Title:

Do complement clauses with first- or third-person perspective support false-belief reasoning? A training study with English-speaking three-year-olds

Authors and affiliations:

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

[1] Department of Nordic Studies and Linguistics, University of Copenhagen
[2] Division of Psychology, Communication, and Human Neuroscience, University of Manchester
[3] Department of Linguistics and English Language, Lancaster University

Corresponding author:

Ditte Boeg Thomsen
Department of Nordic Studies and Linguistics
University of Copenhagen
Emil Holms Kanal 2
2300 København S
Denmark
ditte.boeg@hum.ku.dk

Contact details for all other authors:

Birsu Kandemirci: birsu.kandemirci-2@manchester.ac.uk
Anna Theakston: anna.theakston@manchester.ac.uk
Silke Brandt: s.brandt@lancaster.ac.uk

Submission Date:

27 July 2023

Revision Dates:

First revision submitted 29 January 2024
Second revision submitted 15 May 2024

OSF repository link

https://osf.io/vkw63/

Funding information

Economic and Social Research Council: grant number ES/L008955/1 (to Anna Theakston and Silke Brandt)
Economic and Social Research Council: grant number ES/S007113/1 (to Anna Theakston and Silke Brandt)

Do complement clauses with first- or third-person perspective support false-belief reasoning? A training study with English-speaking three-year-olds

Abstract

To investigate whether acquisition of the perspective-marking grammar of complement-clause constructions supports progression in children's false-belief reasoning, we conducted a training study with 76 English-speaking three-year-olds from the North-West of England (age range: 3;0-3;10 years, 50% female, 80% White). Children were randomly assigned to one of three maximally comparable training conditions, and in a four-week eight-session program, all children participated in the same training activities with mental-state contrasts. Depending on condition, activities were mediated linguistically with either simple clauses, first-person complements or third-person complements. The study addressed critical confounds in previous training studies by avoiding the use of complement clauses in false-belief tests and controlling individual differences in memory, executive functioning, general language and pretest proficiency with complement clauses. The results yielded strong support for the hypothesis of a causal influence of complement-clause exposure on false-belief progression, as children trained with first-person complements advanced significantly more in false-belief reasoning from pretest to posttest than children trained with simple clauses. Examining the roles of first- and third-person complements, a direct comparison between progression in the two complement-clause conditions showed no significant difference, but only children trained with first-person complements progressed significantly more than children in the control condition trained with simple clauses. Follow-up analyses suggested that first- and third-person complements each support false-belief progression at different stages of development.

Keywords: complement clauses; false belief; language acquisition; social cognition; thinking for speaking; executive functioning.

Public significance statement

Understanding that people see situations in different ways is a major achievement in children's social-cognitive development, and in an intervention study, we investigated the role of language in supporting this type of social-cognitive understanding. We found that using sentences indicating perspective like "I thought/knew that the book was in the drawer" when talking about mistakes and disagreement helped three-year-olds develop a more flexible understanding of their own and others' beliefs.

Developing robust abilities to represent and reason about mental phenomena such as belief and knowledge is an important phase in young children's sociocognitive development. Mental-state reasoning helps children to see situations from others' perspectives, to acknowledge that they themselves can be wrong and to understand and interact in social situations involving disagreement, surprise, misunderstanding and pretence. A core aspect of mental-state reasoning is the ability to remember one's own belief and imagine others' beliefs and to explain behaviour as motivated by these invisible beliefs about reality, rather than by reality itself (Wellman et al., 2001). Such belief-reasoning skills are especially useful in situations involving *false belief*, i.e., beliefs that do not match reality, and during the preschool years, children's abilities to represent and reason about false beliefs become increasingly stable and flexible (Tomasello, 2018). While children typically begin to pass explicit false-belief tests between the ages of three and five years, there is substantial individual variation, and this variation appears to be tightly related to development in other areas, especially executive functioning and language. For language, both lexical and grammatical experience and knowledge appear to affect belief reasoning (Lohmann & Tomasello, 2003; Milligan et al., 2007; Ruffman et al., 2002).

One aspect of linguistic development that has been especially thoroughly investigated as a factor in sociocognitive development is acquisition of complement clauses. The complement-clause construction is a crosslinguistically widespread grammatical tool for expressing persons' perspectives on ideas, as in *She thinks [it's her book]* and *I know [it's my book*]. In complement-clause constructions, a complement clause expressing an idea (*it's her book…*) is embedded in another clause (*She thinks, I know…*) that specifies whom the idea belongs to (*she, I, mum, my friends…*) and which perspectival relation holds between person and idea (e.g., *think, know, hope, regret, remember…*). When children acquire the perspective-marking grammar of complement-clause constructions, they gain access to a flexible and explicit community-shared format for making all sorts of invisible mental states, including false beliefs, the focus of joint attention.

Several studies have investigated the hypothesis that acquisition of these explicit perspective-marking constructions supports typically developing children's emerging abilities to represent and reason about false beliefs. Two longitudinal studies found earlier complement-clause proficiency to predict later false-belief reasoning (Boeg Thomsen et al., 2021; de Villiers & Pyers, 2002), and five training studies found exposure to complement-clause constructions to promote false-belief development (Gola, 2012; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003; Mo et al., 2014; San Juan & Astington, 2016). While training studies in principle offer the strongest evidence for causal relationships, conclusions from the five training studies are uncertain due to use of complement clauses in false-belief tests and control conditions and because none of the studies tested the effects of individual differences in executive functioning or in pretest proficiency with complement clauses. The present study was designed to test whether we can replicate the finding that complement-clause training promotes sociocognitive development in a training study where complement clauses exclusively occur in complement-clause training, not in false-belief tests or control conditions, and where individual differences in executive functioning and complement-clause proficiency are controlled.

If complement-clause acquisition does indeed promote sociocognitive development, a second important question is whether different types of complement-clause constructions provide children with different types of sociocognitive support. One hypothesis that has attracted much attention is that experience with constructions with third-person subjects (e.g., <u>*She*</u> *thinks it's a swan*) is more likely to advance belief reasoning than experience with first-person constructions (e.g., <u>*I*</u> *think it's a swan*). This hypothesis is based on theoretical arguments that early-acquired first-person complements do not have genuine mental-state reference and do not present clear cues to differences between belief and knowledge (e.g., Diessel & Tomasello, 2001; Howard et al., 2008). The only experiments directly comparing relations between belief reasoning and complement-clause constructions with first- vs. third-person subjects all suggest an advantage for third-person constructions (Boeg Thomsen et al., 2021; Brandt et al., 2016; Gola, 2012), but as discussed in more detail below, the evidence these studies provide is still preliminary. The role of person in the relationship between complement clauses and belief reasoning is therefore still uncertain, and alternative scenarios are also possible, as theoretical arguments could also support a similar role of first- and third-person complements (Verhagen 2005) or even an advantage of first-person complements. The current training study therefore tests directly whether experience with third-person complements really is more likely to propel false-belief development than experience with first-person complements.

Below we review the evidence from previous complement-clause training studies in more detail and present the controversies regarding the role of person more thoroughly before outlining the current study.

Complement clauses and false belief: training studies

The most direct way to investigate the hypothesis that complement-clause acquisition supports false-belief reasoning is by testing whether children who are exposed to complement clauses show more progression in false-belief reasoning than children who are not. Five studies in three languages have used a training methodology to address this question. For German, Lohmann and Tomasello (2003) found that children (3;3-4;10 years) who had deceptive experience mediated linguistically with contrasting nouns and complement clauses showed greater progression in false-belief reasoning than children who had the same experience mediated only with contrasting nouns. Children who were exposed to complement clauses without any deceptive experience also progressed more than children who experienced deceptive scenarios with minimal linguistic mediation. For English, Hale and Tager-Flusberg (2003) found that children (3;0-4;10 years) advanced significantly from pretest to posttest if they had been trained with either complement clauses or direct false-belief training, but not with another type of clause embedding: relative clauses. In another training study with English-speaking children (3;0-4;9 years), Gola (2012) also found improvement after complement-clause training, but only with complement clauses with second- or third-person subjects. Providing evidence from a different language family, Mo et al. (2014) found that Mandarin Chinese-speaking children (3;4-4;7 years) trained with complement clauses with communication verbs performed better on false-belief posttests than children trained with minimal language.

In contrast, San Juan and Astington (2017) did not find an effect of complement-clause training on explicit false belief (measured with verbal responses) in English-speaking children (2;6-4;5 years), but they did find an effect on implicit false belief (measured with eye tracking). As the study differed from previous studies by combining a passive video-watching format with short dosage (two training sessions), training may have provided children with less opportunities for engaging actively with the training materials, which could explain why only effects on implicit false belief were found.

In sum, five training studies in three languages support the hypothesis that exposure to the perspective-marking syntax of complement-clause constructions promotes development in false-belief reasoning. But, as we will discuss below, there are critical limitations to the conclusions of these studies.

Confounds and uncertainties in previous training studies

Complement clauses in false-belief tests: All five training studies used complement clauses in their false-belief posttests, as in "What did you think was in the box when you first saw it?" (Hale & Tager-Flusberg, 2003) or "What does Aunt Sally think is inside this box, band-aids or ribbon?" (Gola, 2012). When passing false-belief posttests requires complement-clause mastery, this entails the risk that training effects are simply performance effects: complement-clause training could help children advance in false-belief *performance* by strengthening their linguistic ability to perform the false-belief tasks without promoting conceptual development. Indeed, the results pattern in San Juan and Astington (2017) could support such an interpretation. This study did not use complement-clause constructions in its explicit target questions and did not find a training effect for explicit false belief. The implicit false-belief measure on the other hand used complement-clause constructions as verbal prompts to elicit anticipatory looking ("I wonder where [C2] is going to put it/what [C2] thinks/gorps that it is" and "I wonder where [C2] will look first for the cake"), and on implicit false belief, a beneficial effect of training with complements was found. This critical confound in previous studies makes it crucial to replicate the finding that complement-clause training promotes false-belief development in a study that avoids the use of complement clauses in false-belief tests[1].

Complement clauses in control and training conditions: Three of the studies further presented children in their supposedly complement-clause-free control conditions with several complement clauses during training. Judging from their protocol, Mo et al. (2014) appear to have presented children in their visual False Representation condition with 10-18 complement clauses (e.g., "When this boy/girl saw this picture, does he/she know there is a horse in this picture?"). Similarly, children in San Juan and Astington's (2017) control group heard 12 complement clauses with the perception verb *see* during training: "[C2] sees that something has been left on the table" (vs. 24 complement clauses in the complement conditions). Finally, children in Lohmann and Tomasello's (2003) Discourse Only condition appear to have heard the complement-clause construction "Let's hear what he says" 16 times. This apparent design error makes it problematic to use comparisons between conditions to inform us about the actual difference between being exposed to complement clauses or not. The training study by Hale and Tager-Flusberg (2003) presents the opposite problem, as their complement-clause condition only exposed children to two genuine complement clauses all in all (the majority of speech reports employed direct speech, e.g., "The boy says, 'I kissed Grover.'"), making it uncertain what caused the training effect. To obtain an accurate understanding of the role of complement-clause training in supporting social cognition, we should control the occurrence of complement clauses in different conditions stringently: avoiding the use of complement clauses in control conditions and making sure the complement-clause conditions present plenty of genuine complement-clause constructions.

The role of individual differences: None of the studies included measures of executive functioning, entailing a risk that uninvestigated individual differences in ability to benefit from training affected training outcomes. Developments in executive functioning and false-belief reasoning are related (e.g., Carlson et al., 2004; Carlson & Moses, 2001; Hughes & Ensor, 2007), and, crucially, Benson et al. (2013) demonstrated that differences in three-year-olds' executive-functioning skills predicted their *ability to benefit from false-belief training*. The better the children performed on executive-functioning tasks, the more and the faster they learned from training sessions with false-belief scenarios, and as Benson et al. (2013) argue, ability to suppress one's own salient perspective may be a prerequisite for attending to the contrasting perspectives false-belief training activities are rich in. The present study aims to examine the effects of linguistic mediation of perspective-rich situations while controlling the influence from individual differences in receptiveness to training, and we therefore include a broad spectrum of executive-functioning, memory and language measures.

The role of pretest complement-clause proficiency*:* One final uncertainty regards the role of proficiency with complement clauses *before* training. We can expect large variation in the three- and four-year-olds targeted by the training studies, and susceptibility to benefit from complement-clause training is unlikely to be independent from pretest level. Children with weaker schemas for complement-clause constructions could be expected to profit more from intensive exposure than children who have already developed a more stable grasp of the construction. Apart from Gola (2012), all studies tested pretest proficiency with complement clauses, but with mixed sensitivity to variation. Hale and Tager-Flusberg (2003) and Mo et al. (2014) assessed pretest proficiency with just two items while Lohmann and Tomasello (2003) and San Juan and Astington (2017) used more sensitive tests (max score: 12). In those four studies, scores were used to check training group comparability and progression from pretest to posttest, but none of the studies examined whether pretest proficiency *interacts* with training condition to predict false-belief progression. To assess the role of complement-clause training in supporting social cognition accurately, we need precise measures of pretest proficiency, and we must test whether this proficiency plays a different role in training conditions with and without complement clauses.

First- vs. third-person complements

If acquisition of complement clauses supports sociocognitive development, an interesting question is whether different types are equally beneficial. The question is relevant because complement-clause acquisition is a process that unfolds over years, with children gradually constructing increasingly abstract schemas by generalising over the concrete expressions they hear and use (Boeg Thomsen, 2015; Brandt et al., 2010; Diessel & Tomasello, 2001). In this process, different construction types could affect social cognition differentially, and one question that has attracted much discussion is whether complement clauses with first-person subjects (such as ***I** think they're on holiday*) provide less support for false-belief development than complement clauses with third-person subjects (such as ***She** thinks they're on holiday*). There are various theoretical arguments why this might be the case. Most importantly, there is a long tradition of suspecting early-acquired first-person complement-clause constructions to be formulaic entities without genuine mental-state reference that children can understand and use without representing mental states. The processing of third-person complements, on the other hand, has been argued to require representation of mental states and thus to be more intimately connected with mental-state-reasoning development. Below, we sum up the empirical evidence supporting this expected difference between first- and third-person complements before considering two alternative hypotheses: that first- and third-person complements could be equally beneficial for false-belief development and that first-person complements could provide more support for false-belief development than third-person complements.

Based on a child-language corpus study, Diessel and Tomasello (2001) suggested that high-frequency strings with first-person subjects like *I think* and *I hope* function as adverb-like epistemic markers, like *maybe* and *hopefully*. Contrary to constructions with third-person subjects (*He thinks it's a car*), constructions with these strings (*I think it's a car*) were analysed as not being genuine complement-clause constructions and not having genuine mental-state reference. Howard et al. (2008) further suggested that first-person complements provide children with fewer linguistic cues to the difference between belief and knowledge, as mothers often used *I think* in certainty contexts in child-directed speech (e.g., *I think you've had enough peas for now*), making the contrast between the mental verbs *think* and *know* less clear-cut for first-person complements.

Gola (2012) found that only children trained with complement-clause constructions with other-person subjects (combined second and third person), not first-person subjects, advanced significantly in false belief, and correlational and longitudinal studies also suggest that first-person complements play a less central role in false-belief development. Four-year-olds' false-belief performance correlates with their understanding of third-person complements, not first-person complements (Brandt et al. 2016), and for two- and three-year-olds, understanding of third-person but not first-person complements predicts false-belief reasoning six months later (Boeg Thomsen et al. 2021).

Evidence from corpus, correlational, longitudinal and training studies thus align in suggesting a less beneficial role of first-person complements in promoting false-belief development, but nevertheless, this evidence must still be considered preliminary. First, the training study by Gola (2012) had little power to compare false-belief progression, with small training groups (12 children per condition) and only three false-belief questions. All three false-belief tests further involved complement-clause constructions with second- or third-person subjects, making it impossible to exclude that any apparent false-belief advancement in the other-person conditions simply depended on the same constructions being used in training and false-belief tests. Moreover, the training scripts mixed up the variables of subject *person* and subject *variability*. Children trained with first-person complements only heard constructions with the subject *I*, whereas children trained with other-person complements heard more varied complement-clause constructions with second-, third- and even first-person subjects, supporting generalisation and abstraction. Pragmatic appropriateness also varied across conditions. Whereas second-person questions like *Do you think this is a golf ball?* are pragmatically fine and highly frequent, the first-person version, *Do I think this is a golf ball?*, is pragmatically strange and highly infrequent in both child speech and child-directed speech. All in all, with its many critical confounds, the results in Gola (2012) are not informative about the role of person in complement-clause training.

There are also reasons to view results from other studies suggesting a weaker link between first-person complements and false belief with caution. Even though Boeg Thomsen et al. (2021) found children's understanding of third-person complements and not first-person complements to predict later false-belief reasoning, they also stress that the Hidden Objects task they used may not be ideal for measuring children's understanding of first-person complements. The test measures sensitivity to contrasts in certainty (*I think* vs. *I know*, *He thinks* vs. *He knows*), but given that *I think* often signals certainty in adult speech (Howard et al., 2008), it is unclear what children's performance with first-person complements in the test reflects.

Given that there is no strong empirical evidence supporting a less important role for first-person complements in supporting false-belief development, we can also consider alternative hypotheses. One possibility is that first- and third-person complements are equally beneficial for false-belief development. Such a scenario would be theoretically supported by the functional-cognitive analysis of complement-clause constructions by Verhagen (2005), who argues that first-person complements are not functionally different from third-person complements, as both basically perform the same function: anchoring ideas in persons. This may explain why Moore et al. (1990) *did* find significant correlations between false-belief reasoning and comprehension of first-person complements in four-year-olds, indicating that these complements do not lack genuine mental-state reference.

Another possibility is that first-person complements provide *more* support for false-belief development than third-person complements. Based on the literature, this is not a scenario we would expect, but there could be some theoretical arguments in favour of an advantage of first-person complements. First, in a corpus analysis of the Manchester CHILDES corpus (Theakston et al., 2001), we found first-person complements to be much more frequent than third-person complements in child-directed speech. The higher frequency of first-person complements may make them easier to process, freeing resources to process their sociocognitive content, and there is some empirical support for the possibility of such an effect. In their training study with Mandarin Chinese, Mo et al. (2014) found that only children trained with high-frequency communication verbs, not low-frequency mental verbs, exhibited better false-belief performance than the control group at posttest. Second, the perspectival structure of first-person complements might offer children simpler opportunities to practice representing others' perspectives than third-person complements. When children hear a speaker saying: *I hope he's on his way*, they only have to process one perspective in addition to their own, the speaker's, but if the speaker says *She hopes he's on his way*, they arguably have to process two perspectives in addition to their own: the speaker's and the perspective represented by the third-person subject (cf. van Duijn & Verhagen, 2018, on the increased sociocognitive complexity involved in communicating about third-person perspectives). Further, the child will often have more non-verbal cues to estimating the mental state of a speaker than to estimating the mental state of a not necessarily present third person, suggesting that it will often be easier for a child to gauge the relationship between first-person complements and the mental states they represent than between third-person complements and the mental states they represent. At least at the early stages of developing routines of orienting towards and reasoning about others' perspectives, the perspectival simplicity of first-person complements compared to third-person complements might thus offer children more sociocognitive support. All in all, the evidence from studies comparing the roles of first- and third-person complements in sociocognitive development must currently be considered inconclusive. There are theoretical reasons for hypothesising a stronger influence from third-person complements, as they may present clearer cues to building separate categories of belief and knowledge (cf. Howard et al., 2008), but there are also theoretical reasons for hypothesising a similar influence from first- and third-person complements, as both invite children to anchor ideas in persons (Verhagen 2005). The possibility that first-person complements could even exert a stronger influence can also not be excluded, as their higher frequency (cf. Mo et al. 2014) and simpler perspectival structure could theoretically be beneficial, especially at early stages of development. To investigate these contrasting hypotheses, the current study compares the effect on false-belief progression of training with first- and third-person complements in maximally similar training conditions.

The current study

The current study used a training paradigm to investigate whether experience with complement-clause constructions supports development in false-belief reasoning in typically developing English-speaking three-year-olds. Previous training studies yield preliminary support for such a causal influence from language on social cognition, but the current study tests the relation more stringently. Children were randomly assigned to training conditions that were maximally comparable, in the sense that all children took part in exactly the same activities, and the only difference between conditions was whether these activities were mediated linguistically with complement-clause constructions (*I think that that's a tiger*) or without complement-clause constructions (*That's a tiger*). Addressing critical confounds in previous training studies, the current study: 1) avoided the use of complement clauses in pre- and posttests of false belief, 2) avoided the use of complement clauses in the control condition, and 3) controlled individual differences in memory, executive functioning and general language. Contrary to previous studies, we further collected comprehensive measures of pretest complement-clause proficiency and tested whether individual differences in pretest proficiency affected receptiveness to training differentially in conditions with and without complement-clause exposure. Finally, while some studies suggest that complement-clause constructions with first-person subjects are less likely to promote false-belief development than third-person constructions, there is still no solid evidence. Therefore, we directly compared false-belief progression in conditions exposing children to either first-person complements (*I think that that's a tiger*) or third-person complements (*He thinks that that's a tiger*).

Method

Transparency and openness

All test and training materials, data and scripts for analysis are publicly available on the Open Science Framework project (https://osf.io/vkw63/). Sound and video recordings are available in the UK Data Archive [study number: 855166]. The study was not pre-registered. We report how we determined sample size, data exclusions and all measures in the study. The study received ethical approval from the Faculty of Arts and Social Sciences and Lancaster Management School's Research Ethics Committee at Lancaster University (UREC: FL16283).

Participants

Our final sample consisted of 76 English monolingual three-year-olds (38 boys, 38 girls, M_{age}: 3;5 years, *SD*: 2.62 months, *range*: 3;0-3;10 years at pretest) who completed pretest, training and posttest. The children were recruited from ten nurseries and preschools in Lancaster and Manchester. For the Lancaster subset (*n* = 45), 95.6% were White, representative of the resident population in the area (Lancashire County Council, 2011). For the Manchester subset (*n* = 31), 58% were White, 22% Black and 20% Asian, reflecting the more ethnically mixed resident population in the metropolis (Manchester City Council, 2011). An additional 6 children were recruited, but not included in the final sample because they did not respond to the pretests and were therefore excluded from training (*n* = 2), responded correctly to all false-belief questions at pretest and thus were not eligible for training (*n* = 2), participated in training, but went on holiday before the posttests (n = 1), or turned out to have English as a second language (n = 1). Based on a previous lab experiment with children down to age 2;9, we also recruited five two-year-olds (2;9-2;11) for our first training batch. However, as the two-year-olds did not show susceptibility to training and appeared to be too young to benefit, we did not include them in the final sample and only recruited three-year-olds for the remaining batches. Thus, 87 children were seen in total, whereof 76 were included in the final sample[2].

Pretests, posttests and control measures

The study included tests of false belief, complement-clause proficiency, general language, memory and executive functioning (see Table 1).

Table 1

Overview of tests administered

False Belief (pretest and posttest)	Complement-Clause Constructions (pretest and posttest)	Individual Differences (control)
Unexpected Location: Other	Complement-Clause Proficiency	Short-Term Memory
Unexpected Contents: Self	Mental Verbs: 1st person	Working Memory
Unexpected Contents: Other	Mental Verbs: 3rd person	Inhibitory Control 1
Unexpected Identity: Self		Inhibitory Control 2
Unexpected Identity: Other		Cognitive Flexibility
		Receptive Vocabulary

[2]Sample size was determined by following the recommendations by Brysbaert & Stevens (2018) and the preprint tutorial in Kumle et al. (2021). We used the simR package (Green & MacLeod, 2016) to generate a dataset in R (version 3.6.2, R Core Team 2019), estimating means and standard deviations based on a study using the same tests (Boeg Thomsen et al., 2021) and estimating training effects from two false-belief training studies (Benson et al., 2013; Boeg Thomsen, 2016). For effects that had not previously been examined in a training study, we cautiously estimated effect as low: 0.2. Running 1000 simulations per predictor indicated that a sample of 72 participants would yield adequate power for all target predictors, and we therefore continued data collection until at least 72 participants had completed posttest.

False belief

To evaluate changes in false belief, we collected measures of false-belief understanding at pretest and posttest. The tasks required children to remember their own previous false belief (two Self questions) or to predict another's false belief (three Other questions). At pretest, three false-belief tasks were administered: Unexpected Location, Unexpected Contents and Unexpected Identity. At posttest, three tasks with exactly the same structure, but new content, were administered. We originally developed this two-part three-task battery for a longitudinal lab study, adapting the tasks from standard false-belief tests, and detailed motivations for the adjustments can be found in Boeg Thomsen et al. (2021). The most crucial modification was the complete avoidance of complement-clause constructions in the tests. Further, the tasks were more pragmatically transparent (cf. Hansen, 2010) and had lower memory demands (cf. Rubio-Fernández & Geurts, 2013). More information on the tests can be found in the Appendix, and scripts for the tests can be found in Part 1 of the online supplemental material.

Responses to each of the five false-belief questions (items) were binary-coded, with children scoring 1 point per item for correct responses and 0 points for incorrect responses, provided that they passed the control question(s) for the task. All mixed-effects regression analyses included items as random factors and were therefore conducted at item level. In the focal change-score analysis examining progression from pretest to posttest, the dependent variable False Belief: Change Score took values of 1, 0 and -1 at item level (i.e. the result of binary-coded posttest score (1/0) minus binary-coded pretest score (1/0) for each individual item). Analyses not involving regression (e.g., the pretest correlation matrix) used the total score (0-5: the summed score of responses to all five questions). If children failed one or both control questions for a task, their false-belief response(s) on that specific task were discarded as invalid (NA). These invalid responses were not included in the regression analyses (but the child's valid responses on other false-belief tasks were), and children with one or more invalid false-belief responses did not achieve a total 0-5 score (NA).

Complement clauses

To evaluate whether individual differences in pretest proficiency with complement clauses would affect children's receptiveness to training with or without complement clauses, we measured pretest proficiency comprehensively, with two tests: Complement-Clause Proficiency and Hidden Objects. The first measures constructional flexibility in children's schemas for complement-clause constructions while the second measures understanding of the mental verbs *think* and *know* in complement-clause constructions. To further assess whether progression in complement-clause proficiency from pretest to posttest would predict false-belief progression, we measured complement-clause proficiency again at posttest, using tests with exactly the same structure, but new contents.

Complement-Clause Proficiency is a 16-item test designed to measure fine-grained differences in proficiency with complement-clause constructions with varying frequencies. We originally designed this test for a longitudinal lab study, and further details of test construction and administration can be found in Boeg Thomsen et al. (2021). The test has two parts: an elicited-imitation part where children repeated eight pre-recorded stimuli sentences with complement-clause constructions, as in (1), and a comprehension part where children heard complement-clause constructions and distractor clauses and answered eight comprehension questions targeting the embedded complement clauses, as in (2):

(1) *John saw that his mum ate sweets.*

(2) *Liz heard that her parents were awake.* (complement-clause construction)

Then she went into their bedroom. (distractor clause)

What did Liz hear?

Half of the comprehension items presented a contrast between the proposition expressed in the complement clause and the proposition expressed in the distractor clause, as in *Sam found a ball* (distractor clause)*, but he shouted [that he found an apple]* (complement clause), whereas the other half did not, as in (2). The comprehension part was adapted from the original *Memory for Complements* test (de Villiers and Pyers, 1997), with a series of adjustments, posing lower demands on inhibition and working memory and using many more different complement-taking verbs (see Boeg Thomsen et al., 2021).

To capture different stages in complement-clause proficiency, the test presented constructions ranging from early-acquired high-frequency constructions to late-acquired low-frequency constructions. Constructions at the most basic level presented high-frequency complement-taking verbs in high-frequency strings with 1st- and 2nd-person subjects (*I think, you think, I know, you know*), as in (3). Constructions at the middle level presented high-frequency complement-taking verbs (*say*, *see*, *hope*, *pretend*) in low-frequency clauses with 3rd-person subjects, as in (4), and constructions at the most advanced level presented low-frequency complement-taking verbs (*notice*, *hear*, *shout*, *read*) in low-frequency clauses with 3rd-person subjects, as in (5):

(3) ***I know*** *the boot is really muddy* (basic)

(4) ***Jack hoped*** *that the cat was sweet* (middle)

(5) ***Luke noticed*** *that his sister spilt milk* (advanced)

Repetition and comprehension responses were coded following a detailed coding manual (see Part 2 of the online supplemental material). Assessment of responses to each of the 16 individual items was graded (0-2 points), yielding a maximum test score of 32.

Hidden Objects was used to examine understanding of complement-clause constructions with mental verbs in more depth, and to scrutinise potential differences between constructions with first- and third-person subjects. This forced-choice task follows the basic structure from Moore et al. (1990) and Brandt et al. (2016), where a child chooses where to look for stickers based on advice from puppet helpers, and we included the pragmatic adjustments from Boeg Thomsen et al. (2021). In each of 12 trials, the puppets (a pig, a cow and a boy) used complement-clause constructions with *think* and *know* to express contrasting statements about the location of the sticker, as in (6) and (7):

(6) **Mental Verbs: 1st person (6 trials)**

Pig: "I **think** that the sticker is in the blue box."

Cow: "I **know** that the sticker is in the red box."

(7) **Mental Verbs: 3rd person (6 trials)**

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

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

All sentences were pre-recorded and played from a laptop. We counterbalanced how often each colour and puppet were associated with *know* and *think*, as well as the trial-internal order of *know* and *think* statements. Children did not open any boxes before completing the trials, and all boxes contained stickers.

As complement-clause constructions with *know* convey higher certainty than constructions with *think* (although *I think* may, in fact, also signal certainty in everyday use), the child demonstrates understanding of the meaning of complement-clause constructions with mental verbs by consistently choosing the box indicated by the verb *know*. For each of these choices the child scored one point. Points for Mental Verbs: 1st person and Mental Verbs: 3rd person were counted separately, yielding raw scores of 0-6 for each type. However, raw scores are not informative in themselves, as children by chance get 3 out of 6 correct if they show biases such as always choosing their favourite colour, the pig's suggestion, or the first or last box mentioned. In a previous lab study (Boeg Thomsen et al., 2021), such inflexible colour, puppet and order strategies were highly frequent in 2- and 3-year-olds, and as they reflect complete lack of attention to verb meaning, it was important not to award them any points: a raw score of 3 does not reflect better understanding of mental verbs than a raw score of 0. Therefore, any raw score from 0 to 3 was given 0 points. Raw scores of 4 yielded 1 point for nascent attention to mental verbs, and raw scores of 5 to 6 yielded 2 points for stable understanding of mental verbs in complement-clause constructions.

Language, memory and executive functions

To assess to what degree training outcomes would depend on individual differences in general language, memory and executive functions, we tested children's receptive vocabulary, short-term memory, working memory, cognitive flexibility and inhibitory control.

Receptive Vocabulary was assessed with a picture-selection task with progressive difficulty: the *British Picture Vocabulary Scale (BPVS) Third Edition* (Dunn et al., 2009), which requires the child to identify pictures depicting target words among distractor pictures. **Short-Term Memory** was tested with Recall of Digits Forward from the *British Ability Scales II* (Elliott, 1996), a digit-span task with progressive difficulty requiring the child to repeat strings of digits in the same order.

Working Memory was tested with the Missing Scan Task (Roman et al., 2014), a scan-and-retrieve task with progressive difficulty requiring the child to report missing items in sets of increasing size. For ease of administration in the nurseries and pre-schools, we used photos of animals instead of toy animals as items. **Cognitive Flexibility** was tested with the Dimensional Change Card Sort (DCCS, Zelazo, 2006), which tests ability to switch between conflicting rules for sorting (6 items). **Inhibitory Control** was tested with two conflict tasks testing abilities a) to *selectively suppress a dominant response* and b) to *produce a subdominant response* while suppressing a dominant response. Selective suppression was assessed with the Bear/Dragon task (Carlson et al., 2004; Kochanska et al., 1996) where two puppets (in our case a sheep and a crocodile) take turns commanding the child to perform simple actions, and the child must selectively ignore commands from one puppet (5 target items) while following commands from the other (5 control items). Ability to suppress a dominant response to initiate a conflicting subdominant response was tested with the 16-item task Grass/Snow (Carlson & Moses, 2001), which requires the child to point to a photo of snow when the experimenter says "Grass" and to a photo of grass when the experimenter says "Snow".

Coding reliabilities

For our focal experimental measures, False Belief, Complement-Clause Proficiency and Hidden Objects, we calculated interrater reliability. Fifteen percent of both pretest and posttest responses were coded by a second rater who was blind to condition. In the False Belief and Hidden Objects tests, there was no disagreement between raters. For Complement-Clause Proficiency, a weighted Kappa statistic indicated near perfect agreement ($\kappa = 0.95$, weights: squared, utterances: 352, raters: 2, $z = 17.8$).

Training

Training conditions

Children were randomly assigned to one of three training conditions: 1) Simple Clauses, 2) First-Person Complements or 3) Third-Person Complements[3]. In all three conditions, children were exposed to exactly the same training materials and participated in exactly the same training activities, with the only difference between conditions being the experimenter's verbal mediation. In the Simple Clauses condition, children heard simple clauses such as *This is a deer*, and in the complement-clause conditions, these same simple clauses were embedded in complement-clause constructions with first-person subjects (First-Person Complements: *I say that this is a deer*) or third-person subjects (Third-Person Complements: *He says that this is a deer*). All complement-clause constructions in the third-person condition used the subject pronoun *he*, corresponding to the single subject pronoun *I* in the first-person condition, to avoid differences in subject-slot variability.

Training activities

All children participated in nine training tasks, distributed across three types of activities:

1) Exploring and communicating about deceptive and non-deceptive objects (3 tasks)
2) Looking in and communicating about peephole books (3 tasks)
3) Watching and communicating about doll's house hiding scenarios (3 tasks)

[3] For a few children entering the study in the last phase, randomisation was constrained, to ensure that the training groups did not differ significantly on target measures at pretest.

We used three activity types to prevent children from getting bored and to stimulate them to build up general, not task-specific mental-state-reasoning skills. All tasks highlighted the mental states of contrasting perspectives, uncertainty, true and false belief, and knowledge.

The Deceptive and Non-deceptive Objects task was adapted from a false-belief training task developed by Slaughter and Gopnik (1996) and subsequently adapted and used in a line of false-belief training studies (Boeg Thomsen, 2016; Gola, 2012; Lohmann & Tomasello, 2003; San Juan & Astington, 2017). The experimenter showed the child a deceptive object: a bar of soap in the shape of a banana, a pencil sharpener in the shape of a piece of chocolate and an eraser in the shape of a lipstick. The experimenter asked the child what the object was and then demonstrated and verbalised its true identity. The experimenter then introduced a finger puppet who encountered the object, expressed his initial belief about its identity ("I think that that's a nice banana"), explored the object and stated his knowledge ("I know that this is a bar of soap!") and addressed his previous false belief ("I thought that it was a banana!", examples from First-Person Complements condition). For every deceptive object, there was also a non-deceptive object. Here, the interaction structure between experimenter, child and finger puppet was the same, but with statements of belief confirmation instead of belief revision.

The Peephole Books task was adapted from another false-belief training task by Slaughter and Gopnik (1996). The experimenter and the child looked in a peephole book (1. *Who am I? Baby animals*, 2. *Who am I? Farm animals*, 3. *Who am I? Wild animals*) where holes in the pages revealed parts of the animals on the next pages. For each hole, the experimenter first asked the child to guess which animal was coming up and then asked a finger puppet what he would guess. The finger puppet stated a belief and supported it with preliminary evidence ("That's a rabbit. It has a rabbit ear"). Turning the page, the puppet expressed knowledge about the identity of the animal, either confirming or revising his initial belief (“That **was** a rabbit. It had a rabbit ear”, examples from Simple Clauses condition).

The Doll’s House Hiding Game was a newly developed training task partly inspired by Papafragou et al.’s (2007) test of evidentials in Korean children. The experimenter had hidden various objects in a doll’s house (see Figure 1) and asked a doll character, Matthew, to find them. For each object, Matthew’s initial belief was expressed (“He thinks that the socks are under the coffee table”), then he checked that location (“He can see that the socks are NOT under the coffee table”), expressed a new belief (“Then he thinks that the socks are in the piano”), checked the next location, expressed knowledge (“Yes! He knows that the socks are in the piano”) and addressed his initial false belief (“He thought that the socks were under the coffee table”, examples from Third-Person Complements condition). Sometimes Matthew’s initial belief was true and he found the object in the first location, and sometimes he looked in one or two wrong locations (based on false beliefs) before guessing the true location.

Figure 1

Examples of training activities: Deceptive Object, Peephole Book and Doll’s House Hiding Game

All the experimenter’s actions and utterances were scripted; details can be found in the Appendix. The scripts are publicly available on the Open Science Framework project (https://osf.io/vkw63/).

All activities provided children with ample experience with contrast and alignment between mental states, making the use of complement clauses and simple clauses for communication about mental states appropriate and relevant. All activities involved the child's own experience of uncertainty, true and false belief, belief revision and confirmation, and knowledge as well as that of a doll or puppet. The doll's or puppet's mental states were the ones that were verbalised in the target training sentences differing between conditions, i.e. with simple clauses, first-person complements or third-person complements.

Within each of the nine tasks, half of the scenarios presented *belief revision* (i.e., false beliefs that are subsequently revised), the other half *belief confirmation* (i.e., true beliefs that are subsequently confirmed). In this way, we ensured that we were training children to attend to mental states and not simply to learn a low-level strategy of expecting the opposite from reality.

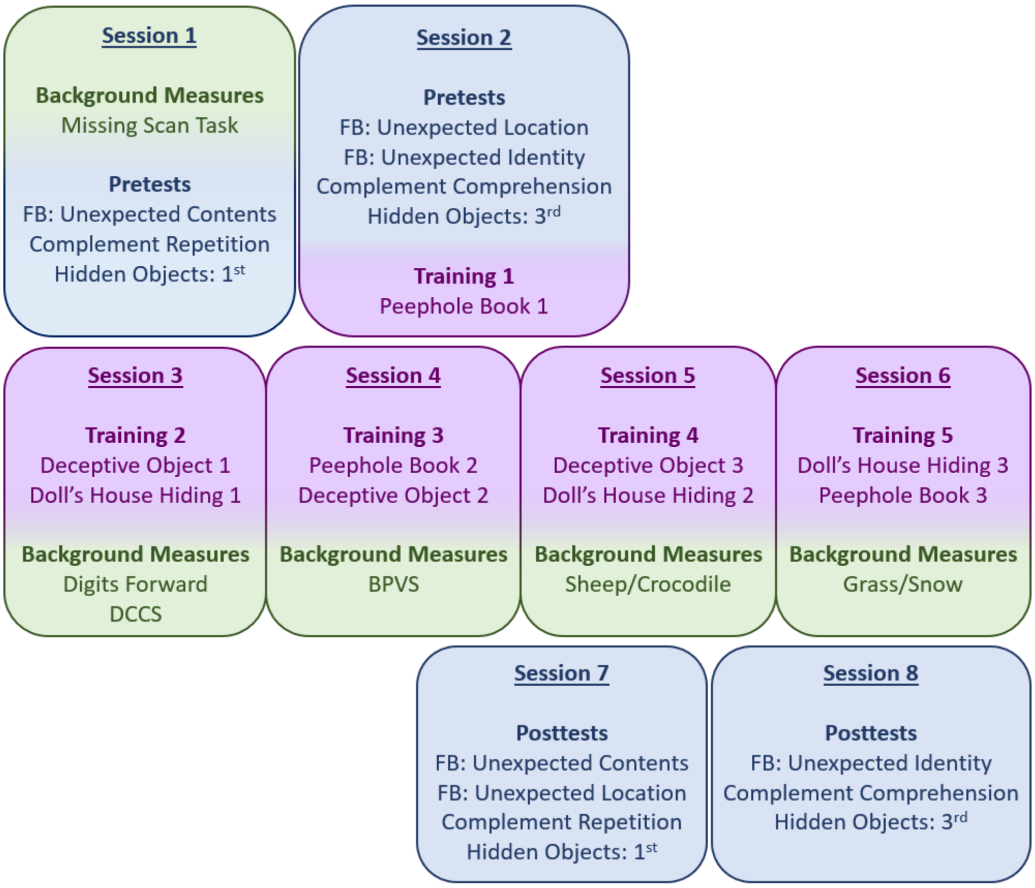
Similarly, to ensure that improvement from pretest to posttest did not simply reflect improvement in performance on trained tasks, structural overlap between training activities and posttests was limited. The Deceptive and Non-deceptive Objects task was similar to the posttest Unexpected Identity, as both involved communication about deceptive objects whose unexpected identities caused belief revision, but the children were not asked any false-belief questions during training, i.e., they were not trained on the format of the posttest question. The other training activities differed substantially from all posttests.

Procedure

Children were tested and trained individually in their nurseries or pre-schools by one of two experimenters (the first and second author). All children received all tests and training tasks in the same fixed order (see Figure 2). The nine training tasks were spread over five different sessions over three weeks. Pretests were administered over two sessions before training, posttests over two sessions after training. Background measures (general language, memory and executive functioning) were administered during the pretest and training phase. In total, there were eight sessions, distributed over four weeks, with two visits per child per week. The first posttest was administered on average 4 days after the last training session. The sessions ranged in length from approximately 15 minutes to 25 minutes, and the mean study duration from session 1 to session 8 was 23 days (duration was included as a control measure in the statistical analysis).

Figure 2

Session overview



Note. FB: False Belief; DCCS: Dimensional Change Card Sort; BPVS: British Picture Vocabulary Scale.

Results

Our primary goals were to test whether children showed more progression in false-belief reasoning in the two complement-clause conditions than in the simple-clause condition when individual differences in memory, executive functioning and general language were controlled, whether there was a difference between being trained with first- vs. third-person complements, and whether training effects were moderated by proficiency with complement clauses prior to training. We evaluate these questions by means of mixed-effects regression analyses.

Before turning to the focal analyses of progression, we present pretest correlations, examine predictors of false-belief reasoning prior to training and check training group comparability. Only the 76 children who completed the training study were included in the focal analyses of progression, and our preliminary analyses were also run on this focal sample.

Pretest analyses

Correlational analyses

The correlation matrix in Table 2 shows that many of our measures of language, social cognition, memory and executive functioning were significantly correlated with each other and with age prior to training[4].

[4] Descriptive statistics for pretest and control measures can be found in Part 3 of the online supplemental material.

Table 2

Spearman's correlations among pretest and control measures

Pretest and Control Measures	1	2	3	4	5	6	7	8	9	10
1. False Belief (total score)										
2. Age (in months)	.18									
3. Short-Term Memory	.14	.19								
4. Working Memory	-.11	**.33 ****	.16							
5. Cognitive Flexibility	.11	.07	.19	-.11						
6. Inhibitory Control (Sheep/Crocodile)	**.28 ***	**.27 ***	**.24 †**	.01	.05					
7. Inhibitory Control (Grass/Snow)	.10	**.40 *****	**.21 †**	.06	.03	**.48 *****				
8. Receptive Vocabulary	**.22 †**	**.35 ****	**.30 ****	.03	.18	**.46 *****	**.54 *****			
9. Complement-Clause Proficiency	.20	**.33 ****	**.61 *****	**.19 †**	**.22 †**	**.31 ***	**.35 ****	**.34 ****		
10. Mental Verbs: First person	-.16	**-.30 ****	-.18	.19	-.07	**-.21 †**	-.12	**-.21 †**	-.13	
11. Mental Verbs: Third person	**.27 ***	.16	-.01	-.16	**.20 †**	**.25 ***	.11	.18	-.05	.01

Note. † $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Prior to training, False Belief (total score: 0-5) was significantly associated with Inhibitory Control (Sheep/Crocodile) and Mental Verbs: Third person, with positive correlations of intermediate strength. There was no significant relationship between Complement-Clause Proficiency and False Belief (r_s .20, p = .11), but for the Comprehension subpart, the positive correlation was significant (r_s = .32, p = .009) [5].

There were significant positive correlations of intermediate strength between Complement-Clause Proficiency and Age, Short-Term Memory, Inhibitory Control (both tasks) and Receptive Vocabulary. Complement-Clause Proficiency was also marginally associated with Working Memory and Cognitive Flexibility. Thus, Complement-Clause Proficiency is correlated to some degree with each of our executive-functioning, memory and general-language measures, demonstrating the importance of testing the influence of these measures when examining relationships between complement-clause acquisition and false-belief reasoning.

Mental Verbs: Third person, i.e., performance with *think* and *know* in complement-clause constructions with third-person subjects, showed positive intermediate correlations with False Belief and Inhibitory Control (Sheep/Crocodile), whereas Mental Verbs: First person was not significantly correlated with False Belief. Mental Verbs: First person was significantly *negatively* associated with Age and – marginally so – with Receptive Vocabulary and Inhibitory Control (Sheep/Crocodile). In the Discussion, we discuss plausible reasons why higher age, larger vocabulary and better inhibitory control should be related to *worse*, not better, performance on Mental Verbs: First person[6].

---

[5] Note that half of the Comprehension items presented true complements. For both types, the correlation with False Belief was significant (true complements: r_s .31, p = .011; false complements: r_s .26, p = .034). Separate correlations for the Comprehension and Repetition subparts of the test can be found in Part 4 of the online supplemental material. In Part 5, we also present posttest correlations. At posttest, Complement-Clause Proficiency and False Belief were significantly correlated (r_s .25, p = .037). For the Comprehension subpart alone, the correlation was still significant (r_s .27, p = .025), and for the Repetition subpart, it now approached significance (r_s .23, p = .059).

[6] Interestingly, these negative correlations disappeared after training where Mental Verbs: First person was positively associated with age, inhibitory control and other measures (see Part 5 of the online supplemental material).

Predictors of pretest False Belief

To examine predictors of false-belief reasoning *before* intervention, we fitted a generalised linear mixed-effects regression model to the dependent variable Pretest False Belief (at item level, binary-coded: 0-1) in the statistical environment R (version 3.6.2, R Core Team 2019), using the package lme4 (Bates et al. 2015). We started from the full model including age, gender and all our language, memory and executive-functioning measures as fixed effects and participant and item as random effects. We then fitted the model using the principle of backwards selection and a significance-based approach (Gries, 2013), and after stepwise pruning of the model of non-significant predictors, the final model (Table 3) included only two significant predictors: Pretest Complement-Clause Proficiency and Pretest Mental Verbs: Third person.

Table 3

Summary of Generalised Linear Mixed-Effects Model fitted to the dependent variable Pretest False Belief

Random effects			
Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	< 0.0001	< 0.0001
Item	(Intercept)	0.6193	0.7870

Number of observations: 373, Participants: 76, Items: 5

Fixed effects				
	Estimate	Std. Error	*z* Value	Pr(>|*z*|)
(Intercept)	-1.23	0.44	-2.781	0.0054 **
Pretest Complement-Clause Proficiency	0.05	0.02	2.149	0.0316 *
Pretest Mental Verbs: Third person	0.43	0.17	2.487	0.0129 *

Significance codes: '**' .01; '*' .05

As a preliminary piece of evidence, the mixed-effects analysis thus replicates previous findings of close relationships between complement clauses and mental verbs used with third-person complements on the one hand and false-belief reasoning on the other. While the concurrent correlations are compatible with the hypothesis that acquisition of complement clauses supports false-belief reasoning, they cannot inform us about direction of causality.

Equivalence of training groups

To examine training group equivalence, we compared the groups pairwise on all our predictor measures: age, gender, duration (days from first to last session) and performance on all pretests. Group means as well as results from the comparisons are presented in Table 4. There were no significant differences between any of the three groups on any of these measures.

Table 4

Training group characteristics and group comparisons

	Simple (*n* = 25)		First (*n* = 25)		Third (*n* = 26)			Simple-First		Simple-Third		First-Third	
	Mean	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	**Test**	***t/U/χ2***	***p***	***t/U/χ2***	***p***	***t/U/χ2***	***p***
False Belief (total score)	1.96	(1.11)	1.95	(1.09)	2.00	(1.02)	Mann-Whitney	256.5	.94	249	.93	233.5	.84
Age (months)	40.56	(2.89)	40.52	(2.37)	40.54	(2.69)	*t*-test	.05	.96	.03	.98	-.03	.98
Short-Term Memory	8.96	(5.34)	8.52	(3.25)	8.73	(3.42)	*t*-test	.35	.73	.18	.86	.35	.73
Working Memory	2.54	(1.14)	2.28	(0.74)	2.27	(1.12)	Mann-Whitney	333.0	.47	357.5	.35	341.5	.74
Cognitive Flexibility	1.20	(2.31)	1.40	(2.45)	0.96	(2.12)	Mann-Whitney	299.5	.75	311	.77	323.5	.54
Inhibitory Control (Sheep/Crocodile)	4.00	(1.64)	3.86	(1.88)	3.14	(2.12)	Mann-Whitney	217.5	.72	269.5	.31	288	.23
Inhibitory Control (Grass/Snow)	10.13	(4.88)	10.71	(4.18)	8.68	(4.56)	Mann-Whitney	277.5	.84	359.5	.24	378.5	.12
Receptive Vocabulary	39.80	(13.21)	39.00	(12.20)	37.54	(11.81)	Mann-Whitney	323.5	.84	360.5	.51	369.5	.41
Complement-Clause Proficiency	10.36	(5.77)	9.76	(4.44)	9.12	(5.19)	*t*-test	.41	.68	.81	.42	.48	.64
Mental Verbs: First person	0.24	(0.44)	0.36	(0.64)	0.38	(0.64)	Mann-Whitney	294	.65	297	.51	317	.86
Mental Verbs: Third person	0.40	(0.58)	0.40	(0.65)	0.65	(0.75)	Mann-Whitney	320	.87	268	.23	263.5	.19
Gender (boys; girls)		13; 12		13; 12		12; 14	Chi Square	.00	1.00	.17	.68	.17	.68
Duration in days (Ses. 1-Ses. 8)	22.88	(6.15)	24.04	(4.60)	22.08	(3.93)	Mann-Whitney	271	.42	357.5	.54	399	.16

Note. **Simple**: group trained with simple clauses, **First**: group trained with first-person complements, **Third**: group trained with third-person complements.

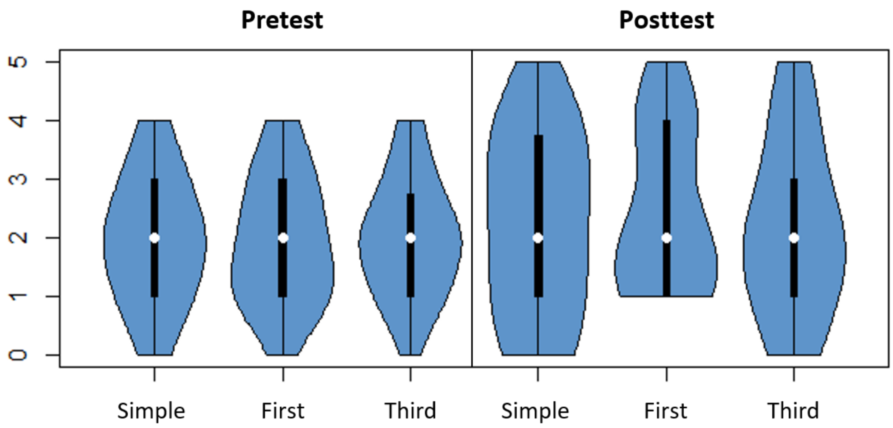
False-belief progression

Pretest-posttest comparisons (total score)

At pretest, children on average got 1.97 out of five false-belief questions correct (*SD* = 1.06, range: 0–4). At posttest, three weeks later, children on average got 2.29 out of the five questions correct (*SD* = 1.48, range: 0–5), with an average change score of 0.38 (*SD* = 1.53, range -3–4). Figure 3 illustrates pretest-posttest performance in the three conditions. There was considerable within-group variation in training outcome for all conditions, with wide ranges of change scores (Simple Clauses: -3–3; First-Person Complements: -2–4; Third-Person Complements: -2–3). This variation in training outcomes shows the importance of disentangling the effects of condition and of individual differences on false-belief progression, and to this end, we now turn to regression analysis.

Figure 3

False Belief scores (total score) in the three training conditions at pretest and posttest



Note. Violin plot showing density curves (blue), median scores (white dots) and interquartile ranges (black boxes) for composite False Belief scores (0-5) at pretest and posttest in the three training conditions.

Predictors of false-belief progression

Our main questions were whether children trained with complement clauses would advance more in false-belief reasoning than children trained with simple clauses when individual differences in executive functioning, memory and general language were controlled, and whether the effects of complement-clause training would be moderated by pretest proficiency with complement clauses. We evaluated these questions using linear mixed-effects regression analysis. Our dependent variable was advancement in false-belief reasoning, operationalised as the change score at item level (from -1 to 1), i.e. for each item the result of subtracting the binary-coded pretest score (0-1) from the corresponding binary-coded posttest score (0-1) for test questions with the same structure, but different contents.

We fitted a linear mixed-effects regression model to the dependent variable False Belief: Change Score in the statistical environment R (version 3.6.2, R Core Team 2019), using the packages lme4 (Bates et al. 2015) and lmerTest (Kuznetsova et al., 2017). We started by fitting the full model, which besides random effects of Participant and Item included the following fixed effects: Pretest False Belief (binary-coded), Duration (days from first to last session), Age (in months), Gender, Short-Term Memory, Working Memory, Cognitive Flexibility, Inhibitory Control (Sheep/ Crocodile), Inhibitory Control (Grass/Snow), Receptive Vocabulary, our experimental manipulation Condition and three measures of pretest proficiency with complement clauses: Complement-Clause Proficiency, Mental Verbs: First Person, Mental Verbs: Third Person. For these three pretest complement-clause variables, we examined interactions with training condition, based on the hypothesis that children who were already proficient with complement clauses at pretest would benefit less from complement-clause training.

Since we were interested in the effects of complements training, our model takes performance in the Simple Clauses condition as its reference level and tests whether False Belief change scores are higher in either of the two complements-training conditions. Starting from the full model and following the principle of stepwise backwards regression and a significance-based approach (Gries 2013), we fitted the model by successively testing whether eliminating the least significant predictor or interaction would make the model significantly worse, and deleting it from the model if it would not. Table 5 summarises the final model.

Table 5

Summary of Linear Mixed-Effects Model fitted to the dependent variable False Belief: Change Score

Random effects			
Groups	Name	Variance	Std. Dev.
Participant	(Intercept)	< 0.0001	0.0002
Item	(Intercept)	0.0205	0.1431
Residual		0.1898	0.4357

Number of observations: 314, Participants: 64, Items: 5

Fixed effects	Estimate	Std. Error	df	*t* Value	Pr(>|*t*|)
(Intercept)	0.17	0.15	67.83	1.194	.2365
Pretest False Belief	-0.91	0.05	301.93	-16.669	< .0001 ***
Inhibitory Control (Sheep/Crocodile)	0.06	0.02	298.51	3.776	.0002 ***
Working Memory	-0.07	0.03	298.46	-2.674	.0079 **
Receptive Vocabulary (scaled)	0.08	0.03	298.42	2.947	.0035 **
Pretest Complement-Clause Proficiency	0.02	0.01	298.51	1.804	.0723 †
Pretest Mental Verbs: First person	-0.11	0.04	298.41	-2.595	.0099 **
Condition: First-person Complements	0.35	0.15	298.43	2.331	.0204 *
Condition: Third-person Complements	0.21	0.14	298.50	1.425	.1553
Condition: First-person Complements × Pretest Complement-Clause Proficiency	-0.03	0.01	298.41	-2.130	.0340 *
Condition: Third-person Complements × Pretest Complement-Clause Proficiency	-0.01	0.01	298.44	-0.734	.4633

Effect size (whole model) R^2marginal = 0.50
R^2conditional = 0.55

Significance codes: ‘***’ .001; ‘**’ .01; ‘*’ .05; ‘†’ 0.1

Note: The intercept represents the *Simple Clauses* level of the factor *Condition* as reference level, while the other two levels, *First-person Complements* and *Third-person Complements*, are represented in the model. The estimate for *Condition: First-person Complements* (0.35) represents the estimated advantage of being trained with first-person complements relative to being trained with simple clauses. The number of participants in the final model is 64 due to missing data points from one participant on the Working Memory test and eleven participants on the Inhibitory Control (Sheep/Crocodile) test.

There was a significant negative main effect of Pretest False Belief, with children with lower initial levels of false-belief reasoning advancing more than children with higher initial levels of false-belief reasoning. We also found significant positive main effects of Inhibitory Control (Sheep/Crocodile) and Receptive Vocabulary as well as significant negative main effects of Working Memory and Pretest Mental Verbs: First Person. We return to these effects in the discussion.

For our main question, we found a significant positive effect of Condition: First-person Complements ($\beta = 0.35$, $t = 2.331$, $p = .0204$) which was moderated by a significant interaction with Pretest Complement-Clause Proficiency ($\beta = -0.03$, $t = -2.130$, $p = .0340$). As hypothesised, this interaction was negative, indicating that the better children were performing with complement clauses before training, the less they benefitted from being trained with first-person complements, whereas the more proficient children in the simple-clauses condition were with complement clauses before training, the more they benefitted from training. For children trained with third-person complements, there was not a significant advantage relative to being trained with simple clauses, and also no significant interaction with pretest proficiency.

To compare the two complement-clause conditions directly, we also fit the model taking Condition: First-person Complements as its reference level. The model showed no difference between the complement-clause conditions (Condition: Third-person Complements, $\beta = -0.15$, $t = -1.059$, $p = .2903$).

Interplay between levels of complement-clause proficiency and training condition

The main analysis showed an interaction between Condition and Pretest Complement-Clause Proficiency. Since our complement-clause test was constructed to capture fine-grained variation at different stages of development, we were able to delve into this interaction by conducting follow-up analyses to explore whether variation at specific levels of complement-clause proficiency interacted differentially with training condition. We therefore fitted exploratory linear mixed-effects regression models to the dependent variable False Belief: Change Score (item level), following the same principles of model selection as in the main analysis and including exactly the same predictors, with one exception: substituting Pretest Complement-Clause Proficiency (total test score) with subscores from each of the levels Basic, Middle and Advanced.

The Basic level tested proficiency with early-acquired constructions with high-frequency complement-taking verbs in high-frequency strings with first- and second-person subjects. The final model for the analysis testing the influence of Basic Complement-Clause Proficiency is almost identical to the one in the main analysis (see Part 6 of the online supplemental material). Again we see a significant interaction between pretest complement-clause proficiency and training condition: The poorer children performed at this basic level of complement-clause proficiency at pretest the more they benefitted from being trained with first-person complements (β = -0.07, t = -2.865, p = .0045), and children trained with first-person complements progressed significantly more than children trained with simple clauses (β = 0.24, t = 2.465, p = .0143).

The Middle level tested proficiency with constructions with high-frequency complement-taking verbs in low-frequency clauses with third-person subjects. At this middle level, there was no interaction between pretest complement-clause proficiency and training condition.

The Advanced level tested proficiency with late-acquired constructions with low-frequency complement-taking verbs in low-frequency clauses with third-person subjects. The final model for the analysis testing the influence of Advanced Complement-Clause Proficiency (see Part 6 of the online supplemental material) again shows a significant interaction between pretest complement-clause proficiency and training condition, but interestingly, at this advanced level, the significant relationship is found with third-person complements. The poorer children performed at this highly abstract level of complement-clause proficiency at pretest the more they benefitted from being trained with third-person complements (β = -0.08, t = -2.636, p = .0088), and children trained with third-person complements progressed significantly more than children trained with simple clauses (β = 0.32, t = 2.508, p = .0127).

While these analyses are exploratory, they suggest an interplay between level of complement-clause proficiency and the type of linguistic input that is especially useful for promoting false-belief development, with first-person complements driving development at early stages of complement-clause acquisition, third-person complements at later, more advanced stages.

Progression in complement-clause proficiency and belief reasoning

To ascertain whether progression in false-belief reasoning was related to advances in complement-clause proficiency for children trained with complement clauses, we assessed complement-clause proficiency at both pretest and posttest with the 16-item Complement-Clause Proficiency test, presenting items with the same structure but new contents at posttest. We operationalised advancement in complement-clause proficiency as a change score (posttest performance minus pretest performance) and fitted linear-mixed effects models to the dependent variable False Belief: Change Score (item level), including the predictor Complement-Clause Proficiency: Change Score as well as all control measures. For children trained with complement clauses, progression in complement-clause proficiency significantly predicted progression in false-belief reasoning (β = 0.03, t = 2.058, p = .045).

Discussion

The main goal of this training study was to investigate the hypothesis that children's acquisition of complement-clause constructions promotes sociocognitive development. Specifically, we tested whether typically developing English-speaking three-year-olds would show greater advances in false-belief reasoning if they had perspective-rich activities mediated linguistically with complement clauses compared to having the same activities mediated with simple clauses. Our results strongly support this hypothesis of a causal influence from language on social cognition: children trained with complement clauses progressed significantly more in false-belief reasoning than children trained in the maximally similar training condition without complement clauses. The current study yields more solid evidence for a causal relation than previous training studies, as it controlled individual differences in memory, executive functioning and general language, avoided the use of complement clauses in all false-belief tests as well as in the control condition and tested the influence of pretest proficiency with complement clauses on receptiveness to training.

Our study also examined whether experience with third-person complements provides children with more sociocognitive support than experience with first-person complements. In an internal comparison between the two complement-clause conditions, the results showed no significant difference, but compared to the baseline of children trained with simple clauses, only children trained with first-person complements progressed significantly more. Contrary to expectations, this result indicates that experience with first-person complements is more beneficial for sociocognitive development than experience with third-person complements. Exploratory follow-up analyses suggested that first-person complements may be especially useful for false-belief development at a more basic stage of complement-clause acquisition, third-person complements at a more advanced stage.

Below, we discuss the influence from complement-clause acquisition on sociocognitive development, the unexpected advantage of first-person complements and the potential interplay with level of complement-clause proficiency, the role of individual differences and future directions.

Complement clauses in false-belief development

Compared to children trained *without* complement clauses, children trained with first-person complements advanced significantly more in false-belief reasoning, and, crucially, progression was moderated by an interaction between proficiency with complement clauses *before training* and the linguistic input children experienced *during training*. The negative interaction between pretest proficiency and training with first-person complements indicates that – as hypothesised – children with initially weaker complement-clause schemas profited more from having perspective-rich situations spelled out linguistically with complements than children who were already proficient with the construction beforehand. The pattern was the opposite in the simple-clauses condition: Here, children who were already proficient with complement clauses beforehand advanced more in their false-belief reasoning, suggesting that these children were utilising their already robust schemas to profit from the perspective-rich training input. As a further confirmation of the supportive role of complement clauses in false-belief development, a follow-up analysis demonstrated that progression in complement-clause proficiency predicted progression in false-belief development for children trained with complement clauses.

Presenting strong support for a causal influence from complement-clause acquisition on sociocognitive development, our results replicate five training studies where complement-clause training was found to promote false-belief development, but where use of complement clauses in false-belief tests and other confounds made the conclusions preliminary (Gola, 2012; Hale & Tager-

Flusberg, 2003; Lohmann & Tomasello, 2003; Mo et al. 2014, San Juan & Astington, 2016). Our results also converge with training studies in atypical development where false-belief progression after complement-clause training has been demonstrated using false-belief tests without complement clauses (Boeg Thomsen, 2016; Durrleman et al., 2019, 2022a, 2022b; Durrleman & Delage, 2020). Further, the results align with conclusions from longitudinal studies of natural development where children's proficiency with complement-clause constructions was found to predict their later false-belief reasoning (Boeg Thomsen et al., 2021; de Villiers & de Pyers, 2002), or where variation in maternal mental-state utterances, including complement-clause constructions, predicted variation in children's later theory of mind (Ruffman et al., 2002).

Complement-clause constructions are a crosslinguistically widespread grammatical tool for making invisible relations between persons and ideas the focus of joint attention, and there are good reasons why acquisition of this flexible and explicit perspective-marking construction should support mental-state reasoning abilities. First, repeated exposure to the same linguistic structures (subject + verb + clause) across a variety of contexts is likely to invite children to compare contexts to identify and abstract structural similarities, and complement-clause constructions may thus act as linguistic cues to recognising and generalising over recurring perspectival relations between persons and ideas (San Juan & Astington, 2017, cf. Bowerman & Choi, 2003). Second, as demonstrated by crosslinguistic experiments in the *thinking for speaking* tradition (Slobin 1996), acquisition of the grammar of one's native language affects routine attention to and memory for those aspects of experience that grammar will recurringly require one to communicate about. Thus, children's acquisition of complement clauses can be expected to stimulate them to pay routine attention to and remember mental aspects of experience.

First- vs. third-person complements

Our study also investigated whether exposure to third-person complements is more likely to promote false-belief development than first-person complements. Surprisingly, we found that only children trained with first-person complements progressed significantly in false-belief reasoning compared to children trained with simple clauses, whereas children trained with third-person complements did not.

The positive effect of first-person complements is not in itself surprising. As argued by Verhagen (2005) and Boeg Thomsen (2016), by pointing to a human conceptualiser (the subject *I*), these constructions perform the central function of complement-clause constructions: anchoring ideas in persons, in this case the speaker. Providing empirical support for a relation with social cognition, Moore et al. (1990) reported concurrent correlations between false-belief performance and comprehension of first-person complements. While some experiments had reported an advantage of third-person over first-person complements, their conclusions were uncertain due to design flaws (Gola, 2012) or due to the use of the Hidden Objects task (Boeg Thomsen et al., 2021; Brandt et al., 2016) which is better suited pragmatically for measuring understanding of third- than first-person complements. Comparing the impact of experience with first- and third-person complements on false-belief progression directly, the current study transcends the uncertainties from previous studies and refutes the hypothesis that third-person complements are generally more beneficial than first-person complements for sociocognitive development.

Nevertheless, while it is neither unexpected that first-person complements do support false-belief reasoning nor that third-person complements are not *more* beneficial, we did not expect third-person complements to be *less* beneficial than first-person complements. Here, it is important to stress again that in a direct comparison, there was no significant difference between the two complement conditions – but at the same time there was also no significant advantage of third-person complements relative to the baseline simple-clause condition.

There are two likely (not mutually exclusive) explanations why the main analysis only found an effect of first-person complements, one pertaining to frequency, one to perspectival complexity and stages in complement acquisition. First, as substantiated by a corpus analysis of the Manchester CHILDES corpus (Theakston et al., 2001), complement-clause constructions with first-person subjects are much more frequent than third-person complements in child-directed speech. Therefore, it is likely that children who heard first-person complements during training had richer opportunities to generalise their training experience to everyday contexts during the training weeks, with the high-frequency first-person complements acting as hooks between training and everyday interaction. Further, the higher frequency of first-person complements may make them quicker for children to process, freeing processing resources for other aspects of the complex training activities. This is all the more likely, as differences in frequency have previously been found to moderate the supportive influence of complement-clause constructions: Mandarin-speaking children trained with complement clauses with high-frequency communication verbs progressed more in false-belief reasoning than children trained with low-frequency mental verbs (Mo et al. 2014). A second probable explanation is that first- and third-person complements play their key role at different stages, with complements presenting speaker perspective primarily supporting social cognition at initial stages of complement-clause acquisition, and third-person complements promoting development at more advanced stages. The advantage of first-person complements at early stages could be explained by their simpler perspectival structure as comprehending them only requires children to represent one perspective in addition to their own, the speaker's, whereas comprehension of third-person complements arguably requires children to represent two additional perspectives (the speaker's and the one represented by the third-person subject, cf. van Duijn &

Verhagen, 2018). Our follow-up analyses yield preliminary support for this interpretation by showing a significant effect of third-person training in an analysis testing the influence of pretest complement-clause proficiency at the most abstract level, in contrast to a significant effect of first-person training in the analysis testing the influence of pretest complement-clause proficiency at the most basic level. These exploratory results suggest that studies comparing the influence of different complement-clause constructions could profit from asking not only *whether* specific construction types support false-belief development, but also *when*, at which stages of development, they do so.

Evaluating first-person complements with the Hidden Objects task

One final issue regarding the role of first- and third-person complements is the effect of using the Hidden Objects task to compare them. Evaluating mental-verb understanding in complement-clause constructions with this task, both Boeg Thomsen et al. (2021) and Brandt et al. (2016) found significant relations between false belief and third-person, but not first-person complements, and indeed our analysis of *pretest* false belief replicated this finding: only performance with third-person complements on the Hidden Objects task predicted false belief. However, as argued by Boeg Thomsen et al. (2021), the Hidden Objects task is not ideal for first-person complements because it depends on *certainty* contrasts, and there is no clear-cut certainty contrast for first-person complements with the mental verbs *think* and *know* in child-directed speech (Howard et al., 2008). Our pretest correlations highlight the complexities involved in using the Hidden Objects task for evaluating first-person complements. If the task measured understanding of first-person complements reliably, we would expect a positive association with age, but on the contrary, we found a significant *negative* association: the older children were, the more often they trusted the speaker saying *I think* over the speaker saying *I know*. A possible explanation is that, as children get older, they begin to evaluate speaker credibility and politeness and potentially interpret statements with *I know* (in uncertain contexts like the sticker-hiding game) as signs of cocksureness and *I think* as marking a more reflected speaker stance. This is only a speculative interpretation, but it finds some support in the two other – marginally significant – negative correlations: with Receptive Vocabulary and Inhibitory Control. Thus, the more advanced vocabularies children have *and* the better they are at inhibiting spontaneous responses, the more likely they are to trust the statement marked with *I think* over the statement marked with *I know*. These findings suggest that performance with first-person complements on the Hidden Objects task cannot be taken as a pure measure of children's understanding of the different degrees of certainty marked by *think* vs. *know*, but may also tap into children's growing understanding of credibility and politeness, with lower scores potentially reflecting pragmatic maturity[7]. This also makes it difficult to interpret the negative main effect of Mental Verbs: First person in our main analysis of false-belief progression. The straightforward interpretation would be that the poorer understanding children have of mental verbs expressing speaker perspective, the more they benefit from being trained with scenarios with conflicting perspectives. Nevertheless, since the negative pretest correlations indicate that poorer performance may actually reflect more advanced pragmatic understanding, we cannot say anything conclusive about this effect.

Individual differences: executive functioning and language

Besides the role of complement clauses in false-belief development, our results also cast light on the role of individual differences, showing clear effects of inhibitory control, receptive vocabulary and

[7] Interestingly, all negative correlations for the first-person version of Hidden Objects disappeared after training, where performance was positively correlated with Complement-Clause Proficiency, Age, Inhibitory Control and Short-Term Memory (see Part 5 of the online supplemental material). Now, there was further a marginally significant positive correlation between performance on the first- and third-person versions. This indicates that after training – and perhaps after experiencing the third-person version of the task (administered after the first-person version at pretest), children understood that the task was to compare the certainty level of the two mental verbs and used their abilities to do so, thus making posttest performance a valid measure of understanding of mental verbs in first-person complements.

working memory on training outcomes. The better children were at selectively suppressing a dominant response in the Sheep/Crocodile inhibitory control task (the Bear/Dragon task, cf. Carlson et al., 2004; Kochanska et al., 1996), the more they profited from training ($\beta = 0.06$, $t = 3.776$, $p = .0002$). This result replicates Benson et al.'s (2013) finding that performance on response-conflict executive-functioning tasks predicted ability to profit from false-belief training, and it corroborates the hypothesis that children's ability to suppress their own perspective supports their abilities to learn from situations with contrasting perspectives (Benson et al., 2013; Carlson et al., 2004). The result also aligns with longitudinal studies of involvement of executive functions in Theory of Mind development (Carlson et al., 2004; Hughes & Ensor, 2007).

Receptive vocabulary, as measured by the *BPVS* (Dunn et al., 2009), also predicted receptiveness to false-belief training ($\beta = 0.08$, $t = 2.947$, $p = .0035$), aligning with results from the training studies by Benson et al. (2013) and Gola (2012). The facilitative relation can be interpreted in two ways: On the one hand, better lexical skills may directly support ability to learn from training, by allowing children to process the linguistic input faster and more easily. On the other hand, better lexical skills may reflect better underlying sociocognitive skills, since efficient word learning depends on good intention-reading skills, in which case it is not – or not only – children's vocabularies *per se* that support ability to profit from training, but (also) the better intention-reading skills they represent.

Finally, we found a negative effect of working memory ($\beta = -0.07$, $t = -2.674$, $p = .0079$), measured by the Missing Scan Task (Roman et al., 2014), indicating that children with poorer abilities to actively retain and manipulate multiple items in working memory profited more from training than children with better working memory. At a first glance, this finding is surprising, given that we would expect children with better executive functions to learn faster, but a likely explanation is that children with poorer working memory may be much more challenged than their peers in learning from complex social situations in everyday life. Such children may otherwise be sociocognitively mature and ready to learn about false beliefs, but be hindered by their poor working memory in extracting and generalising over complex information in their natural environment, and the training sessions with one-on-one interaction and recurring activity types may have provided them with advantageous learning contexts.

Future directions

Our follow-up analyses showed a significant effect of third-person complements at the most advanced level of complement-clause proficiency and of first-person complements at the most basic level. As these analyses were exploratory, they should be followed up with new studies investigating directly whether constructions anchoring ideas in the speaker are indeed more beneficial for sociocognitive development at the earliest stages of complement-clause acquisition, and third-person constructions more helpful at more advanced stages.

A second important question concerns the role of constructional variability. Hearing many different *types* of complement-clause constructions would be expected to support generalisation and abstraction more than hearing the same amount of *tokens* representing few types, and more general schemas for complement clauses could in turn be hypothesised to yield stronger support for sociocognitive development. Our complement conditions presented children with a variety of complement-taking verbs to support generalisation, but the subject slot was invariable (*I* or *he*). The effect of complement-clause training on false-belief progression may therefore be even stronger in conditions with more natural subject variability, and follow-up studies could compare progression in conditions with invariable subject slot to conditions with high subject variability.

Finally, knowledge about the advantages of using complement-clause constructions to communicate about perspective-rich situations could potentially be useful for practitioners aiming to support sociocognitive development in everyday contexts, and new studies could examine to which degree our experimental results could be usefully implemented in practice. Given that we conducted training in daycares with background noise, in the midst of other activities, the study has higher ecological validity than a lab study, but new studies could investigate whether our activities designed for one-on-one interaction could be adapted to group formats more suitable for daycares, and whether the beneficial effects would also be seen when activities were led by staff.

Conclusions

Developing robust and flexible abilities to remember, imagine and reason about one's own and others' beliefs is an important aspect of preschool children's sociocognitive development, helping them to understand and navigate their social worlds. Within typical development, the age at which children begin to pass false-belief tests varies substantially, and one of the factors likely to explain this variation is differences in children's exposure to complement-clause constructions, as these are flexible and explicit grammatical constructions dedicated to making invisible relations between persons and ideas the focus of joint attention. Addressing uncertainties in previous training studies, the current study provides solid evidence for a causal influence of complement-clause exposure on sociocognitive development. Controlling individual differences in complement-clause proficiency, general language, memory and executive functioning and avoiding the use of complement clauses in false-belief tests, we found that English-speaking three-year-olds trained with complement clauses progressed significantly more in false-belief reasoning than children trained in a maximally similar condition without complement clauses. Importantly, having perspectives spelled out linguistically with complements was especially beneficial for children with initially weaker complement-clause schemas. Our results also clearly show that complement-clause exposure is only one factor among many in false-belief development, as we found significant effects of inhibitory control, working memory and receptive vocabulary too. Finally, our results cast new light on the role of subject person in complement-clause constructions. Previous studies had suggested a less supportive influence from first-person complements, but investigating causality stringently, we found a clear effect of exposure to first-person complements on false-belief progression, indicating that it is highly useful for children to have the perspective of the speaker explicated linguistically in situations with conflicting perspectives. For third-person complements our results are less clear: in the main analysis, they were not significantly more beneficial than simple clauses, but as suggested by our follow-up analysis, they may play their key role at more advanced stages of development.

References

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

Bates, D., Maechler, 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

Benson, J.E., Sabbagh, M.A., Carlson, S.M., & Zelazo, P.D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*(9), 1615-1627.

Boeg Thomsen, D. (2015). Ontogenetic paths to the parenthetical construction. In S. Schneider, J. Glikman & M. Avanzi (Eds.), *Parenthetical verbs*. *Linguistische Arbeiten,* 557, (pp. 189-222). Berlin: De Gruyter.

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

Boeg Thomsen, D., Kandemirci, B., Theakston, A. L., & Brandt, S. (2024, May 22). Complement clauses and false belief: a training study with English-speaking three-year-olds. Retrieved from osf.io/vkw63

Boeg Thomsen, D., Theakston, A., Kandemirci, B., & Brandt, S. (2021). Do complement clauses really support false-belief reasoning? A longitudinal study with English-speaking 2- to 3-year-olds. *Developmental Psychology, 57*(8), 1210-1227.

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). Cambridge, Massachusetts: The 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,* 131–43. 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,* 9. http://dx.doi.org/10.5334/joc.10

Carlson, S.M, Louis, J.M., & 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*, 299-319.

Carlson, S.M., & Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032–1053.

de Villiers, J. G., & Pyers, J.E. (1997). Complementing cognition: the relationship between language and theory of mind. *Proceedings of the annual Boston University Conference on Language Development, 21,* 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*, 1037-1060.

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

Duijn, M.v., & A. Verhagen. (2018). Beyond triadic communication: A three-dimensional conceptual space for modelling intersubjectivity. *Pragmatics & Cognition, 25*(2), 384–416. https://doi.org/10.1075/pc.18004.dui

Dunn, L., Dunn, D., Styles, B., & Sewell, J. (2009). *British Picture Vocabulary Scale (BPVS) Third Edition*. GL Assessment.

Durrleman, S., Bentea, A., Prisecaru, A., Thommen, E., & Delage, H. (2022a). Training syntax to enhance Theory of Mind in children with ASD. *Journal of Autism and Developmental Disorders* https://doi.org/10.1007/s10803-022-05507-0

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., & Delage, H. (2020). Training complements for belief reasoning in developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(6), 1861–1877. https:// doi.org/ 10. 1044/ 2020_ jslhr- 19- 00075

Durrleman, S., Dumont, A., & Delage, H. (2022b). Syntactic strategy training for Theory of Mind in deaf children. *Journal of Deaf Studies and Deaf Education 27*(1), 89–100. https://doi.org/10.1093/deafed/enab034

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

Gola, A.H. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development, 27,* 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,* 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*, 493– 498. http://dx.doi.org/10.1111/2041-210X.12504

Gries, S.T. (2013). *Statistics for Linguistics with R: A Practical Introduction*, 2nd ed. Berlin/Boston: de Gruyter.

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

Hansen, M.B. (2010). If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology, 1*(23), 1–7. https://doi.org/10.3389/fpsyg.2010.00023

Harrell. F.E. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. New York: Springer Science and Business Media.

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,* 375–402. https://doi.org/10.1177/0142723708091044

Hughes, C., & R. Ensor. (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

Kochanska, G., Murray, K.T., 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

Kumle, L., Võ, M.LH., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

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

Lancashire County Council. (2011). Lancashire's population by ethnicity, 2011 Census of Population: https://www.lancashire.gov.uk/media/903536/census-2011-districts-ethnicity.pdf

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

Manchester City Council. (2011). Ethnic Group Summary: 2011 Census of Population: https://secure.manchester.gov.uk/download/downloads/id/25531/q05b_2011_census_summary_-_ethnic_group.pdf

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

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253–299. https://doi.org/10.1016/j.cognition.2006.04.001

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

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL 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., & 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

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734–751. https://doi.org/10.1111/1467-8624.00435

San Juan, V. & Astington, J.W. (2017). Does language matter for implicit theory of mind? The effects of epistemic verb training on implicit and explicit false-belief understanding. *Cognitive Development 43*, 19-32. DOI: http://dx.doi.org/10.1016/j.cogdev.2016.12.003.

Slaughter, V. & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*, 2967-2988. https://doi.org/10.2307/1131762

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.

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, 115*(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

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

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

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

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

False-belief tests

Unexpected Location builds on Wimmer and Perner (1983) and Baron-Cohen et al. (1985), testing ability to predict another's false belief about the location of an object that has been moved from its original location without the protagonist witnessing this transfer. The task format is narrative, with the experimenter acting out the short story with dolls and props (pretest: mum moves teddy bear from toy box to pram; posttest: dad moves carrot from bag to pot). The child is posed a reality control question (current object location), a memory control question (original object location) and a target false-belief question (Other question).

Unexpected Contents builds on Hogrefe et al. (1986) and Perner et al. (1987), presenting the child with a familiar container with unexpected contents (pretest: ball in raisin box, posttest: spoon in crayon box). Having discovered its surprising contents, the child is posed a reality control question (true contents) and two false-belief questions testing ability to predict another's false belief about the contents of the container (Other question) and to remember their own previous false belief (Self question).***Unexpected Identity*** builds on Gopnik and Astington (1988) and presents the child with a deceptive object (pretest: apple-shaped candle, posttest: flower-shaped pen). Having discovered its surprising identity, the child is posed a reality control question (true identity) and two false-belief questions testing ability to remember their own previous false belief about the object (Self question) and to predict another's false belief (Other question).

Training scripts

All the experimenter's actions and utterances were scripted, with scripts consisting of training parts and framing parts. The training parts were the sentences that differed between conditions by verbalising mental states with either simple clauses, first-person complements or third-person complements. The framing parts were identical across conditions. These sentences were used for introducing and framing activities and for keeping children engaged and attentive with simple-clause questions, such as *Who's hiding here, [child]? Where were the socks, [child]?*

To keep sentences in all conditions pragmatically felicitous, the training parts only presented declarative clauses. In the complement-clause conditions, the training sentences presented a variety of high-frequency and low-frequency complement-taking verbs designating cognition (*think*, *know*, *believe*, *realise*), utterance (*say*) and perception (*see*, *smell*, *feel*, *notice*), in both present and past tense. All complement clauses were introduced with the complementizer *that,* formally marking them as embedded clauses.