes the final model.

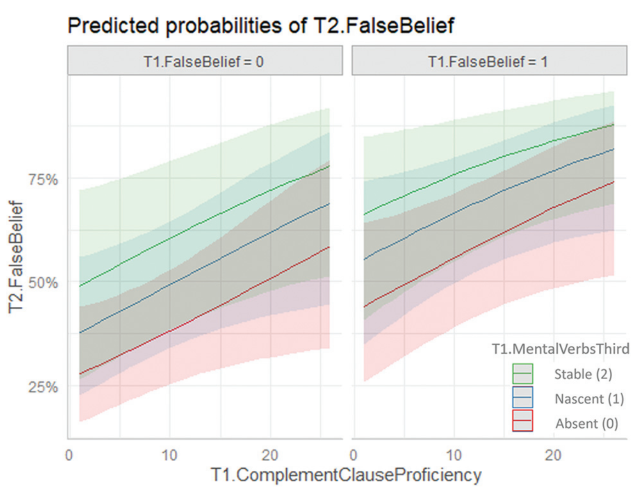
Table 6 demonstrates continuity in Complement-Clause Proficiency from T1 to T2, with T1 performance being a significant predictor of T2 performance. Crucially, however, False Belief at T1 also contributed uniquely to explaining variance in T2 Complement-Clause Proficiency (illustrated in Figure 3). This result suggests a bidirectional relationship where developments in attention to mental states support developments in children's gradual construction of abstract schemas for complement-clause constructions, as well as the other way around.

Among the other significant predictors, it is worth noting that children's performance with Mental Verbs: First person in complement-clause constructions at T1 was also a significant independent predictor of Complement-Clause Proficiency 6 months later (see Figure 3).

We also fitted models to each of our three T2 measures of verb knowledge (Mental Verbs: First person; Mental Verbs: Third person; and Epistemic Modals) to check whether T1 False Belief was a significant predictor of later understanding of the certainty-marking function of these verbs. T1 False Belief did not emerge as a significant predictor in any of these models.

Figure 2

Marginal Effects Plot Showing the Effects of the Three Predictors: False Belief, Complement-Clause Proficiency, and Mental Verbs



Note. The plot is generated from the *glmer* model using the *ggeffects* package (Lüdecke, 2018) in R. The plot is divided into panels to illustrate the advantage of passing (right panel) over failing (left panel) a false-belief question at Time 1 (T1) and for passing the corresponding false-belief question at Time 2 (T2). See the online article for the color version of this figure.

[7] In principle, the influence from mental-verb knowledge could be explained as a superficial task factor. Whereas our false-belief questions avoided complement clauses, two out of five questions did use the verb *think*, in the representational-change "self" questions (e.g., "What did you first think about this box? Did you think about raisins or a ball first?"). To check whether performance on these two self questions could be driving the correlation, we tested interactions between Mental Verbs and Presence of the verb *think*. This follow-up analysis showed no difference in the role of mental-verb knowledge for false-belief questions with and without *think*, thus confirming that the effect of mental-verb knowledge is genuine and not a task factor.

Table 6
Summary of Linear Mixed Effects Model Fitted to the Dependent Variable T2 Complement-Clause Proficiency: Final Model

Random effects

Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	0.0385	0.1961
Item	(Intercept)	0.0843	0.2904
Experimenter	(Intercept)	<0.0001	<0.0001
Set	(Intercept)	<0.0001	0.0001
Order	(Intercept)	<0.0001	0.0001
Residual		0.4347	0.6593

Fixed effects

	Estimate	*SE*	*df*	*t* value	Pr(>|*t*|)	Power [95% CI]
(Intercept)	−0.44	0.21	42.86	−2.126	.0393*	
False Belief	0.07	0.03	35.20	2.414	.0211*	.67 [.64, .70]
Short-Term Memory	0.03	0.02	35.75	1.853	.0722†	.46 [.43, .49]
Working Memory	0.06	0.03	35.49	1.851	.0725†	.47 [.44, .50]
Receptive Vocabulary	0.02	0.00	36.35	4.634	<.0001***	.99 [.98, 1.0]
Complement-Clause Proficiency	0.12	0.04	641.02	2.904	.0038**	.83 [.81, .86]
Mental Verbs: 1st person	0.27	0.08	35.33	3.609	.0009***	.93 [.92, .95]
Epistemic Modals	−0.27	0.08	34.90	−3.395	.0017**	.91 [.89, .92]

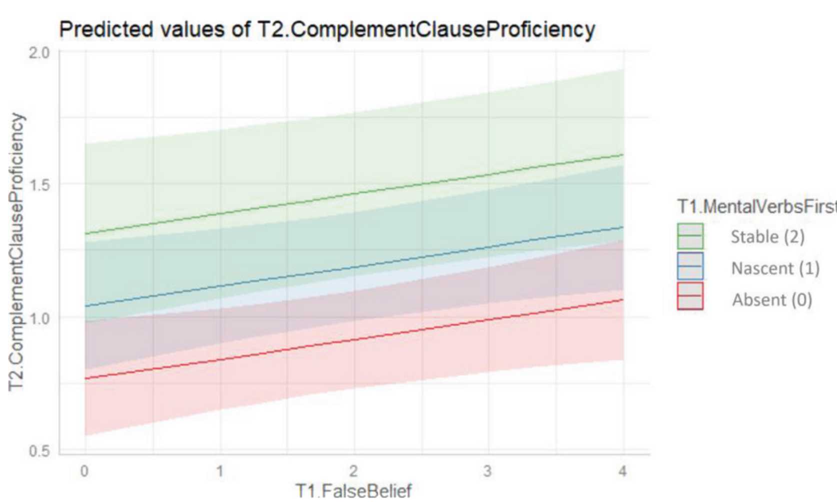
Note Number of observations = 680; number of items = 16; number of experimenters = 2; number of sets = 2; number of orders = 2.
The number of participants in the final model is 43 because one participant did not respond to the Complement-Clause Proficiency task at T1, and another did not respond to the Complement-Clause Proficiency task at T2.
† $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The main aim of this longitudinal study was to investigate whether acquisition of perspective-marking language facilitates development in children's social cognition. Specifically, we examined whether children's proficiency with complement-clause constructions around the age of 3 years would explain variance in their false-belief reasoning 6 months later. Our results showed that this was indeed the case. Both children's performance with false belief and with complement clauses at T1 (*M* = 3 years, 1 month) contributed independently to predicting false-belief performance at T2 (*M* = 3 years, 7 months). However, the relationship appears to be bidirectional, with advances in belief-reasoning also supporting developments in children's gradual acquisition of complement-clause constructions.

Two aspects of complementation, structural flexibility in children's schemas for complement-clause constructions as well as

Figure 3
Marginal Effects Plot Showing the Effects of the Three Predictors: False Belief, Complement-Clause Proficiency, and Mental Verbs



Note. The plot shows the effects of the Time 1 (T1) predictors False Belief and Mental Verbs: First person on Time 2 (T2) Complement-Clause Proficiency while holding the other predictors in the model constant. The plot is generated from the *lmer* model using the *ggeffects* package (Lüdecke, 2018) in R. See the online article for the color version of this figure.

understanding of mental verbs used in complement-clause constructions both contributed uniquely to predicting later false-belief reasoning. We discuss the complementary roles of these two aspects of complementation in more detail below.

The current study provides a more stringent test of the developmental relationship between complement-clause acquisition and false-belief development than usually seen, as it avoided the widespread confound of using complement clauses in the false-belief tests. Further, it included a series of executive-functioning, memory and general-language measures, allowing us to exclude with some confidence the possibility that the relationship between complement clauses and false belief could simply be explained by success on both tasks depending on the same executive-functioning, memory or linguistic skills. We return to the lack of effects of our executive-functioning and memory measures below.

The Role of a General Schema for Complement-Clause Constructions

Why would children who were better at understanding and repeating complement-clause constructions like "Ann said [that she was very hungry]" around the age of 3 years be more likely to answer false-belief questions correctly 6 months later? To answer this question, it is useful to look at the basic communicative function of complement-clause constructions. The fact that this type of syntax has evolved in the majority of languages of the world (Dixon, 2006; Noonan, 2007) indicates that this construction serves a fundamental function in human communication (cf. Moore, 2020). According to cognitive linguistics, the central function of complement-clause constructions is to provide language users with a flexible format for explicating the relationship between ideas and the persons these ideas belong to, that is, to anchor ideas about the world in human conceptualizers (Verhagen, 2005). Whether we say, "I think it's in the cupboard" or "Mum says it's in the cupboard," we are specifying that the idea that something is in the cupboard is not just an objective fact about the world, but an idea that belongs to a person with a specific attitude toward and perspective on that idea. In this way, complement-clause constructions are a specialized grammatical means for talking about invisible mental states and to highlight conflicts between different mental states as well as between mental states and reality.

Complement-clause constructions are certainly not the only linguistic way to communicate about relationships between persons and ideas or to highlight contrasts between ideas and reality, but as part of the *grammar* of a language, they offer children constantly recurring evidence for the communicative relevance of specifying mental aspects of situations, that is, whom ideas belong to and how. Crosslinguistic experiments in the thinking-for-speaking tradition (Slobin, 1996) have provided evidence that when a grammatical phenomenon in a language requires attention to certain aspects of situations, language users are likely to monitor and store such aspects of experience routinely so that they will be easy to communicate about later (Bowerman & Choi, 2003; Okuno et al., 2020). In the same way, growing up in a linguistic community where complement clauses are part of the grammar may train children to pay attention to and keep track of relationships between persons and ideas, and such routine monitoring of mental aspects of experience may support children's developing abilities to represent and reason about beliefs.

The Role of Mental Verbs in Complement-Clause Constructions

To examine whether not only the complement-clause construction itself, but also familiarity with the specific mental verbs frequently used in the construction supports developments in mental-state reasoning, we tested children's ability to differentiate between *think* and *know*. The results confirmed a unique contribution from mental-verb knowledge (measured with the Hidden Objects task) in addition to proficiency with the complement construction itself (measured with Complement-Clause Proficiency). Our results thus confirm the relationship between mental-verb knowledge and false belief found by Moore et al. (1990) and Brandt et al. (2016) using false-belief tests without complements and substantiate their hypothesis of a causal influence from mental-verb acquisition on mental-state reasoning with longitudinal data. When proficiency with *think* and *know* in complement-clause constructions predicts later false-belief performance, a likely explanation is that hearing these mental verbs in situations with different degrees of certainty invites children to compare and generalize over these situations, thus supporting them in building up stable concepts of knowledge and belief.

Further, the facilitatory role of mental verbs with complements appears to be superior to another type of verbal epistemic marking, that is, epistemic modal verbs. Children's performance with the two high-frequency modals *might* and *must* in the same Hidden Objects task at T1 did not explain any variance in false-belief performance at T2. A straightforward explanation of this lack of effect was children's poor performance with modals at T1. Our results do not exclude the possibility that epistemic modals help consolidate categorization of mental states of certainty and uncertainty at older ages, but any potential support would appear to be secondary to the one provided by mental verbs with complements. A likely explanation for the privileged role of the latter is that they explicitly anchor the proposition presented in the complement clause in a conceptualizer (the subject of the complement-taking verb; e.g., "**The cow** thinks that the sticker is in the red box"), whereas the epistemic modals leave the presence of a certain versus uncertain conceptualizer implicit (e.g., "The sticker might be in the red box").

While our results establish a special role for mental verbs, they also suggest a possible interplay with the specific types of complement-clause constructions they occur in. Whereas children's sensitivity to the contrasting degrees of certainty marked by mental verbs in third-person complements (e.g., "**the cow** thinks") did explain variance in later mental-state reasoning, their sensitivity to the same contrast in first-person complements (e.g., "**I** think") did not (for similar results, see Brandt et al., 2016). As with the epistemic modals, children were generally performing poorly with first-person complements at T1, with only 4 out of 45 children (9%) consistently differentiating between "I think" and "I know," in line with previous findings of chance performance in 3-year-olds (Brandt et al., 2016; Moore et al., 1990).

Why were the 2- to 3-year-olds performing so poorly with the first-person complements? The key reason is probably that the

Hidden Objects test measures sensitivity to *certainty* contrasts. This makes it a useful task for testing children's differentiation of expressions with a clear certainty contrast (e.g., "he thinks/knows"; *might/must*), but it is less clear what it can tell us about children's understanding of "I think" and "I know," because, as demonstrated by Howard et al. (2008), children very often hear "I think" in situations where the speaker is absolutely certain. In real life, children may depend on contextual cues to discern whether the first-person complement is used for indicating certainty or uncertainty.

The use of "I think" in certainty contexts makes it likely that present-tense first-person complements with *think* and *know* provide children with less reliable cues for carving out distinct categories of belief and knowledge than third-person complements. Nevertheless, our finding that T1 Mental Verbs: First person was a significant predictor of T2 Complement-Clause Proficiency suggests that children do not simply process constructions with "I think" as adverb-like markers autonomous from the complement-clause construction. Because the Hidden Objects task may not be the most appropriate test for capturing proficiency with first-person complements, we have examined the role of first-person complements in supporting false-belief reasoning directly in a new training study (Boeg Thomsen, Kandemirci, et al., 2021). Here, children who had mental-state-rich experience mediated with first-person complements advanced significantly more in false-belief reasoning than children trained with simple clauses, supporting the interpretation that the lack of relationship with false-belief reasoning in the current study may be a task effect.

Executive Functioning and Memory

None of our T1 executive-functioning or memory measures emerged as significant predictors of later false-belief reasoning. While we had reduced ability to detect the potential influence of motoric inhibitory control due to missing data, this was not the case for our other executive function and memory measures, and the lack of a predictive relationship is somewhat surprising since previous longitudinal studies with English-speaking children found earlier executive-functioning skills to predict later ToM performance (Carlson et al., 2004; Hughes & Ensor, 2007). Part of the explanation may be that these studies used an aggregate ToM measure, for when Carlson et al. (2004) tested the influence of T1 executive functioning on T2 false belief specifically, they did not find a significant predictive relationship. Further, age may play a role, as we targeted children who were older at T1 than the 2-year-olds in these two studies. Thus, differences in executive functioning could play their prime role in sociocognitive development at an earlier stage, and indeed, Hughes and Ensor (2007) did *not* find any significant effect of executive functioning at their T2 (38.5 months) on ToM at T3 (T3; 50.5 months).

Results from correlational studies with older children are also mixed: Burnel et al. (2020) found that complement clauses, flexibility and working memory explain concurrent variance in ToM tasks in French-speaking 3- to 11-year-olds. On the other hand, Durrleman and Franck (2015) did not find correlations between either inhibition or flexibility and verbal false-belief tasks in French-speaking children with autism (6 to 16 years) or typical development (4 to 9 years), and neither did de Villiers and de Villiers (2012) in English-speaking deaf (4 to 7 years) or hearing (3 to 5 years) children. Although the current study shows that the relationship between complement-clause proficiency and false-belief reasoning does not simply depend on shared executive-function requirements, it leaves other questions open regarding the role of executive functioning in ToM development.

The Role of False-Belief Reasoning in Complement-Clause Acquisition

This study examined longitudinal relationships between the acquisition of perspective-marking language and sociocognitive development in children around age 3 and we found evidence for both directions of influence. This result aligns with the findings in Milligan et al. (2007), who conducted a meta-analysis of longitudinal relationships between language ability and false-belief performance and found significant effect sizes for both directions of influence. It also aligns with current proposals for a bidirectional relation between language and ToM in ontogenetic and phylogenetic development. For example, Rubio-Fernández (2020) argued that children's acquisition of deictic markers both depends on and also supports their understanding of others' perspectives, and based on computational modeling, Woensdregt et al. (2020) suggested a bidirectional evolutionary relationship between word learning and sociocognitive development.

On the other hand, the result contrasts with the main conclusion in de Villiers and Pyers' (2002) longitudinal study where early complement-clause proficiency was found to predict later false-belief performance, but not vice versa. Whereas de Villiers and Pyers took this finding to support the hypothesis that false-belief representation depends on complement-clause syntax, it is worth keeping in mind that their complement-clause test was not designed to capture fine-grained differences, making it a less sensitive tool for evaluating longitudinal relationships. Further, the highlighted unidirectional relationship is only found from T2 (age range = 3 years, 5 months–4 years, 2 months) to T3 (age range = 3 years, 9 months–4 years, 6 months), where the children are generally older than in our study. De Villiers and Pyers did not report any general analyses of predictive relationships from T1 (age range = 3 years, 1 month–3 years, 10 months) to T2 (age range = 3 years, 5 months–4 years, 2 months), but they do analyze the relationship between one location-change false-belief task and complement clauses, and here they do find a bidirectional relationship.

Thus, in both de Villiers and Pyers (2002) and the current study, we see evidence of developments in complement-clause acquisition and social cognition supporting each other mutually in 3-year-olds. If complement-clause mastery and false-belief understanding are thought of as absolute abilities that a child either has or has not acquired, such an interplay might seem implausible, and it becomes important to determine which of the two steps is a prerequisite for the other. On the other hand, if both developmental processes are gradual and stretched out over a long time—as demonstrated for complement-clause acquisition by Brandt et al. (2016) and Diessel and Tomasello (2001) and for false-belief development by Amsterlaw and Wellman (2006)—the two processes would have plenty of time to affect and interact with each other.

The preceding sections have focused on the ways in which children's increasing experience with complement clauses and mental verbs is likely to promote development in their representation of mental states. As for the reverse direction of influence, it is highly likely that having a nascent understanding of mental states would both be a strong incentive for children to acquire linguistic tools for communicating about these mental phenomena in increasingly nuanced ways and make it easier for them to discern the function of the complement clauses they encounter in their input (for a similar argument about such an interplay in language evolution, see Moore, 2020).

Future Directions

Although this study provides strong evidence that children's proficiency with complement-clause constructions and mental verbs around age 3 contributes to explaining differences in their false-belief reasoning 6 months later, it also leaves some questions open.

First, even though finding a longitudinal relationship supports the hypothesis of a causal influence from acquisition of perspective-marking language on sociocognitive development, training studies can yield clearer evidence of causation. Whereas previous training studies have addressed this question (Gola, 2012; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Mo et al., 2014), their conclusions are uncertain because they used complement clauses in their posttest false-belief questions and did not control for differences in executive functioning (but see Durrleman et al. (2019) for promising training effects in different diagnostic groups). We have therefore also conducted a new training study with English-speaking 3-year-olds, using the same battery of false-belief and complement-clause tests as the current study (Boeg Thomsen, Kandemirci, et al., 2021).

Second, the current study showed a clear difference between mental verbs occurring with first-person versus third-person complements, with only third-person mental verbs significantly predicting later false-belief reasoning. Although this result aligns with previous findings (Brandt et al., 2016; Gola, 2012), the Hidden Objects test may not be ideal for assessing understanding of first-person complements, and in our new training study, we therefore compare the influence of being exposed to first- versus third-person complements directly. For mental verbs, we should also emphasize that we only tested them in complement-clause constructions, not in simple clauses (e.g., "The cow thinks so")—because our aim was to discern whether mental-verb knowledge was aiding false-belief reasoning *in addition to* syntactic proficiency with the construction. This focus makes it impossible to conclude anything about mental verbs in general, and new studies are needed to evaluate whether familiarity with mental verbs in simple clauses shows the same predictive relationship with later false-belief reasoning.

Finally, it is worth stressing that though this study focused on the role of children's acquisition of complement clauses and mental verbs for developments in false-belief reasoning, we are not arguing that this is the whole story. Learning about the invisible mental world and its possible contrasts with the visible world is a complex process, and we would expect children to attend to whatever verbal and nonverbal cues their experience offers them and to recruit whatever general cognitive abilities they possess in this process. Thus, noticing surprised expressions in others' faces or behaviors such as looking for an object in a place where it is not or hearing utterances that contradict each other or contrasting verbal labels for the same object may all be sources that the child draws on when building up a stable and flexible understanding of beliefs. What the current longitudinal study shows is that acquiring the perspective-marking grammar of complement clauses with its recurring anchoring of ideas in persons and its flexible format for communicating about false beliefs is likely to support the child in this complex process.

Conclusions

The majority of languages in the world offer speakers a linguistic tool for communicating flexibly and explicitly about their own and others' invisible perspectives on ideas: the complement-clause construction. In languages such as English, this construction further hosts mental verbs that categorize epistemic states such as belief and knowledge. For young children in the process of learning about the invisible mental world, previous studies have suggested that acquisition of the perspective-marking grammar of complement clauses supports their developing ability to reason flexibly about false beliefs. Addressing critical methodological uncertainties in previous studies and targeting children at a younger age, the current longitudinal study provides strong support for the hypothesis that complement-clause acquisition supports development in false-belief reasoning. Using false-belief tests without complement clauses, testing complement-clause proficiency comprehensively, and controlling for individual differences in executive functioning, memory, and general language, we found that two aspects of complement-clause acquisition (structural proficiency and mental verb knowledge) around age three play independent roles in predicting false-belief performance 6 months later. Further, as we also found false-belief reasoning around age three to predict later complement-clause proficiency, our study yields evidence for a bidirectional relationship between linguistic and sociocognitive development.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of False Belief understanding. *Journal of Cognition and Development*, *7*(2), 139–172. https://doi.org/10.1207/s15327647jcd0702_1

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind." *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boeg Thomsen, D. (2016). *Linguistic perspective marking and mental-state reasoning in children with autism: A training study with complement clauses* [Unpublished doctoral dissertation]. University of Copenhagen.

Boeg Thomsen, D., Kandemirci, B., Theakston, A., & Brandt, S. (2021). *Perspective-marking grammar supports mental-state reasoning: A training study with English 2- to 3-year-olds* [Manuscript in preparation]. Department of Cross-Cultural and Regional Studies, University of Copenhagen.

Boeg Thomsen, D., Theakston, A., Kandemirci, B., & Brandt, S. (2021). *A longitudinal study of complement clauses and false belief*. https://osf.io/qkjn6/

Bowerman, M. M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 388–427). MIT Press. https://doi.org/10.7551/mitpress/4117.003.0021

Brandt, S., Buttelmann, D., Lieven, E., & Tomasello, M. (2016). Children's understanding of first- and third-person perspectives in complement clauses and false-belief tasks. *Journal of Experimental Child Psychology*, *151*, 131–143. https://doi.org/10.1016/j.jecp.2016.03.004

Brandt, S., Lieven, E., & Tomasello, M. (2010). Development of word order in German complement-clause constructions: Effects of input frequencies, lexical items, and discourse function. *Language*, *86*(3), 583–610. https://doi.org/10.1353/lan.2010.0010

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Burnel, M., Durrleman, S., Reboul, A., Arnaud, C., Baciu, M., & Perrone-Bertolotti, M. (2020). Theory-of-mind during childhood: Investigating syntactic and executive function contributions. *Social Development*. Advance online publication. https://doi.org/10.1111/sode.12471

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319. https://doi.org/10.1016/j.jecp.2004.01.002

Cheung, H., Chen, H. C., & Yeung, W. (2009). Relations between mental verb and False Belief understanding in Cantonese-speaking children. *Journal of Experimental Child Psychology*, *104*(2), 141–155. https://doi.org/10.1016/j.jecp.2009.05.004

Dąbrowska, E. (2004). *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh University Press.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of False Beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 189–226). Psychology Press. https://doi.org/10.4324/9781315784670

de Villiers, J. G., & Pyers, J. E. (1997). Complementing cognition: The relationship between language and theory of mind. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 136–147). Cascadilla Press.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*(1), 1037–1060. https://doi.org/10.1016/S0885-2014(02)00073-4

de Villiers, P. A., & de Villiers, J. G. (2012). Deception dissociates from False Belief reasoning in deaf children: Implications for the implicit versus explicit theory of mind distinction. *British Journal of Developmental Psychology*, *30*(1), 188–209. https://doi.org/10.1111/j.2044-835X.2011.02072.x

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics*, *12*(2), 97–141. https://doi.org/10.1515/cogl.12.2.97

Dixon, R. M. W. (2006). Complement clauses and complementation strategies in typological perspective. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Explorations in linguistic typology* (Vol. 3, pp. 1–48). Oxford University Press.

Dunn, L., Dunn, D., Styles, B., & Sewell, J. (2009). *British Picture Vocabulary Scale* (3rd ed.). GL Assessment.

Durrleman, S., Burnel, M., De Villiers, J. G., Thommen, E., Yan, R., & Delage, H. (2019). The impact of grammar on mentalizing: A training study including children with autism spectrum disorder and developmental language disorder. *Frontiers in Psychology*, *10*, 2478. https://doi.org/10.3389/fpsyg.2019.02478

Durrleman, S., Burnel, M., Thommen, E., Foudon, N., Sonié, S., Reboul, A., & Fourneret, P. (2016). The language cognition interface in ASD: Complement sentences and False Belief reasoning. *Research in Autism Spectrum Disorders*, *21*, 109–120. https://doi.org/10.1016/j.rasd.2015.10.003

Durrleman, S., Burnel, M., & Reboul, A. (2017). Theory of mind in SLI revisited: Links with syntax, comparisons with ASD. *International Journal of Language & Communication Disorders*, *52*(6), 816–830. https://doi.org/10.1111/1460-6984.12317

Durrleman, S., & Franck, J. (2015). Exploring links between language and cognition in autism spectrum disorders: Complement sentences, False Belief, and executive functioning. *Journal of Communication Disorders*, *54*, 15–31. https://doi.org/10.1016/j.jcomdis.2014.12.001

Elliott, C. D. (1996). *The British Ability Scales II*. NFER-NELSON.

Gola, A. H. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development*, *27*(1), 64–76. https://doi.org/10.1016/j.cogdev.2011.10.003

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of False Belief and the appearance-reality distinction. *Child Development*, *59*(1), 26–37. https://doi.org/10.2307/1130386

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction (*2nd ed.). de Gruyter. https://doi.org/10.1515/9783110307474

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit False Belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, *6*(3), 346–359. https://doi.org/10.1111/1467-7687.00289

Hansen, M. B. (2010). If you know something, say something: Young children's problem with False Beliefs. *Frontiers in Psychology*, *1*, 23. https://doi.org/10.3389/fpsyg.2010.00023

Hansen, M. B., & Markman, E. M. (2005). Appearance questions can be misleading: A discourse-based account of the appearance-reality problem. *Cognitive Psychology*, *50*(3), 233–263. https://doi.org/10.1016/j.cogpsych.2004.09.001

Harrell, F. E. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. Springer Science+Business Media. https://doi.org/10.1007/978-1-4757-3462-1

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus False Belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*(3), 567–582. https://doi.org/10.2307/1130337

Howard, A. A., Mayeux, L., & Naigles, L. (2008). Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language*, *28*(4), 375–402. https://doi.org/10.1177/0142723708091044

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Kidd, E., Lieven, E., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*(1), 132–142. https://doi.org/10.1016/j.langsci.2009.05.002

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507. https://doi.org/10.2307/1131828

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. M. (1996). Social influences on False Belief access: Specific sibling influences or general apprenticeship? *Child Development*, *67*(6), 2930–2947. https://doi.org/10.2307/1131760

Lewis, S., Hacquard, V., & Lidz, J. (2017). Think pragmatically: Children's interpretations of belief reports. *Language Learning and Development*, *13*(4), 395–374. https://doi.org/10.1080/15475441.2017.1296768

Lind, S. E., & Bowler, D. M. (2009). Language and theory of mind in autism spectrum disorder: The relationship between complement syntax and False

Belief task performance. *Journal of Autism and Developmental Disorders*, *39*(6), 929–937. https://doi.org/10.1007/s10803-009-0702-y

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of False Belief understanding: A training study. *Child Development*, *74*(4), 1130–1144. https://doi.org/10.1111/1467-8624.00597

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development*, *81*(2), 597–615. https://doi.org/10.1111/j.1467-8624.2009.01418.x

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Lust, B., Flynn, S., & Foley, C. (1996). What children know about what they say: Elicited imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee, & H. Smith Cairns (Eds.), *Methods for assessing children's syntax* (pp. 55–76). MIT Press.

Miller, C. (2004). False belief and sentence complement performance in children with specific language impairment. *International Journal of Language & Communication Disorders*, *39*(2), 191–213. https://doi.org/10.1080/13682820310001616994

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Mo, S., Su, Y., Sabbagh, M. A., & Xiu, J. (2014). Sentential complements and False Belief understanding in Chinese Mandarin-speaking preschoolers: A training study. *Cognitive Development*, *29*, 50–61. https://doi.org/10.1016/j.cogdev.2013.11.001

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*(3), 722–730. https://doi.org/10.2307/1130957

Moore, R. (2020). The cultural evolution of mind-modelling. *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-020-02853-3

Noonan, M. (2007). Complementation. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 52–150). Cambridge University Press. https://doi.org/10.1017/CBO9780511619434.002

Okuno, A., Cameron-Faulkner, T., & Theakston, A. (2020). Crosslinguistic differences in the encoding of causality: Transitivity preferences in English and Japanese children and adults. *Language Learning and Development*, *16*(1), 61–88. https://doi.org/10.1080/15475441.2019.1685387

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with False Belief. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). Want that is understood well before say that, think that, and False Belief: A test of de Villiers's linguistic determinism on German-speaking children. *Child Development*, *74*(1), 179–188. https://doi.org/10.1111/1467-8624.t01-1-00529

R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved 18th June, 2019, from https://www.R-project.org/

Roman, A. S., Pisoni, D. B., & Kronenberger, W. G. (2014). Assessment of working memory capacity in preschool children using the missing scan task. *Infant and Child Development*, *23*(6), 575–587. https://doi.org/10.1002/icd.1849

Rubio-Fernández, P. (2020). Pragmatic markers: The missing link between language and theory of mind. *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-020-02768-z

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, *7*(4), 835–850. https://doi.org/10.1007/s13164-015-0290-z

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate False Belief understanding. *Developmental Psychology*, *34*(1), 161–174. https://doi.org/10.1037/0012-1649.34.1.161

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, *78*(2), 376–396. https://doi.org/10.1111/j.1467-8624.2007.01004.x

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of False Beliefs. *Cognition*, *38*(1), 1–12. https://doi.org/10.1016/0010-0277(91)90020-5

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

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

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*(1), 127–152. https://doi.org/10.1017/S0305000900004608

Tomasello, M. (2018). How children come to understand False Beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, *18*(2), 121–147. https://doi.org/10.1111/1468-0017.00217

Vendetti, C., Kamawar, D., Podjarny, G., & Astle, A. (2015). Measuring preschoolers' inhibitory control using the black/white stroop. *Infant and Child Development*, *24*(6), 587–605. https://doi.org/10.1002/icd.1902

Verhagen, A. (2005). *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford University Press.

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., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *82*(3), 780–792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

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

Wiig, E. H., Semel, E. M., & Secord, W. (2006). *Clinical evaluation of language fundamentals preschool 2*. Pearson Education/Psychorp.

Williams, D. M., & Happé, F. (2009). What did I say? Versus what did I think? Attributing False Beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders*, *39*(6), 865–873. https://doi.org/10.1007/s10803-009-0695-6

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

Woensdregt, M., Cummins, C., & Smith, K. (2020). A computational model of the cultural co-evolution of language and mindreading. *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-020-02798-7

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46

Appendix

Experimental Items

Complement Comprehension

(C) Jack hoped that the cat was sweet. Then he stroked it. What did Jack hope?
(D) Nick found a rope, but he shouted that he found a snake. What did Nick shout?
(D) The park was still open, but Mrs. Scott read that it was closed. What did Mrs. Scott read?
(C) Pam saw that her dad picked flowers. Then she found a vase for them. What did Pam see?
(C) The car was blue, but Mr. Smith said that it was red. What did Mr. Smith say?
(D) Sue heard that her parents were angry. Then she went into her playroom. What did Sue hear?
(D) Luke noticed that his sister spilt milk. Then he got a cloth for her. What did Luke notice?
(C) Claire met a girl, but she pretended that she met a queen. What did Claire pretend?

Complement Repetition

(B) Jean hoped that the tea was hot.
(A) You know the boy has a drum.
(B) John saw that his mum ate sweets.
(A) You think the doll is very pretty.
(B) Dan pretended that he brushed his hair.
(A) I think the monkey likes the fruit.
(B) Ann said that she was very hungry.
(A) I know the coat is really dirty.

Received May 1, 2020
Revision received May 5, 2021
Accepted May 19, 2021 ■