Title:

Putting Complement Clauses into Context: Testing the effects of story context, false-belief understanding, and syntactic form on children's and adults' comprehension and production of complement clauses

Authors:

Silke Brandt (Lancaster University, Department of Linguistics and English Language)

Stephanie Hargreaves (University of Manchester, School of Health Sciences)

Anna Theakston (University of Manchester, School of Health Sciences)

Correspondence concerning this article should be addressed to:

Silke Brandt

Department of Linguistics and English Language

County South

Lancaster University

Lancaster, UK, LA1 4YL

Email: s.brandt@lancaster.ac.uk

Keywords:

Complement clauses, complementizer, false belief, form-function mapping, task effects

Abstract

A key factor that affects whether and at what age children can demonstrate an understanding of false belief and complement-clause constructions is the type of task used (whether it is implicit/indirect or explicit/direct). In the current study, we investigate, in an implicit/indirect way, whether children understand that a story character's belief can be true or false, and whether this understanding affects children's choice of linguistic structure to describe the character's belief or to explain the character's belief-based action. We also measured children's understanding of false belief in explicit false-belief tasks. English- and German-speaking young 4- and 5-year-olds as well as English- and German-speaking adult controls heard complement-clause constructions in a story context where the belief mentioned in the complement clause (e.g., "He thinks that she's not feeling well") turned out to be false, true, or was left open. After hearing the test question ("Why does he not play with her?"), all age groups were most likely to repeat the whole complement-clause construction when the belief turned out to be false. That is, they tended to explicitly refer to the character's perspective and say "He thinks…". When the belief turned out to be true, participants often reverted to a simple clause ("She's not feeling well"). Furthermore, children with better short-term memory were more likely to repeat the whole complement-clause construction. However, children's performance in explicit false-belief tasks showed no relation to their performance in our novel, more implicit/indirect, task. Whether or not the complement clause was introduced by a *that* complementizer only had a small effect on the German adults' responses, where leaving out the complementizer also changes the word order of the complement clause. Overall, our results suggest that task characteristics and individual differences in short-term memory affect children's ability to demonstrate false-belief understanding and to express this understanding linguistically.

The raw data is available via this OSF link:
https://osf.io/3uc64/?view_only=47144d846885440caa22ec5a9f13057c

Introduction

According to most current linguistic theories, language processing and acquisition interact with other (socio-)cognitive skills and processes, such as categorization, analogy, or joint attention, and with our experience of the world (e.g., Bergen, 2015; Langacker, 2008; Tomasello, 2003). In language acquisition research, a lot of attention has been focused on how children's (socio-)cognitive skills facilitate language learning (e.g., Kidd, 2012; Tomasello, 2003). At the same time, there is evidence that language can also affect children's and adults' perception of the world, as has been shown, for example, by studies looking at how linguistic labels drive categorization processes in infants (e.g., Althaus & Westermann, 2016), how cross-linguistic differences in word meaning impact spatial cognition (e.g., Choi, 2006), and the existence of cross-linguistic differences in the extent to which speakers selectively highlight or defocus the agent of a causal action (e.g., Okuno, Cameron-Faulkner & Theakston, 2019). In the current study, we focus on the role of language in children's understanding of other minds. In particular, we investigate how learning complement-clause constructions, such as "She thinks that it's raining", interacts with children's understanding that people can have false beliefs about events, state of affairs, or the identity of entities. In addition, we explore whether cross-linguistic differences in the use of this construction also relate to cross-linguistic differences in the relation between language and understanding of mental states.

Previous studies have argued that complement-clause constructions are a facilitating, or even necessary, linguistic tool that allows children - and adults - to represent others' and their own (past) false beliefs (e.g., de Villiers & Pyers, 2002). What is special about complement-clause construction is that the whole sentence (e.g., "The alien thinks that the earth is flat") can be true even when the complement clause (e.g., "The earth is flat") is false. As has been argued by de Villiers (2007, p. 1869), no other form of linguistic or visual representation would allow us to "represent others' knowledge, or the 'possible world' in someone else's head".

This suggestion that complement-clause constructions are crucial to represent and develop an understanding of false belief has been challenged by cross-linguistic research

and by studies carefully looking at other linguistic tools, such as mental verbs. For example, while in English, finite complements are used with communication and mental verbs, German additionally allows the use of desire verbs together with finite complements (e.g., "She says/thinks/wants that it's raining"). However, only children's comprehension of complement-clause constructions with communication and mental verbs is related to their false-belief understanding, which indicates that it is the semantics of the verbs rather than the structure of the sentence that is related to false-belief development (Perner et al., 2003; for similar findings in Chinese see Cheung, Chen, & Yeung, 2009). This has also been acknowledged by de Villiers (2007), who revised and refined her theory to say that only complement-clause constructions containing realis verbs and tensed complements support false-belief understanding.

In addition, novel research paradigms, employing implicit tests of false-belief understanding, suggest that infants might start to develop an understanding of others' mental states well before they are able to either produce or comprehend mental verbs or complement-clause constructions (e.g., Baillargeon, Scott, & Bian, 2016). At the same time, Lewis, Hacquard, and Lidz (2017) suggested that there are also alternative - and pragmatically more felicitous - ways of testing children's understanding of complement-clause constructions than those used in previous research on language and false-belief development. They argued that children could show an earlier understanding of complement-clause constructions when (conflicting) beliefs are highlighted in the task. More details about different ways of testing children's understanding of both false belief and complement clauses will be provided in Section 1.3.

In the current study, we present a new way of testing children's understanding and elicited production (i.e. repetition) of complement-clause constructions together with their understanding of false belief. Like Lewis et al. (2017), we embedded the test sentences in story contexts supported by picture stimuli, where characters had true or false beliefs. However, whereas Lewis et al.'s task was rather meta-linguistic and explicit in nature (i.e., children had to judge whether a character's statement containing a complement-clause construction was true or false), we developed a more naturalistic and indirect way of testing

children's - and adults' - understanding and production of complement-clause constructions. As will be discussed in more detail below, our participants were introduced to a protagonist who wanted to do something (e.g., play with their friend), but believed that they would not be able to do so because, for example, their friend was not feeling well. The protagonist's belief was always expressed by a complement-clause constructions (e.g., "He thinks that she's not feeling well"). In the following scene, this belief turned out to be false or true, or it was left open whether the protagonist's belief was false or true. Then we asked the participants why the protagonist was not doing what they wanted to do and expected children and adults to repeat the whole complement-clause construction when the protagonist's belief was false (i.e., when she actually did feel well) and to be more likely to only repeat the complement clause (e.g., "She's not feeling well") when the protagonist's belief was true (i.e., when she really did not feel well). In other words, children and adults were expected to explicitly refer to the protagonist's mental state when that protagonists' behavior did not make sense according to the real state of affairs. Before we introduce this novel task, we present a summary of previous research on the form and function of complement-clause constructions and their role in children's developing understanding of false belief.

1.1 Form and Function of Complement Clauses in Spontaneous Speech

Complement-clause constructions contain a main clause or complement-taking phrase (CTP) and a complement clause. In example (1) the CTP is "I think", followed by the complement clause "that it's raining":

(1) I think [that it's raining].

On a formal level, complement-clause constructions have traditionally been classified as subordinate constructions, where the complement clause serves as an argument of the verb in the CTP (e.g., "think") (e.g., Noonan, 1985). Prototypically, main clauses contain new and foreground information, whereas subordinate clauses contain old and background

information. Therefore, classifying complement-clause constructions as subordinate constructions would suggest that the CTP contains foreground information, whereas the complement clause contains background information. That is, when we use complement-clause constructions such as "I think that it's raining", we are mostly focusing on our *thinking* – or other mental states and processes such as knowing or guessing - and less on the proposition expressed by the complement clause ("that it's raining").

However, this analysis of complement-clause constructions has been challenged by analyses of adult spoken discourse (e.g., Simons, 2007; Thompson, 2002; Thompson & Mulac, 1991; Verhagen, 2005). Thompson (2002) offers compelling arguments against classifying complement clauses as subordinate, on both the formal and the functional level. We will mainly focus on the functional level here. Most importantly, Thompson (2002, p. 152) argues that "the interactional work that conversationalists are engaged in doing with an utterance that appears to contain a complement involves the complement just as much as the CTP-phrase". This functional analysis is supported by the observation that what is expressed in the complement clause is often the focus of the conversation, as convincingly shown in this example from Thompson (2002, p. 132):

(2) (talking about a photo collage on the wall; parentheses added)

Terry: I think [it's cool].

Abbie: It is cool.

Maureen: It is great.

The exchange in (2) is not about Terry's *thinking*, but about the content of the complement clause ("it's cool"). In fact, what Thompson (2002, p. 131) and others (e.g., Kärkkäinen, 2003; Simons, 2007; Thompson & Mulac, 1991; Verhagen, 2005) suggest is that the function of most CTPs occurring in spoken discourse is to express an epistemic, evidential, or evaluative stance. That is, CTPs are used as parenthetical markers to express speaker certainty, source of information, or speaker attitude in relation to a given proposition, with that proposition expressed by the complement clause being the actual

focus of the conversation or main point of utterance. However, this is not to say that CTPs are always used as parenthetical markers. They can also contain the main proposition and occur together with a complement clause containing background information, as shown in example (3). CTPs lie on a continuum, where they can function as parenthetical markers at one end, or express the main proposition at the other end.

(3) I don't care [that we missed the flight].

In the current study, we investigate whether children and adults are more likely to interpret the CTP as the main proposition and include it in their repetition of a complement-clause construction when it refers to a character's false belief, and whether they are more likely to interpret the CTP as a parenthetical marker and leave it out of their repetition when it refers to a character's true belief.

Previous studies also suggest that certain CTP forms are more likely to function as parenthetical markers than others. Formulaic CTPs, which often contain first-person subjects and a frequent verb in present tense, and no other linguistic material (e.g., "I guess", "I know", "I think"), are most likely to function as parenthetical markers (e.g., Diessel & Tomasello, 2001; Thompson & Mulac, 1991). However, parenthetical markers can also contain second- or third-person subjects, less frequent verbs, and past-tense verbs. For example, Simons (2007) shows that CTPs such as "Henry thinks", "I regret", or "I heard" can also be used and interpreted as parenthetical markers in specific contexts.

Another formal feature shared by many CTPs that can be interpreted as parenthetical markers is that the following (or preceding) complement clause is not introduced by a *that* complementizer (see example (2) above) (e.g., Diessel & Tomasello, 2001; Thompson & Mulac, 1991; see also Dor, 2005; Kaltenböck, 2006). Leaving out the *that* complementizer makes complement clauses formally indistinguishable from main clauses that occur without CTPs. This formal feature is quite subtle in English. Note, however, that when complement clauses are preposed in English, they are always used without a complementizer (e.g., "It's still raining, I think"). These preposed complement clauses have been discussed as a form of

embedded root transformation, which are typically used to add emphasis to the moved element (the complement clause in this case). And Hooper and Thompson (1973, p. 470) argued that "since the complementizer is not present, the preposed sentence is taken as the main assertion, and thus receives more emphasis than if it were in the usual complement position".

In a verb-second language such as German, complement clauses also display a different word order when the complementizer is left out. Complement clauses introduced by a complementizer are verb-final (see (4a)), whereas those without complementizers are verb-second, which is the same word order as that used in main clauses (see (4b)):

(4a)	Ich	glaube,	dass	es	noch	**regnet**.
	I	believe	that	it	still	rains
	I believe that it is still raining.					

(4b)	Ich	glaube,	es	**regnet**	noch.
	I	believe	it	rains	still
	I believe it is still raining.				

Similar to English, for German and other Germanic languages, it has been argued that verb-second subordinate clauses without complementizers express the main point of utterance (e.g., Brandt, Lieven, & Tomasello, 2010). However, there does not seem to be a clear relation between absence of complementizer, verb-second word order and main point of utterance. In Frisian, for example, verb-second complement clauses can also be introduced by a complementizer, and complement clauses can also express the main point of utterance when they are introduced by a complementizer (see Heycock, 2017).

In addition, it has also been argued that the verb-second word order in complement clauses without complementizers is mainly a syntactic phenomenon, unrelated to the function and information status of the complement clause (e.g., den Besten, 1983). That is, in German and other Germanic languages, the complementizer position in a syntactic tree is

either occupied by a complementizer or the finite verb. When the complementizer is left out, the finite verb moves into the complementizer position, creating the verb-second word order. For language acquisition, it has been argued that as soon as German-speaking children acquire the complementizer position, they also follow this syntactic rule (e.g., Poeppel & Wexler, 1993). However, Fritzenschaft, Gawlitzek-Maiwald, Tracy, and Winkler (1990) also found that German-speaking children sometimes use verb-second subordinate clauses with complementizers. Similarly, it has been found that both children and adults produce verb-second subordinate clauses together with relativizers or conjunctions. Rather than seeing this as a syntactic-rule violation, it has been suggested that verb-second subordinate clauses are used to express the main point of utterance and that the presence or absence of the complementizer or other material in the complementizer position does not matter (e.g., Antomo & Steinbach, 2010: Brandt, Diessel, & Tomasello, 2008).

In the current study we investigate and compare English and German participants' interpretation and elicited production of complement clauses with and without complementizers. For both languages, it has been argued that complement clauses without complementizers are more likely to express the main point of utterance than complement clauses with complementizers (e.g., Brandt et al., 2010; Hooper & Thompson, 1973). For German, however, the main factor driving the information status of complement clauses might be their word order. Verb-second complement clauses, with or without complementizers, tend to express the main point of utterance. In addition, the verb-second word order in complement clauses without complementizers might be a syntactic rather than pragmatic-semantic phenomenon (e.g., Heycock, 2007; Poeppel & Wexler, 1993). Therefore, it is possible that the presence or absence of complementizers shows different effects in English and German, and we will come back to this point in the Discussion.

To summarize, data from adult spoken discourse suggest that CTPs can take on different functions. They can function as parenthetical markers that are used together with complement clauses expressing the main proposition. Alternatively, they can express the main proposition and occur with complement clauses containing background information. Only in the latter case do interlocutors primarily focus on mental states and processes such

as *thinking, knowing* or *remembering.* These functional differences do, to some extent, also correlate with formal differences: CTPs that function as parenthetical markers are often expressed by formulaic phrases, such as "I think" or "I guess", and the following complement clause often lacks a *that* complementizer, which, in verb-second languages such as German, also leads to word-order changes.

1.2 Development of Complement Clauses in Spontaneous Speech

When looking at English-speaking children's first spontaneous use of complement-clause constructions around the age of 3 years, Diessel & Tomasello (2001) found that they tend to only use CTPs that function as parenthetical markers. Only later do they also use CTPs to refer to mental states and processes such as *knowing* and *thinking* (for similar findings see Bartsch and Wellman, 1995; Shatz, Wellman, and Silber, 1983). More recently, Harris and colleagues (e.g., Harris, Yang, & Cui, 2017) have argued that 20-month-olds do not just use mental verbs as parenthetical or discourse markers, but already use them in a more sophisticated way, to signal knowledge or ignorance. However, their analysis was based on only one mental verb ("know") and it was not restricted to complement-clause constructions, but also included simple sentences, such as "I don't know" or "Do you know that".

Given that adults are also most likely to use CTPs as parenthetical markers rather than to foreground mental states and processes, this developmental pattern is perhaps not too surprising (see also Dudley, Rowe, Hacquard, & Lidz, 2017; Hacquard & Lidz, 2019; Harrigan, Hacquard, & Lidz, 2019). However, it has also been argued that, early in development, children might not use CTPs to refer to mental states and processes because, up until the age of 5 years, they do not seem to have a full understanding of their own and others' mental states (e.g., Bartsch & Wellman, 1995; Shatz et al., 1983), as has been shown in a great number of studies looking at children's false-belief understanding (for an overview and meta-analysis see Wellman, Cross, & Watson, 2001). Further support for the developmental link between complement-clause constructions and mental-state understanding comes from studies that have found longitudinal and concurrent relations

between children's comprehension of complement clauses and their false-belief understanding, which will be discussed in more detail below.

1.3 Comprehension of Complement Clauses and False Belief

A good number of correlational, longitudinal, and training studies have shown a link between children's acquisition of complement-clause constructions and their false-belief understanding. For example, de Villiers and Pyers (2002) used a longitudinal design to demonstrate that English-speaking children's understanding of the subordinate structure of complement-clause constructions predicts their false-belief understanding. To test children's understanding of the structure of complement-clause constructions, children heard, for example: "He thought he found his ring, but it was really a bottle cap. What did he think?" And they were expected to say "he found his ring" or just "ring". Children who did not understand the subordinate structure of complement-clause constructions often wrongly responded with "bottle cap". Alternatively, it has been suggested that children failed to provide the correct answer because of processing difficulties rather than lack of syntactic knowledge. In particular, Lutken, Legendre, and Omaki (2020, p. 40) suggested that children struggle with complex questions such as "What did he say he stole" because memory limitations make it difficult or impossible for them to integrate the complement clause into the matrix (e.g., "What did he say") (see also De Mulder, Wijnen, & Coopmans, 2019). In addition, children with limited short-term memory might also find it difficult to remember the details from the story context, which are needed to answer the question. We will come back to memory limitations in the Discussion.

Furthermore, the study by de Villiers and Pyers (2002) has also been criticized for potential confounds between the complement-clause and false-belief tasks. In particular, the complement-clause task also involved a false-belief scenario (see Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003 for detailed discussion). In fact, de Mulder et al. (2019) developed a test of complement-clause understanding that did not involve any false-belief scenario and found no longitudinal relationship between children's understanding of complement-clause constructions and their understanding of false belief. At the same time,

Boeg Thomsen, Theakston, Kandemirci, and Brandt (2021) also tested children's comprehension of complement-clause constructions without a false-belief scenario and were able to replicate the original findings by de Villiers and Pyers (2002). Whether or not complement-clause constructions play a unique role in children's false-belief development thus remains an open question and we will come back to this in the Discussion.

The correlation and developmental link between children's processing of complement-clause constructions and their understanding of false belief has also been found in training studies with English-speaking children (Hale & Tager-Flusberg, 2003), German-speaking children (Lohmann & Tomasello, 2003), Mandarin-speaking children (Mo, Su, Sabbagh, and Xiu, 2014), children with autism (Tager-Flusberg & Joseph, 2005), and children with developmental language disorder (Durrleman, Burnel, and Reboul, 2017). However, recent correlational, longitudinal, and training studies (Boeg Thomsen, et al., 2021; Brandt, Buttelmann, Lieven, & Tomasello, 2016; Howard Gola, 2012) suggest that not all types of complement-clause constructions are equally related to children's false-belief understanding. In particular, children's understanding of formulaic CTPs with first-person subjects (e.g., "I think") show a weaker link to false belief than their understanding of less formulaic CTPs with third-person subjects (e.g., "the cow thinks") (see also Howard, Mayeux, & Naigles, 2008). This finding from comprehension studies with children is in line with the production data from both children and adults summarized above. There seems to be converging evidence suggesting that the CTPs that are used most frequently mostly function as parenthetical markers rather than to draw attention to mental states and processes, and that they are thus only weakly linked to or dependent on children's understanding of mental states and processes.

However, recent studies on children's (and even infants') understanding of false belief challenge the conclusion that children's false-belief development depends on their acquisition of complement-clause constructions. In particular, in specific types of false-belief tests, infants seem to be able to demonstrate an understanding of false belief long before they comprehend or produce complement-clause constructions. Over the last 15-20 years, researchers have developed more implicit test paradigms to investigate infants' and young

children's understanding of false belief (for overviews see Baillargeon et al., 2016; Scott & Baillargeon, 2017). In these implicit test paradigms, participants are not asked to make any explicit predictions about belief-dependent behaviors or verbal actions. Instead, they are shown belief-congruent or belief-incongruent actions. For example, they are shown an animation or a video where a protagonist puts an object into container A. Then the protagonist leaves the scene and, in their absence, the object is moved into container B. When they return, the protagonist either looks for the object in container A (belief congruent) or in container B (belief incongruent). In these violation of expectation paradigms, it has been found that infants (under the age of 2 years) look longer at belief-incongruent actions than at belief-congruent actions (e.g., Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007). In anticipatory looking paradigms, infants look at the container or location that a protagonist with a true or false belief would approach before they actually go to that location (e.g., Southgate, Senju, & Csibra, 2007). And in helping paradigms, infants adapt their helping behavior according to the protagonist's belief (e.g., Buttelmann, Carpenter, & Tomasello, 2009).

These findings have been interpreted as evidence that infants (under the age of 2 years) have an understanding of others' mental states, and that they understand that others' behavior is driven by their mental states. Even though this rich interpretation has been challenged by various researchers (e.g., Priewasser, Rafetseder, Gargitter, & Perner, 2018; Rakoczy, 2012; Ruffman, 2014), and the results from these implicit false-belief tests have failed to replicate (e.g., Dörrenberg, Rakoczy, & Liszkowski, 2018), it is possible that infants have at least some rudimentary understanding of others' mental states.

Alongside these recent developments in the design of new false-belief tests, some researchers have also developed new ways of testing children's understanding of complement-clause constructions. In particular, Lewis et al. (2017) suggest that children as young as 3 years are able to interpret CTPs as referring to mental states and not just as parenthetical markers. They suggest that children are most likely to interpret CTPs as parenthetical markers because that is the most common function of CTPs in their input (see also Dudley et al., 2017; Hacquard & Lidz, 2019; Harrigan et al., 2019). However, when the

task is designed in such a way that mental states are the focus of the conversation, children are, to some extent, able to understand CTPs as referring to these mental states. In Experiment 1 (Lewis et al., 2017), children between the ages of 3;10 (3 years; 10 months) and 4;5 were told and shown a story where one character was hiding and one or two seekers had to find them. Both seekers expressed their beliefs (e.g., Dora: "Hmm, where should I look? Oh, I see a yellow tail behind the toy box. I know! Swiper's there! I'll look for Swiper behind the toy box!" Boots (in the two-seeker condition): "Hmm, where should I look? Oh, I see a yellow tail behind the curtain. I know! Swiper's there! I'll look for Swiper behind the curtain!"). Then children heard complement-clause constructions from another puppet (e.g., "Dora thinks that Swiper is behind the toy box.") and had to judge whether this statement was true or false. The finding most relevant for the current study was that children performed better when there were two seekers with two different beliefs about the hiding location than when there was only one seeker with one belief. The authors suggest that having two conflicting beliefs draws children's attention to these beliefs and makes them focus on the CTPs that encode the beliefs. When there was only one seeker, children were more likely to interpret the complement-clause constructions without paying any attention to the CTP (e.g., "Dora thinks that Swiper is behind the toy box" was understood as if it just said "Swiper is behind the toy box"). This meant that in a false-belief scenario (where Swiper was *not* behind the toy box), children often incorrectly rejected the complement-clause statement.

In Experiment 2, children between the ages of 3;1 and 4;2 always saw two seekers, who held true or false beliefs about the hiding place of a third character. In the key condition, one seeker had a false belief and said, for example, "Hmm, where should I look? Oh, I see a yellow tail behind the toy box. I know! Swiper's there! I'll look for Swiper behind the toy box!" Then the children heard either a literally true or false sentence from another puppet (literally true: "Dora thinks that Swiper is behind the toy box"; literally false: "Dora thinks that Swiper is behind the curtain"). In the literally false sentence, when the seeker had a false belief, the complement clause "Swiper is behind the curtain" corresponded to the real state of affairs (i.e. Swiper really is behind the curtain, but Dora falsely thinks that he's

behind the toy box). Therefore, if children simply ignored the CTP, they should accept the literally false sentence because the complement clause correctly describes the current state of affairs. However, Lewis et al. (2017) found that children were less likely to accept the literally false sentence than the literally true sentence in the false-belief condition and argue that this indicates that children are able to take into account the CTP when they interpret complement-clause constructions.

It should be noted, however, that children still struggled with this task in the false-belief condition. The authors found that they generally performed better when the main seeker had a true belief. In addition, even though children were good at rejecting a literally false sentence in the false-belief condition, they only showed at-chance performance when they had to accept a literally true sentence in this condition. This indicates that, even though children might have at least started to develop an understanding of CTPs and mental states by the age of 3 years, they still find it difficult to coordinate their linguistic knowledge and their understanding of mental states. That is, children still need to learn the correct mapping between specific linguistic constructions and different types of mental states. The authors also tested participants on two explicit false-belief tasks, but found no correlations between false-belief understanding and the ability to judge whether complement-clause statements were literally true or false, thus questioning previous research on the developmental link between children's understanding of complement-clause constructions and their understanding of false belief.

1.4 The Current Study

The current study was designed to further investigate the relationship between children's understanding of false belief and their interpretation of complement-clause constructions. As in Lewis et al.'s (2017) study, we were interested to see whether children are able to focus on CTPs when the task is designed in such a way that it draws their attention to different beliefs and mental states. However, we designed a novel and more naturalistic task, which did not involve any meta-linguistic truth-value judgment. As will be described in more detail below, children heard stories containing complement-clause

constructions and our test questions indirectly asked them to repeat these. This enabled us to investigate children's comprehension and elicited production (i.e. repetition) within the same task. We hypothesized that this task would show that children around the age of four years are able to focus on and repeat CTPs in their own production (cf. Lewis et al., 2017).

In addition, we wanted to investigate whether (i) children's ability to focus on the mental states and processes encoded by CTPs interacts with their general understanding of mental states and processes, as measured in explicit false-belief tasks, and (ii) if and when children are able to choose and produce appropriate linguistic constructions for different mental states. To this end, we embedded different belief conditions in the stories. Story characters held a false-, true-, or open belief (i.e., it was left open whether the protagonist's belief was false or true). We hypothesized that children would be more likely to focus on and repeat CTPs when they also showed a general understanding of mental states and processes in explicit false-belief tests, and when the character's belief was false. Adults and older children might also be more likely to focus on and repeat the CTP because they have a better understanding of mental states than younger children, and adults and older children might also show more sensitivity to the different belief conditions presented in the stories. In addition, older children and adults might also be more likely to produce complement-clause constructions because they have better short-term memory.

Another aim of the current study was to take a closer look at form-function mappings in children's and adults' interpretation of complement-clause constructions. As summarized above, data from spontaneous speech suggest that when the complement clause is not introduced by a complementizer, the CTP is more likely to be interpreted as a parenthetical marker (e.g., Diessel & Tomasello, 2001; Thompson & Mulac, 1991). In order to establish whether this is the case for both children and adults in a more controlled, experimental setting, we presented our participants with complement clauses with and without complementizers. We hypothesized that both children and adults would be more likely to focus on and repeat the CTP when the complement clause is introduced by a complementizer.

Finally, we also noted that the difference between complement clauses with and without complementizers is more marked in verb-second languages such as German, where it also leads to word-order differences in the complement clause. Therefore, we tested both German- and English-speaking children and adults and hypothesized that German-speaking participants would show a greater sensitivity to the presence or absence of the complementizer and the accompanying changes in word order.

2. Methods

2.1 Participants

We piloted the complement-clause task described below with German-speaking children aged 3;5 (3 years; 5 months) (*n* = 9), but most of them struggled to understand the task. Therefore, in both English and German, we tested children around the age of 4-years (henceforth referred to as 4-year-olds, range = 3;9-4;3), and around the age of 5-years (henceforth referred to as 5-year-olds, range = 4;9-5;3) as well as adults. For the English sample, we recruited 27 4-year-olds, 27 5-year-olds, and 25 adults. Two 4-year-olds and two 5-year-olds had to be excluded due to experimenter error (*n* = 1), general difficulty understanding the complement-clause task (*n* = 2)[a], or giving a large number of unintelligible responses on the complement-clause task (*n* = 1). One adult participant was excluded due to the voice recorder malfunctioning. In total, there were 25 participants in each of the child participant groups (4-year-olds: mean age = 4;0, range = 3;9-4;3; 11 males; 5-year-olds: mean age = 5;0, range = 4;9-5;3; 13 males) and 24 in the adult group (mean age = 36;7, range = 19;9 – 61;11; 4 males) in the final sample. Children were recruited from a state school in the Greater Manchester area. Adults were sampled from the same geographical area and from a range of occupational backgrounds. All participants were native speakers of English with no known developmental disorders. Children were tested individually in their school and adults were tested at the University of Manchester or in a quiet room at their workplace.

[a] Children who did not understand the task either always provided the same answer (e.g., "because it's raining") or said something that was not related to or mentioned in the story.

For the German sample, we recruited 27 4-year-olds, 24 5-year-olds, and 24 adults. Three 4-year-olds had to be excluded due to general difficulty understanding the complement-clause task (*n* = 2) or the short-term working memory test described below (*n* = 1). In the final sample, there were 24 participants in each of the child participant groups (4-year-olds: mean age = 4;0, range = 3;9-4;2; 12 males; 5-year-olds: mean age = 5;0, range = 4;9-5;3; 12 males) and 24 in the adult group (mean age = 29;3; range = 19;6-44;3; 5 males). Children were recruited from nurseries in Leipzig. Adults were sampled from the same geographical area. The adult participants were studying towards or already in possession of a postgraduate degree at the University of Leipzig. All participants were native German speakers with no known developmental disorders. Children were tested individually in a quiet room in their nursery and adults were tested at the University of Leipzig or in a quiet room at their workplace.

2.2 Design

We tested two factors within participants: Within a story context, complement-clause constructions were presented with or without a *that* complementizer (with complementizer vs. without complementizer). The belief encoded by the complement clause turned out to be true or false, or it was left open whether it was true or false (true belief vs. false belief vs. open belief). This 2x3 design resulted in a total of six conditions. We conducted separate and combined analyses for the English and German sample. In the separate analyses, the only between-participants factor was age (4-year-olds vs. 5-year-olds vs. adults). In the combined analysis, we added language group as a between-participants factor. All participants completed the complement-clause task. Children then completed four explicit false-belief tasks, to test their general understanding of mental states, and one sentence-repetition task, to test their short-term memory as a control variable.

2.3 Materials

2.3.1 Complement-clause task: We created 18 short stories, so that each of the six conditions could be tested in three different stories creating three items per condition. The

stories all followed the same basic script, but differed slightly in content (for an overview see Appendix A). An additional three stories were created for use as warm-ups. These warm-up stories were similar in format and content to the test stories; however, they were less complex in that they did not contain any complement-clause constructions or other mental-state language.

For the test items, each story included three context sentences, two manipulation sentences, and a test question. The first manipulation sentence manipulated the presence of the *that* complementizer (with complementizer vs. without complementizer). The second manipulated whether the belief encoded by the complement clause was true (e.g., "Oh look! She really isn't feeling well"), false (e.g., "Oh look! She actually is feeling well"), or open. Sentences which appeared in the open belief condition offered no additional information (e.g., "Oh look, there's a yellow star"). The star was either red, blue or yellow, and this was counterbalanced across items. Table 1 shows an example story in the 'with complementizer' and 'true-belief' condition (for a full list of test items see Appendix A).

Table 1

Test item in the 'with complementizer' and 'true belief' condition[b]

Context sentence: Introduction of main character	*This is Tom.*	
Context sentence: Introduction of second character or a specific location	*Tom likes Anne.*	

[b] The sentences were pre-recorded and presented together with simple images

Context sentence: Desirable activity with second character or at a specific location	*Tom wants to play tennis with Anne.*	
Complementizer manipulation: With vs. without complementizer	*But he thinks **that** she is not feeling well.*	
Belief manipulation: True belief vs. false belief vs. open belief	*Oh look, **she really isn't feeling well!***	
Test question	*Why doesn't Tom play tennis with Anne?*	

Six characters featured in the stories, with a maximum of two characters appearing together in each story. Character pairings were always the same: Tom and Anne (Tom and Lisa in German); Sam and Jane (Max and Sophie); and Dan and Sue (Paul and Ida). In half of the stories, the main character wanted to engage in an activity with the other character (character stories; see example in Table 1). In the remaining half, the main character wanted to engage in an activity at a specific location (location stories; e.g., "Tom wants to play in the garden"). For both character and location stories, activities were limited to "play with/at", "go with/to", and "eat with/at" (see Table 2).

The presentation of the 18 stories was semi-randomized: Participants received each desirable activity twice within each belief manipulation, once paired with a complement clause with a complementizer and once paired with a complement clause without a complementizer. In addition, for each desirable activity, half of the stories were character stories, and half were location stories. The order of presentation ensured that character

pairs did not appear in consecutive stories. Based on this, we created six different semi-randomized lists (see Table 2 for an overview and Appendix A for a full list of test items).

Table 2

Sentence stimuli and story types in complement-clause task

	True belief	**False belief**	**Open belief**
With that-complementizer	play story	play story	play story
	go story	go story	go story
	eat story	eat story	eat story
Without that-complementizer	play story	play story	play story
	go story	go story	go story
	eat story	eat story	eat story

Sentence length was controlled by adding adverbials (e.g., “still”, “today”, “outside”), possessive pronouns (e.g., “his”) and adjectives (e.g., “new”, “very”, “red”) to increase sentence length where necessary. For the English participants and German adults, items were presented using Microsoft PowerPoint on a 13-inch laptop. The German children saw the items on a 15-inch laptop. Stories consisted of pre-recorded context and manipulation sentences spoken by a native speaker of English or German, and each sentence was supported by a single image, created using Photoshop (see example in Table 1). The experimenter always asked the test question live - using a puppet for child participants.

2.3.2 False belief tasks: The child participants completed four explicit false belief tasks, consisting of two unexpected-contents tests (Perner, Leekam & Wimmer, 1987) and two change-of-location tests (Wimmer & Perner, 1983; Brandt et al., 2016). Both sets included a first-person version (i.e. assessing understanding of children’s own beliefs) and a third-person version (i.e. assessing understanding of others’ beliefs). The order of task completion was counterbalanced and each sequence was paired with one of the six randomized lists

that were created for the complement-clause task. So, for example, all children assigned to List 1 completed the false belief tasks in a given order.

2.3.3 Sentence repetition task: Children also completed a sentence repetition task to assess their short-term memory. The test items consisted of six sentences ranging from six to ten words using varied sentence structures, including subject relative clauses and simple transitive sentences with adverbs (see Appendix B). Test sentences and scoring were adapted from the *Sprachentwicklungstests für drei- bis fünfjährige Kinder* (Language Development Test for Children aged 3 to 5; SETK 3-5) (Grimm, Aktas, & Frevert, 2001) for German and the *Clinical Evaluation of Language Fundamentals* (CELF-Preschool 2 UK) (Wiig, Secord, & Semel, 2006) for English. The order of sentence presentation was the same for all participants, since the complexity of sentences increased for each successive item.

2.4 Procedure

All English participants were tested by the same female native speaker of English. The German children were tested by a female, native speaker of German and the German adults were tested by another female, native speaker of German. For the complement-clause task, the experimenter (E) sat next to the participant at a table with the laptop positioned in front of the participant. E first told the participant that they were going to listen to some stories through the laptop and there would be pictures on-screen. For child participants, before starting, E introduced a human looking puppet named Eddie (Nils in German). Children were told that Eddie was feeling tired and if he fell asleep, he would miss parts of the stories, so they had to listen carefully in case Eddie had some questions to ask. After indicating they were happy to help the puppet, children were told they could win a sticker for every three stories they heard, providing they listened carefully. E then went on to start the three warm-up stories. Each warm-up item introduced one of the characters that also appeared in the test stories and a desired activity, which could not be achieved (see Appendix A). After hearing the first sentence of a story item (e.g., “This is Tom”), the puppet always fell asleep. Once a story had played through, the puppet woke up and asked the child the test question

(e.g., “Why doesn’t Tom eat cereal?”). If children gave no response, E used prompts such as: “Hmmm, can you think? Why doesn’t Tom eat cereal?” and “Can you remember what was said in the story?” For the warm-up items, provision of incorrect answers was corrected by E (e.g., “He hasn’t got any milk. Do you remember hearing that in the story? That’s why he doesn’t eat cereal!”). For the adult groups, E asked the test questions directly and there were no stickers.

After the warm-up stories, participants completed the 18 test items. Each slide was pre-set to stay on screen for 5.25 seconds before automatically transitioning to the next, allowing enough time for participants to view the visual stimuli and for the pre-recorded sentences to play. For the children, during the test items, the puppet would fall asleep and wake up at the same points as during the warm-up items. E gave positive feedback after each trial with regard to their listening and concentration only, to keep the children motivated. Responses for all tasks were audio-recorded to allow for reliability checks with coding.[c]

When the children had completed the complement-clause task, they were asked to move their chair to the other side of the table so that they faced E during the next tasks. The experimenter then started the first of the four false belief tasks. For the explicit change-of-location test (Wimmer & Perner, 1983), E used two small dolls to introduce two new characters (Peter and Susie). Children had to identify the name given to each doll before E continued with the test. The story narrative was told in present tense. It explained that: “Peter puts his ball in his box, he then goes away, where he isn’t able to see or hear us anymore. Next Susie comes along and moves the ball from the box to her basket, and then she goes away, too”. E acted out the story whilst telling the narrative, then asked the test question: “When Peter comes back, he’s looking for his ball. Show me, where will Peter look

[c] Even though having 18 similar stories could be taxing for the concentration levels of the child participants, our data suggest that children remained attentive to the stimuli throughout the testing period: For the German and English 4-year-olds, the number of ‘other’ and ‘null’ responses decreased throughout the complement-clause task. For the German and English 5-year-olds the number of ‘other’ and ‘null’ responses stayed around the same level throughout the complement-clause task.

for his ball?" This was followed by the reality control ("Where is the ball really now?"), and finally the memory control ("Where did Peter put the ball in the beginning?").

The other version of the change-of-location task tests children's understanding of their own revised beliefs (see Brandt et al., 2016). This test involved E placing a closed opaque red box and a closed opaque blue box on the table in front of the child. She then presented a small toy to the child and informed them that she was going to hide it in one of the boxes. An occluder was placed on the table between the boxes and the child whilst E conducted the manipulation: The toy was always placed in the blue box and the lid of the red box was repositioned to create a false belief. So, when the occluder was removed, it looked as if E had put the toy in the red box. E then asked the control question ("Can you tell in which box I've hidden the toy?"), before revealing the true location of the toy (the blue box) by removing it from the box. It was imperative that the child responded to this control question correctly (stating the red box) to ensure they had the appropriate false belief. After replacing the toy and closing the box lid, E asked the test question ("Where did you think the toy first was?"), and the reality control question ("Where is it really now?").

Children also took part in an unexpected-contents task which included a test question assessing understanding of own beliefs and a test question assessing understanding of others' beliefs (Perner et al., 1987). For this task, E placed a *Smarties* tube on the table in front of the participant and then asked the child: "What do you think is in here"? After responding with *Smarties*, *sweets*, *chocolate*, or similar, E proceeded to reveal the contents of the tube, explicitly stating that no Smarties were inside, but in fact crayons were inside. E then replaced the crayons in the tube and closed it up again. Pointing at the tube, E asked the reality control question ("What's in here now?"), followed by the two test questions: the first-person version ("What did you first think was in here?") and the third-person version ("What will [participant's teacher's/friend's name] think is inside the box?"). The order of the two questions was counterbalanced across lists.

Finally, children completed the sentence repetition task (see Appendix B for full list of sentences). The task was introduced as a parrot game. Two warm-up examples were given to ensure the child understood the task: "So, when I say, 'I like cake', the parrot says, 'I like

cake'". E asked the child if they could pretend to be a parrot and presented the practice statement for them to repeat: "The car goes fast". After the child had attempted to repeat the practice sentence, E continued with the test sentences. There were six test sentences in total and E continued to the next sentence after receiving a response from the child. Children were offered the opportunity to have the sentence repeated once per trial, on the condition that they had made no initial attempt at repeating the target sentence, or because there had been a disruption in the room (e.g. the school bell rang).

Test sessions with adults lasted between 10 and 15 minutes, and with children between 25 and 35 minutes.

2.5 Coding and Scoring

For the complement-clause task, we expected participants to repeat the complement-clause manipulation sentence with or without the CTP. For example, when asked "Why doesn't Tom go to the playground?" we expected an answer such as "Because (he thinks that) it's raining outside". Responses that contained the CTP and the complement clause were coded as 'main-subordinate'. Responses that only repeated the complement clause (e.g. "Because it's raining outside") were coded as 'subordinate-only'. Responses – with or without CTPs – that contained different beliefs or reasons, not mentioned in the story (e.g., "Because (he thinks that) she doesn't want to play") were coded as 'other' and further analyzed in the error analysis. Finally, giving no response, saying "I don't know", or providing an incomplete utterance such as *"*because*"* was coded as 'NR' (no response). We tolerated changes in tense, leaving out of adverbials, and small lexical changes, including the mental verbs (e.g., "he believed" instead of "he thinks"). A sample of coded data can be found in Appendix C. The German transcripts were coded by the first author; the English transcripts were coded by the second author. Ten percent of the English data were coded by both first and second author, yielding Cohen's Kappa coefficients of 0.96 (97.7% agreement).

Responses on the false belief tasks were coded as '1' (pass) if children correctly answered the test question *and* the control question(s) for that trial, or '0' (fail) if children gave an incorrect response to either the test question or the control question(s), or both,

for that trial. Since there was a total of four false-belief test questions (2 assessing own beliefs; 2 assessing others' beliefs), 4 was the maximum score participants could achieve.

Responses on the sentence repetition task were coded according to the number of correct words repeated by participants for each item. Following the scoring manuals, the order in which the words were repeated did not affect the scoring, but most children repeated them in the same order as in the target sentence. Addition of extra words was not penalized either (see Appendix B for the maximum scores that could be obtained for each item). If children failed to produce any sort of utterance, they were automatically given a '0' for that item. English participants could achieve a maximum score of 47 on the sentence repetition task, whereas for German the maximum score was 46.

We will first present the results from the English participants, followed by the German participants, and finally a combined analysis of both language groups.

3. Results

3.1 English Children and Adults

Figure 1 shows the mean number of main-subordinate responses for the English participants in each condition of the complement-clause task. All age groups were most likely to produce both the CTP and the complement clause in the false-belief conditions, followed by the open-belief conditions. And they were least likely to produce both the CTP and the complement clause in the true-belief conditions. The presence of the complementizer, however, did not seem to have a systematic or large effect on the proportion of main-subordinate responses within the open-, false- or true-belief conditions.

Figure 1

Proportion of main-subordinate responses in each condition for the English-speaking participants. Maximum number of main-subordinate responses in each condition is 3. Compl. = complementizer.

To test whether age, belief, and presence of complementizer had a significant effect on participants' tendency to produce main-subordinate responses, we analyzed these data with binomial generalized linear mixed models (glmer) in R (version 3.6.0, R Core Team, 2019), using the package lme4 (version 1.1-7, Bates et al., 2015). For these main analyses we only considered answers that were coded as either 'main-subordinate' or 'subordinate-only'. Other responses and errors will be analyzed and discussed in section 3.4 below. Following the principle of backward selection and a significance-based approach (Gries, 2013), we started with a full model that contained the random effects of participants and items and all fixed effects (age: 4 vs. 5 vs. adults; belief: open vs. false vs. true; complementizer presence). Note that we were only able to use random intercepts rather than random intercepts and slopes because the models that included random slopes (for subjects) did not converge. Then we took out the least significant factor and checked whether this would significantly decrease the goodness of fit of the model, by comparing models by ANOVA. This was repeated until taking out the least significant effect significantly decreased the goodness of fit of the model.

The final model for the English-speaking participants, presented in Table 3, indicates that the number of main-subordinate responses increased with age, and that participants were more likely to produce main-subordinate responses in the false- and open-belief conditions than in the true-belief conditions.

Table 3

GLMM model for main-subordinate responses by English participants[d]

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.1298	0.6849	3.110	<.01 **
4-year-olds	-5.8915	0.9763	-6.034	<.001 ***
5-year-olds	-3.6521	0.8968	-4.072	<.001 ***

[d] Since we analyzed the children's data twice (see Table 4 below), we have adjusted the significance threshold for the p-values (Bonferroni correction: .05/2 = .025). Adults and true belief were used as reference levels and fixed effects were dummy coded.

Open belief	2.6972	0.2580	10.453	<.001 ***
False belief	2.8393	0.2595	10.942	<.001 ***

To check whether adults and older children were more sensitive to the different belief conditions because they generally have a better understanding of mental states than younger children, we also tested whether the main effect of age interacts with the main effect of belief. Adding this interaction term did significantly improve the model, but caused convergence problems. To investigate this potential interaction, we conducted pairwise comparisons between the different belief conditions within each of the three age groups (collapsing across complementizer presence and absence because we found no main effect for complementizer presence). T-tests indicated that all age groups produced significantly more main-subordinate responses in the false-belief than in the true-belief condition (adults: $t = 4.08$, $df = 23$, $p < .001$; 5-year-olds: $t = 4.69$, $df = 24$, $p < .001$; 4-year-olds: $t = 3.14$, $df = 24$, $p = .004$), and in the open-belief than in the true-belief condition (adults: $t = 3.80$, $df = 23$, $p < .001$; 5-year-olds: $t = 3.50$, $df = 24$, $p = .002$; 4-year-olds: $t = 3.03$, $df = 24$, $p = .006$).[e] In addition, whereas the two child groups did not show a significant difference between the open-belief and false-belief conditions (5-year-olds: $t = 1.30$, $df = 24$, $p = .21$; 4-year-olds: $t = 1.23$, $df = 24$, $p = .23$), the adults showed a tendency to produce more main-subordinate responses in the false-belief condition than in the open-belief condition ($t = 2.01$, $df = 23$, $p = .057$), which explains the interaction between age and belief.

For the children, we were also interested in whether they would be more likely to produce both the CTP and the complement clause when they had good short-term memory (as measured by sentence repetition) and/or showed good understanding of mental states (as measured in explicit tests of false belief – with a maximum total score of 4). First, we built a model looking at the same random and fixed effects as for the full data set (children and adults combined), by following the same principle of backward selection. As before, due to convergence problems, the models only included random intercepts rather than random

[e] The p-values remain significant after Bonferroni correction for multiple comparisons (.05/9 = .006). True belief was the reference level and fixed effects were dummy coded.

intercepts and slopes. In order to have a more fine-grained measure of age, we used age in months when we looked at the children's data only. We then added children's short-term memory and understanding of mental states to this model to check whether these individual-difference measures would explain any additional variance in children's tendency to produce main-subordinate responses. The final model suggests that children with better short-term memory also produced more main-subordinate responses, whereas their understanding of mental states had no effect and was thus removed from the final model (Table 4).

Table 4

Final GLMM model with individual-difference measures for main-subordinate responses by English child participants

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-13.73554	3.39494	-4.046	<.001 ***
Age (months)	0.08728	0.07386	1.182	0.237
Open belief	2.20103	0.28064	7.843	<.001 ***
False belief	2.22144	0.27482	8.083	<.001 ***
Short-term memory	0.24213	0.07592	3.189	<0.01 **

Adding an interaction between age and short-term memory did not further improve the fit of the model (χ^2 = 1.608; *df* = 1; *p* = .205). Finally, even though children's general understanding of mental states in explicit false-belief tests did not seem to contribute to performance, we also checked whether it would interact with belief as embedded in the different story contexts. That is, children who show better understanding of mental states in explicit false-belief tests might also produce more main-subordinate responses in the false- than in the true- or open-belief conditions, compared to children with lower scores in the explicit false-belief tests. However, adding this interaction led to convergence problems. In addition, we found no correlation between children's tendency to produce a main-

subordinate construction and their general understanding of mental states within or across the different belief conditions. Overall, the Spearman correlation co-efficient between children's general understanding of mental states and number of main-subordinate responses was -.05 (-.15 in the true-belief condition, .01 in the false-belief condition, and -.05 in the open-belief condition).

3.2 German Children and Adults

Figure 2 shows the mean number of main-subordinate responses for the German participants in each condition of the complement-clause task. As in English, all age groups were most likely to produce both the CTP and the complement clause in the false-belief conditions, followed by the open-belief conditions. And they were least likely to produce both the CTP and the complement clause in the true-belief conditions. In contrast with the results for English speakers, the presence of the complementizer seemed to markedly increase the 5-year-olds' and adults' production of main-subordinate responses, but only in the true-belief conditions.

Figure 2

Proportion of main-subordinate responses in each condition for the German-speaking participants. Maximum number of main-subordinate responses in each condition is 3. Compl. = complementizer.

In order to test whether age, belief, and presence of complementizer had a significant effect on participants' tendency to produce main-subordinate responses, we used binomial generalized linear mixed effects models (glmer) in R, following the same principle of backward selection and using the same random effects structure as for the English data. Similar to the results from the English-speaking participants, the number of main-subordinate responses increased with age, and the German-speaking participants were also more likely to produce main-subordinate responses in the false- and open-belief conditions than in the true-belief conditions (see Table 5).

Table 5

GLMM model for main-subordinate responses by German participants[f]

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.1396	0.7793	1.462	0.143650
4-year-olds	-7.6650	1.2791	-5.993	<.001 ***
5-year-olds	-4.0709	1.0976	-3.709	<.001 ***
Open belief	4.0286	0.3570	11.285	<.001 ***
False belief	4.0570	0.3566	11.378	<.001 ***

As for the English data, adding an interaction between age and belief caused convergence problems. In addition, t-tests suggest that all age groups produced significantly more main-subordinate responses in the false-belief than in the true-belief condition (adults: t = 5.20, df = 23, p <.001; 5-year-olds: t = 4.73, df = 23, p <.001; 4-year-olds: t = 3.30, df = 23, p = .003), and in the open-belief than the true-belief condition (adults: t = 5.10, df = 23, p <.001; 5-year-olds: t = 4.21, df = 23, p <.001; 4-year-olds: t = 3.08, df = 23, p = .005).[g] The difference between the open-belief and false-belief conditions was non-significant in each age group (adults: t = -1.55, df = 23, p =.135; 5-year-olds: t =.12, df = 23, p =.903; 4-year-olds: t = -1.36, df = 23, p = .188).

Finally, although we did not make any specific predictions about the possible relation between the presence of a complementizer and the belief state, descriptive statistics (see Figure 2) suggest that it played some role in the older children's and the adults' production of main-subordinate responses in the true-belief conditions. When we added an interaction between belief and complementizer and compared this model to a model containing only

[f] Since we analyzed the children's data twice (see Table 6 below), we have adjusted the significance threshold for the p-values (Bonferroni correction: .05/2 = .025). Adults and true belief were used as reference levels and fixed effects were dummy coded.

[g] The p-values remain significant after Bonferroni correction for multiple comparisons (.05/9 = .006).

the main effects of age, belief, and complementizer, the model with the interaction was marginally better than the model without this interaction (χ^2 = 5.84; *df* = 2; *p* = .05). Pairwise comparisons suggest that, in the true-belief conditions, adults were more likely to repeat both the CTP and the complement clause when the complementizer was present than when it was absent (*t* = 3.19; *df* = 23; *p* = .004. For the 5-year-olds, the same trend was not significant (*t* = 1.99; *df* = 23; *p* = .059).

We then built a binomial generalized linear mixed model (glmer) to only look at the children's data and to investigate whether any of the individual-difference measures would improve the fit of the model. First, we built a model looking at the same random and fixed effects as for the full data set (German-speaking children and adults combined), by following the same principle of backward selection and using the same random effects structure. For the German children's data, we had to use age in years, rather than months, because age in months caused convergence problems.

We then added children's short-term memory and understanding of mental states in explicit false-belief tests (with a maximum score of 4) to this model. The final model (Table 6) suggests that children with better short-term memory produced more main-subordinate responses, whereas their general understanding of mental states had no effect. In addition, we found a marginally significant interaction between age and short-term memory, indicating that the older children tended to produce more main-subordinate responses when they also had better short-term memory. The interaction between children's general understanding of mental states and the different belief conditions in the story was not significant. As in English, children who performed better in explicit false-belief tasks did not produce more main-subordinate responses when the story character had a false or open belief. This factor was therefore taken out of the final model.

Table 6

Final GLMM model with individual-difference measures for main-subordinate responses by German child participants[h]

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-9.42425	1.70827	-5.517	<.001 ***
5-year-olds	-9.21004	5.83475	-1.578	0.11446
Open belief	3.55054	0.40359	8.797	<.001 ***
False belief	3.35835	0.38716	8.674	<.001 ***
Short-term memory	0.23699	0.09028	2.625	<.01 **
Age (5)*Short-term memory	0.47847	0.26228	1.824	0.06811 .

3.3 Combined Data Analysis

In order to check whether there were any differences between the English- and German-speaking participants' responses, we built models with language group added as a main effect. Using the same random effects structure and following the same principle of backward selection as before, the final model with both child and adult participants included showed main effects for age and belief, but no main effect for language group or complementizer presence (Table 7). In addition, we also tested for an interaction between language group and complementizer presence, but this turned out to be non-significant, and the model failed to converge. There was thus no difference between the two language groups. Both groups produced more main-subordinate responses when the character had a false or open belief, and the production of main-subordinate responses increased with age.

Table 7

GLMM model for main-subordinate responses by English and German participants

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.6733	0.5154	3.247	<.01 **

[h] When we used Age (months), the model failed to converge. Therefore, we used age groups for this model. True belief was the reference level and fixed effects were dummy coded.

4-year-olds	-6.5213	0.7633	-8.544	<.001 ***
5-year-olds	-3.7931	0.6898	-5.499	<.001 ***
Open belief	3.2519	0.2069	15.721	<.001 ***
False belief	3.3588	0.2073	16.199	<.001 ***

For the children's data, we also checked whether adding short-term memory and general understanding of false belief would improve the fit of the model. Replicating what we found when we analyzed the English and German data separately, only short-term memory turned out to be significant (see Table 8). The interaction between short-term memory and age was non-significant, and testing for an interaction between understanding of mental states and belief was not possible because the model failed to converge.

Table 8

Final GLMM model with individual-difference measures for main-subordinate responses by English and German child participants

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-13.95402	3.17900	-4.389	<.001 ***
Age (months)	0.11386	0.06245	1.823	.068 .
Open belief	2.73544	0.22831	11.981	<.001 ***
False belief	2.66409	0.22088	12.061	<.001 ***
Short-term memory	0.25697	0.06076	4.229	<.001 ***

3.4 Errors and Other Responses

In our novel complement-clause task, all age groups adapted their responses to the knowledge state of the main character. However, partly due to limited memory resources, the child participants were not always able to repeat the main- and subordinate clause or even just the subordinate clause. In the false-belief scenarios, just repeating the subordinate clause is actually not an appropriate response, as we will illustrate in the next

section, where we will present a qualitative analysis of errors and other responses made by children and adults. We will first describe and discuss the most common error types and other responses that children made in the two conditions where the story character had a false belief. As in previous analyses, we collapsed across the complementizer present and absent conditions. We coded the errors and other responses as 'Repeats False Belief', 'Repeats Reality Statement', or 'Describes Image' (see Figure 3 for the distribution of these error types across the two language and child age groups). Adults did not make any of these errors or produce any of these alternative responses.

In the responses coded as 'Repeats False Belief', children just repeated the complement clause. That is, children repeated the main character's belief even though it had turned out that it was a false belief. For example, children heard the following story:

> "This is Sam. Sam likes Jane. Sam wants to play hockey with Jane. But he thinks that she's still at school. Oh look, she actually isn't still at school".

When asked "Why doesn't Sam play hockey with Jane?", they answered "because she's still at school". This type of answer is inappropriate in the given context. It also indicates that children did not fully understand that the character had a false belief. Alternatively, they might not have been able to express this understanding linguistically, or they were unable to coordinate their linguistic behavior with their understanding of false belief. It should be noted that, in the false-belief scenarios, the only appropriate answer was a complex main-subordinate response, whereas in the open- and true belief conditions, responding with just the subordinate clause could be considered appropriate as well. Therefore, providing an appropriate response in the false-belief scenarios was linguistically and cognitively more demanding than in the other conditions. As shown in Figure 3, the error of only repeating the complement clause (Repeats False Belief) was much more common in the 4-year-olds than in the 5-year-olds, especially in the English group, indicating that between the ages of 4 and 5 years, children advance in their understanding of false belief and/or their ability to express this understanding linguistically. That is, once children

understand that the character's belief is false, they should not repeat it without explicitly referring to the character's perspective. However, referring to the character's perspective is only possible by repeating the whole complement-clause construction, which may also be constrained by limited short-term memory. Why the English-speaking 4-year-olds were more likely to produce this error than the German-speaking 4-year-olds is not clear. Overall, both groups seemed to be equally likely (or unlikely) to produce complement-clause constructions (see number of main-subordinate responses in Figures 1 and 2). In terms of their performance in the explicit false-belief tests, we could not see any cross-linguistic differences either (English 4-year-olds *M* score = 2.3; German 4-year-olds *M* score = 2.0). In addition, both English and German 4-year-olds showed similar performance on the short-term memory task. If anything, the English 4-year-olds performed slightly better (English 4-year-olds *M* score = 34.6 out of 47 (74%); German 4-year-olds *M* score = 28.2 out of 46 (61%)).

Another type of response in the false-belief scenarios was to repeat the reality statement from the story (e.g., "because she actually isn't at school"). Again, this response is inappropriate in the given context and could indicate that children did not understand the task. However, it could be argued that children who produced this type of answer seemed to have some understanding that the character had a false belief and chose to draw attention to this by referring to the contrasting true state of affairs, despite this being an inappropriate response to the question. This type of response also indicates that it takes time for children to be able to coordinate their understanding of belief and knowledge states with their linguistic behavior. As suggested above, children might have responded in this way because providing the appropriate response (i.e., a complex main-subordinate construction) was too demanding for them. Unlike for the 'Repeat False Belief' error, we did not see any developmental trends for this response type, but it was also less common overall (see Figure 3).

In the 'Describes Image' category, children described the image that accompanied the reality statement rather than repeating the reality statement. For example, when children heard "she actually isn't at school", they saw a picture of Jane playing hockey on her own.

So, when they were asked “Why doesn’t Sam play hockey with Jane?”, they answered “because she’s playing hockey already”. This response type is similar to the ‘Repeat Reality’ type in that children may have described the image to draw attention to the contrasting state of reality even though this was not an appropriate response to the question. This could also suggest that these children had some understanding of false belief, but were unable to express this with a complex syntactic structure.

Overall, there was a decrease in inappropriate responses, where children just repeated the character’s false belief (i.e. the ‘Repeat False Belief Error’). However, both 4- and 5-year-olds sometimes struggled to respond with a complex main-subordinate construction and described the real state of affairs instead (i.e. ‘Repeats Reality Statement’ and ‘Describes Image’ Errors).

Figure 3

Percentages of children’s errors and other responses in the false-belief scenarios (with and without complementizers) in English and German. Overall, the English children produced a total of 150 responses in the false-belief scenarios, and the German children produced a total of 144 responses. Note that the correct main-subordinate responses are not displayed in this figure.

In addition to the target responses and the errors in the false-belief scenarios analyzed and described above, both children and adults also made ‘Belief Comments’ across all conditions. We further categorized these ‘Belief Comments’ as confirmations, rejections, or epistemic markers. To illustrate these, we provide a list of examples (in bold) from the English child participants below:

Confirmation:	“Because it… he thinks it’s closed **and it is**.”
	“Because she thinks that he’s poorly **but (s)he really is**.”
Rejection:	“Because she thinks he’s on holiday **but he’s not**.”
	“Because it’s raining **but it isn’t raining**.”

Epistemic marker: "Because **maybe** he's out holiday."

"Because the red slide **might** be dirty."

"**I think** Sam's already there."

As can be seen in the examples above, participants sometimes confirmed the character's belief (e.g., "and it is"), or they made it explicit that the main character had a false belief by rejecting that belief (e.g., "but he's not"). Some participants also used adverbials (e.g., "maybe"), modals (e.g., "might") or mental verbs (e.g., "I think") to indicate that they weren't sure whether the main character's belief was true or false. In other words, they used an epistemic marker in their response. Overall, especially the 5-year-olds made these "Belief Comments", and, across age groups, they were most frequent in the false-belief scenarios (see Table 9). These comments were made together with both erroneous responses and main-subordinate or subordinate-only responses.

Table 9

Number of different Belief Comments in English and German.[i]

		Open Belief			False Belief			True Belief		
		confirm	reject	epistemic	confirm	reject	epistemic	confirm	reject	epistemic
English	4-year-olds	1	0	2	1	2	0	5	0	0
	5-year-olds	0	1	7	1	29	4	21	2	0
	Adults	0	0	3	0	7	2	2	0	1
German	4-year-olds	1	4	12	0	33	5	12	1	12
	5-year-olds	0	1	4	0	43	2	24	0	4
	Adults	0	0	0	0	7	2	4	0	4

Table 9 illustrates that when the truth value of the main character's belief was left open (i.e., in the open-belief conditions), the most common – and appropriate – 'Belief Comment' across all age groups was to use an epistemic marker to express some

[i] Overall, for the English 4- and 5-year-olds, there were a total of 150 responses in each condition. For the English adults and both German children and adults, there were a total of 144 items in each condition.

uncertainty about the character's belief. When the main character had a false belief, all age groups were most likely to reject that false belief when they made a 'Belief Comment'. Interestingly, children also sometimes combined this explicit rejection with one of the inappropriate answers that we analyzed in the previous section. For example, in one of the false-belief scenarios, the main character did not play outside because they falsely thought that it was raining. When asked why the character did not play outside, one 5-year-old said "because it's raining, but it isn't raining". This, again, illustrates that it takes time for children to be able to express their understanding of beliefs and knowledge states linguistically and to coordinate their socio-cognitive and linguistic skills. Finally, when the main character had a true belief, the most common – and appropriate - 'Belief Comment' across all age groups was to confirm this belief.

Why the German-speaking children seemed more inclined to make belief comments than the English-speaking children is not clear. As mentioned above, the German- and English-speaking 4-year-olds did not differ in their general understanding of mental states, and the same was found for the older children's performance in the explicit false-belief tasks (English-speaking 5-year-olds *M* score = 2.9; German-speaking 5-year-olds *M* score = 3.0). In addition, the 5-year-olds in both language groups showed similar performance in the short-term memory task (English 5-year-olds *M* score = 40.3 out of 47 (86%); German 5-year-olds *M* score = 42 out of 46 (89%).

4. Discussion

The current study had three main aims: First, similar to Lewis et al. (2017), we wanted to investigate whether children are able – or at least more likely – to focus on (and repeat) CTPs (complement-taking predicates such as "I think") when the task is designed in such a way that it draws their attention to different beliefs and mental states.

Second, we were interested to see whether children would be more likely to focus on (and repeat) CTPs when they show a general understanding of mental states in explicit false-belief tests, and when the character's belief is false. To investigate how the general understanding of mental states affects the likelihood of paying attention to CTPs, we also compared different age groups (i.e. 4- and 5-year-old children and adults), and investigated whether adults and older children and/or children who perform better in explicit false-belief tests would also be more sensitive to whether the character had a true, false, or open belief.

Finally, we asked whether children and adults would be more likely to focus on (and repeat) the CTP when the complement clause is introduced by a complementizer and whether German-speaking participants would show a greater sensitivity to the presence or absence of the complementizer - and the accompanying changes in word order - than their English-speaking peers.

In the remainder of the Discussion, we will summarize and discuss our results for each of these research aims.

4.1 Focusing on CTPs in a Task that Draws Attention to Different Mental States

Similar to the complement-clause task developed by Lewis et al. (2017), our task involved different characters with different beliefs (false, true, and open), in order to draw children's attention to these contrasting mental states. Our results also align with the findings presented by Lewis and colleagues: Just around their fourth birthday, children are sensitive to mental states when they interpret and produce complement-clause constructions, at least in a task that draws their attention to different beliefs. In our task and the one designed by Lewis and colleagues, children were able to take the CTP into

account when they interpreted and produced complement-clause constructions, even when they were at an age at which their understanding of mental states is still developing (Wellman et al., 2001). This challenges the assumption that early in development children do not use or interpret CTPs to refer to mental states because they do not have a full understanding of these mental states up until the age of five years (e.g., Bartsch & Wellman, 1995; Shatz et al., 1983). As suggested elsewhere (e.g., Amsterlaw & Wellman, 2006; Boeg Thomsen et al., 2021; Brandt et al., 2010; Diessel & Tomasello, 2001), both children's Theory of Mind development and their acquisition of complement-clause constructions are gradual processes and are likely to interact with each other before children show a stable understanding of false belief or other mental states in explicit tests. However, as will be discussed in the next section, we also found that older children and adults were more likely to produce the CTPs in their responses than four-year-olds, which could be due to better understanding of mental states and / or more advanced development of other cognitive skills, such as short-term memory.

4.2 Coordinating False-Belief Understanding and Linguistic Knowledge

In our complement-clause task, both children and adults were most likely to repeat the whole sentence (e.g., "Because he thinks that it's raining") and include the CTP ("he thinks") when the story character had a false belief, or when it was unclear whether their belief was true or false. When the story character had a true belief, children and adults often left out the CTP and just repeated the subordinate clause (e.g., "because it's raining"). This suggests that, in our task, even the 4-year-olds were able to adjust their response according to different belief states, which, in turn indicates a developing understanding of false belief. In addition, these results also indicate that all age groups made a distinction between relevant and irrelevant information (cf. Sperber & Wilson, 1986). As suggested by an anonymous reviewer, it is more relevant to repeat the CTP when the belief is false or open than when it is true.

However, our results also showed that older participants were overall more likely to include the CTP in their answer. One explanation for this is that remembering and producing

a complement-clause construction with a CTP makes more demands on short-term memory than just repeating the complement clause. In support of this, we also found a main effect of short-term memory in the children's data. In addition, our analysis of errors and other responses suggests that even when children were sensitive to the characters' belief states, they sometimes struggled to produce a complement-clause construction to express this and used alternative - less complex - linguistic structures instead. For example, children were more likely to reject a false belief than a true belief (e.g., "Because it's raining but it isn't raining") (see Table 9).

Another possible explanation for the finding that the older children and adults were more likely to pay attention to the CTP and include it in their answers is that they have a better understanding of mental states. This is supported by interactions between age and belief: For the English participants we found that only the adults had a tendency to produce more main-subordinate responses (and include the CTP) in the false-belief than in the open-belief condition. For the German participants we found that the adults showed a bigger difference between the open-belief and true-belief conditions in their likelihood of including the CTP in their response. We should acknowledge though that differences between children and adults in the open-belief condition could have also been caused by the fact that some aspects of this condition were pragmatically odd. More specifically, participants heard a comment about the color of a star before they heard the test question. The picture of the star and the comment about its color were not related to the story and might have distracted and confused the child participants in particular.

However, all age groups showed clear differences between the false-belief and true-belief conditions. More importantly, we found no main effect of general false belief in the children's data. This indicates that children who performed better in the explicit false belief tests were *not* more likely to focus on and produce the CTP in their responses. This is in line with findings presented by Lewis et al. (2017), but does not align with earlier studies that found correlations and longitudinal relationships between children's comprehension of complement-clause constructions and their general false-belief understanding (de Villers & Pyers, 2002).

There are (at least) two possible reasons for why we have not found any significant links between children's comprehension and production of complement-clause constructions and their performance in explicit false-belief tests. First, we will consider the possibility that there is no unique relation between complement-clause constructions and false-belief reasoning (cf. Cheung et al., 2004; Perner et al., 2003). Second, as has been suggested before (e.g., Lewis et al., 2017; Papafragou, Li, Choi, & Han, 2007), we will argue that different tasks testing children's linguistic and socio-cognitive skills might yield different results, and that it might take children some time to coordinate their linguistic and socio-cognitive skills across various tasks (i.e., choose and produce appropriate linguistic constructions to map onto different mental states and vice versa).

According to de Villiers and Pyers (2002), complement-clause constructions are a unique tool that allows children (and adults) to represent others' mental states, even when these mental states differ from their own mental states or from reality (see also de Villiers, 2007). Empirical evidence for this claim first came from a longitudinal study, where children's comprehension of the structure of complement-clause constructions predicted their false-belief understanding (see Introduction; for similar results from a more recent study with additional control variables see Boeg Thomsen et al., 2021). However, a number of follow-up studies have come to the conclusion that other linguistic structures and skills can also facilitate false-belief reasoning. In the training study by Lohmann and Tomasello (2003), for instance, children also improved in their false-belief reasoning when they were trained with simple linguistic structures. Similarly, a number of studies with Cantonese- and English-speaking children by Cheung and colleagues (e.g., Cheung et al., 2004; Cheung, 2006) suggest that children's general language comprehension, as measured in standardized tests of grammar and vocabulary, is a better predictor of their false-belief reasoning than the more specific measure of complement-clause comprehension. Similar results have been reported for Dutch, a language closely related to German (and English) (De Mulder et al., 2019). In the current study, we used a sentence-repetition task to measure children's short-term memory. It could be argued that this task also taps into verbal working memory and as such is also dependent to some extent on children's general language skills. For both

English- and German-speaking children, we found that their performance in this sentence-repetition task correlated with their performance in the explicit false-belief tasks (English: r_s =. 44; p = .001; n = 50; German: r_S = .48; p <.001; n = 48). Crucially, the sentence-repetition task did not involve any complement-clause constructions. We would thus argue that our results support the assumption that children's false-belief understanding is related to their general language skills rather than their comprehension or production of complement-clause constructions. Given that we also found a main effect of short-term memory in children's likelihood of repeating the whole complement-clause construction, we could argue that both children's false-belief understanding and their processing of complement-clause syntax is constrained by more general processing skills (cf. Lutken et al., 2020).

Turning to the second possibility, whether we find correlations between children's understanding of false belief and their acquisition of complement-clause constructions can also depend on the tasks that we use to measure these skills. First of all, whereas previous research on children's acquisition of complement-clause constructions and false-belief understanding tested children's *comprehension* of complement clauses, we tested both their *comprehension* and their elicited *production* (i.e. repetition) of this syntactic construction within the same task. As has been suggested for other linguistic markers of mental states, comprehension tasks are often more difficult and meta-linguistic than production tasks. For example, in one of their comprehension tasks, Papafragou et al. (2007) asked children to map an utterance with a specific evidential marker onto one of two characters who had either direct or indirect evidence of the content of a box (see also Davis & Landau, 2021). In another comprehension task, children had to judge whether an utterance with an evidential marker was correct or incorrect, which is similar to how Lewis et al. (2017) tested children's interpretation of complement-clause constructions. They found that Korean children's production precedes their comprehension of evidential morphology (for similar findings with Turkish-speaking children see Ozturk & Papafragou, 2016). In addition, they found positive relations between Korean children's production, but not their comprehension, of evidential morphology and their performance in source-monitoring tasks.

Even though our task involved comprehension and production of complement-clause constructions, it was less meta-linguistic than the comprehension tasks used to test children's acquisition of evidential morphology (e.g., Papafragou et al., 2007) or the truth-value judgment task used to test children's understanding of complement-clause constructions (Lewis et al., 2017; see also Davis & Landau, 2021; Harrigan et al., 2019). In contrast with most previous studies that presented complement-clause constructions in isolation (e.g., de Villiers & Pyers, 2003), we presented them in a story context. In addition, children were not directly asked to repeat or judge the truth value of complement-clause constructions. We tested their comprehension and production in a more indirect way. This gives the task a more implicit flavor and might explain why it did not correlate with the more explicit false-belief measures.

As recently suggested by Rubio-Fernández, Mollica, Ali, and Gibson (2019), adults automatically track beliefs in everyday conversation. And implicit tests of false belief suggest that infants and young children also have an automatic system to track others' beliefs (for overviews see Baillargeon et al., 2016; Scott & Baillargeon, 2017). Most previous studies on false belief and complement-clause constructions have used explicit measures for both skills. Our study is the first one to investigate how a more implicit or indirect measure of complement-clause acquisition might relate to performance in explicit false-belief tests.

There are a few studies that have investigated the relationship between explicit language measures and implicit and explicit false-belief tests, and they show mixed results. Low (2010) used the test of complement-clause comprehension developed by de Villiers and Pyers (2002), which we described and discussed in the Introduction, and found no correlations between children's understanding of complement-clause constructions and their anticipatory looking behavior in an implicit false-belief task. At the same time, similar to the original findings (de Villiers & Pyers, 2002), children's comprehension of complement-clause constructions correlated with their performance in explicit tests of false belief. Similar dissociations between implicit and explicit false-belief tests in their relation to language have been reported by Grosse Wiesmann, Friederici, Singer, and Steinbeis (2017). To our knowledge, the only study, so far, that found a positive correlation between

language and implicit false-belief understanding is a training study by San Juan and Astington (2017). They found that an exposure to epistemic verbs during training led to improvements in an implicit false-belief task (anticipatory looking in a change-of-location scenario). In addition, vocabulary knowledge was also a significant predictor of children's performance in this implicit false-belief task.

Another recent study suggests that children's sensitivity to syntactic features of complement-clause constructions could be related to socio-cognitive precursors of false-belief understanding. In a longitudinal study, Kaltefleiter, Sodian, Kristen-Antonow, Grosse Wiesmann, and Schuwerk (2021) found that German-speaking children who were able to repeat or revert to the right word order in the complement clause (i.e., verb-final for complement clauses with a *that* complementizer and verb-second for complement clauses without a *that* complementizer) were also better at perspective-taking and understanding ignorance three months later. What we don't know yet is whether more implicit tests of language (including complement clauses) also show a relation to implicit tests of false belief. For future research, we should develop additional implicit tasks to look at children's developing understanding of mental-state language. For example, we could use anticipatory-looking paradigms to test how infants and young children interpret complement-clause constructions. This would then allow us to investigate whether and how children 's behavior in implicit language tasks relates to their behavior in implicit and explicit false-belief tasks.

We should also acknowledge that combining measures of false-belief understanding and complement-clause comprehension and production within the same task, as we did in our novel task, might have led to potential confounds. As discussed in the Introduction, previous studies on complement-clause comprehension and false-belief development have been criticized for having false-belief scenarios in their complement-clause task, thus confounding the two measures. Similarly, many of the explicit false-belief tasks also contain complement-clause constructions in their test questions. A recent longitudinal study with English-speaking children suggests that the relationship between complement-clause constructions and false belief holds even when these confounds are excluded (Boeg Thomsen et al., 2021).

Nevertheless, it would be good to also test, for example, whether and how children's performance in our novel task would relate to performance in a task that tests the comprehension or production of complement-clause constructions without a false-belief scenario.

Before we conclude, we will briefly discuss our final research question, namely whether the presence of the *that* complementizer would affect the likelihood that children and adults focus on and repeat the CTP, and whether there are cross-linguistic differences.

4.3 Form-Function Mapping within and across Languages

Based on previous literature (e.g., Diessel & Tomasello, 2001; Thompson & Mulac, 1991), we hypothesized that participants would be more likely to interpret the CTP as a parenthetical marker and leave it out in their responses when the following complement clause was not introduced by a *that* complementizer. In English, we found no main effect for complementizer absence. In German, we did not find a significant main effect for complementizer absence either. In addition, the combined data analysis did not show a significant main effect for complementizer presence or an interaction between language group and complementizer presence. However, in the German data, there was a marginally significant interaction between belief and complementizer and post-hoc pairwise comparisons indicate that, when the character had a true belief, adults were more likely to leave out the CTP when the *that* complementizer was absent.

These findings indicate that complementizer absence might have a bigger effect in verb-second languages, where the absence also leads to word-order changes. As briefly discussed in the Introduction, this finding could also suggest that it is the verb-second word order rather than the lack of complementizer that drives this effect. Future studies could disentangle the effects of complementizer presence/absence and word order. Future research could also investigate whether other formal features (e.g., complement preposing) would make English- and German-speaking children and adults more likely to interpret the complement clause as the main point of utterance and leave out the CTP.

In the current study, the fact that only adults showed some sensitivity to whether the complementizer was present also indicates that it takes time for children to become sensitive to subtle differences in form and how these formal differences might map onto different functions. However, complementizer absence is not the only formal feature that might make it more likely for interlocutors to interpret the CTP as a parenthetical marker that could be left out. Another feature is the use of first-person subjects and frequent verbs in present tense (e.g., "I think") (e.g., Diessel & Tomasello, 2001). In the current study, we used the frequent verb "think", but it was always used together with a third-person subject (e.g., "Tom thinks"). Further research is needed to investigate whether both English- and German-speaking children and adults would be more likely to leave out CTPs with first-person subjects.

5. Conclusions

With our newly developed task, we have demonstrated that English- and German-speaking children, as young as 3;9, and adults are sensitive to story characters' belief states when they interpret and produce complement-clause constructions, as demonstrated by their selective production of full complement clauses, and their explanations and erroneous responses that indicate some partial knowledge. However, this understanding develops further with age and increasing short-term memory. Children's understanding of mental states as measured in explicit false-belief tasks, however, did not affect their behavior in our complement-clause task. Finally, the absence of the *that* complementizer only had a limited effect on the adults in the German group, who were more likely to leave out the CTP when the character had a true belief and the complementizer was missing. These findings suggest (1) that it takes time for children to produce the right linguistic form and map it onto different belief states (2) that children's understanding of both mental-state language and false belief differs across implicit and explicit tasks, and (3) that cross-linguistic differences in the form of mental-state language can also lead to cross-linguistic differences in adults' (and older children's) interpretation of mental-state language.

References

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10 month-old infants. *Journal of Experimental Child Psychology*, *151*, 5-17. https://doi.org/10.1016/j.jecp.2015.11.013

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

Antomo, M. & Steinbach, M. (2010). Desintegration und Interpretation: *Weil*-V2-Sätze an der Schnittstelle zwischen Syntax, Semantik und Pragmatik. *Zeitschrift für Sprachwissenschaft*, *29*(1), 1-37. https://doi.org/10.1515/zfsw.2010.001

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

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* Oxford University Press.

Bergen, B. (2015). Embodiment. In E. Dabrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 10-30). De Gruyter Mouton. https://doi.org/10.1515/9783110292022-002

Besten, Hans den. 1983. On the Interaction of Root Transformations and Lexical Deletive Rules. In W. Abraham (Eds.), *On the Formal Syntax of the Westgermania* (pp. 47-131). John Benjamins.

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*.

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

Brandt, S., Diessel, H., & Tomasello, M. (2008). The acquisition of German relative clauses: A case study. *Journal of Child Language*, *35*(2), 325-348. https://doi.org/10.1017/S0305000907008379

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

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337-342. https://doi.org/10.1016/j.cognition.2009.05.006

Cheung H. (2006). False belief and language comprehension in Cantonese-speaking children. *Journal of Experimental Child Psychology*, *95*(2), 79-98. https://doi.org/10.1016/j.jecp.2006.05.002

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

Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation language in Theory-of-Mind development: Evidence from Cantonese and English. *Child Development, 75*(4), 1155-1170. https://doi.org/10.1111/j.1467-8624.2004.00731.x

Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language*, 26(2), 207–232. doi:10.1177/0142723706060748

De Mulder, H. N., Wijnen, F., & Coopmans, P. H. (2019). Interrelationships between Theory of Mind and language development: A longitudinal study of Dutch-speaking kindergartners. *Cognitive Development*, *51*, 67-82. https://doi.org/10.1016/j.cogdev.2019.03.006

de Villiers, J.G. (2007). The interface between language and Theory of Mind. *Lingua, 117*(11), 1858-1878. doi: 10.1016/j.lingua.2006.11.006

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1)*,* 1037–1060. https://doi.org/10.1016/s0885-2014(02)00073-4

Davis, E. E., & Landau, B. (2021). Seeing vs. seeing that: Children's understanding of direct perception and inference reports. *Experiments in Linguistic Meaning, 1*, 125-135. https://doi.org/10.3765/elm.1.4864

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

Dor, D. (2005). Toward a Semantic Account of *that*-Deletion in English. *Linguistics*, *43*(2), 345-382. https://doi.org/10.1515/ling.2005.43.2.345

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development, 46,* 12-30. https://doi.org/10.1016/j.cogdev.2018.01.001

Dudley, R., Rowe, M., Hacquard, V., & Lidz, J. (2017). Discovering the factivity of *know*. *Proceedings of SALT, 27*, 600-619.

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

Fritzenschaft, A., Gawlitzek-Maiwald, I., Tracy, R. & Winkler, S. (1990). Wege zur komplexen Syntax. *Zeitschrift für Sprachwissenschaft*, *9*(1-2), 52-134. https://doi.org/10.1515/zfsw.1990.9.1-2.52

Grimm, H., Aktas, M., Frevert, S. (2001). *SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder: Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. Hogrefe.

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

Hacquard, V., & Lidz, J., (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language, 34*(1), 73-96. https://doi.org/10.1111/mila.12192

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

Harrigan, K., Hacquard, V., & Lidz, J. (2019). Hope for syntactic bootstrapping. *Language, 95*(4), 642-682. doi:10.1353/lan.2019.0069.

Harris, P. L., Yang, B., & Cui, Y. (2017). 'I don't know': Children's early talk about knowledge. *Mind & Language, 32*(3), 283-307. https://doi.org/10.1111/mila.12143

Heycock, C. (2017). *Embedded root phenomena*. The Wiley Blackwell Companion to Syntax,

Second Edition, 1-37.
Hooper, J. B., & Thompson, S. A. (1973). On the applicability of root transformations. *Linguistic Inquiry, 4*(4), 465-497. https://www.jstor.org/stable/4177789
Howard Gola, A. A. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development, 27*(1), 64–76. https://doi.org/10.1016/j.cogdev.2011.10.003
Howard, A. A., Mayeux, L., & Naigles, L. R. (2008). Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language*, *28*(4), 375-402. https://doi.org/10.1177/0142723708091044
Kaltefleiter, L. J., Sodian, B., Kristen-Antonow, S., Grosse Wiesmann, C., & Schuwerk, T. (2021). Does syntax play a role in Theory of Mind development before the age of 3 years?. *Infant behavior & development*, *64*, 101575. https://doi.org/10.1016/j.infbeh.2021.101575
Kaltenböck, G. (2006). '… That is the question': complementizer omission in extraposed that-clauses. *English Language and Linguistics*, *10*(2), 371-396. http://doi.org/10.1017/S1360674306001961
Kärkkäinen, E. (2003). *Epistemic stance in English conversation: A description of its interactional functions, with a focus on* I think. John Benjamins. https://doi.org/10.1075/pbns.115
Kidd E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171-184. https://doi.org/10.1037/a0025405
Langacker, R.W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
Lewis, S., V. Hacquard & J. Lidz (2017). Think pragmatically: Children's interpretations of belief reports. *Language Learning and Development*, *13,* 357-374. DOI:10.1080/15475441.2017.1296768
Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130–1144. https://doi.org/10.1111/1467-8624.00597
Low J. (2010). Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, *81*(2), 597-615. https://doi.org/10.1111/j.1467-8624.2009.01418.x
Lutken, C. J., Legendre, G., & Omaki, A. (2020). Syntactic Creativity Errors in Children's Wh-Questions. *Cognitive Science*, *44*(7), e12849.
Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*(2), 199-230. https://doi.org/10.1080/15475441.2015.1024834
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*(1), 50-61. https://doi.org/10.1016/j.cogdev.2013.11.001
Noonan, M. (1985). Complementation. In T. Shopen (Ed.), *Language Typology and Syntactic Description Volume 2: Complex Constructions* (pp. 42–140). Cambridge University Press.
Okuno, A., Cameron-Faulkner, T. R., & Theakston, A. L. (2019). Crosslinguistic differences in the encoding of causality: Transitivity preferences in English and Japanese children and adults. *Language Learning and Development, 16*(1), 61-88. https://doi.org/10.1080/15475441.2019.1685387

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, *308*(5719), 255-258. https://doi.org/10.1126/science.1107621

Papafragou, A., Li, P., Choi, Y., & Han, C.-h. (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

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

Poeppel, D., & Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language*, *69*(1)1-33. https://doi.org/10.2307/416414

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development, 46,* 69-78. https://doi.org/10.1016/j.cogdev.2017.08.002

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*(1), 59-74. https://doi.org/10.1111/j.2044-835X.2011.02061.x

Rubio-Fernández, P., Mollica, F., Ali, M. O., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition, 193,* Article 104011. https://doi.org/10.1016/j.cognition.2019.104011

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development, 18*(2), 139–158. https://psycnet.apa.org/doi/10.1016/S0885-2014(03)00002-9

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, 41,* 19-32. https://doi.org/10.1016/j.cogdev.2016.12.003

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

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*(3), 301-321. https://doi.org/10.1016/0010-0277(83)90008-2

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, *117*(6), 1034-1056. https://doi.org/10.1016/j.lingua.2006.05.006

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

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

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

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298-318). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0014

Thompson, S.A. (2002). "Object complements" and conversation: Towards a realistic account. *Studies in Language, 26*(1), 125-163. https://doi.org/10.1075/sl.26.1.05tho

Thompson, S.A, & Mulac, A. (1991). The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics, 15*(3), 237-251. https://doi.org/10.1016/0378-2166(91)90012-M

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

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

Wiig, E. H., Semel, E. M., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals (CELF) Preschool 2*. Pearson Education/Psychorp.

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