



PAPER

Implicit and explicit false belief development in preschool children

Charlotte Grosse Wiesmann,^{1,2} Angela D. Friederici,¹ Tania Singer³
and Nikolaus Steinbeis^{3,4}

1. Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

2. Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Germany

3. Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

4. Department of Developmental and Educational Psychology, Universiteit Leiden, The Netherlands

Abstract

The ability to represent the mental states of other agents is referred to as Theory of Mind (ToM). A developmental breakthrough in ToM consists of understanding that others can have false beliefs about the world. Recently, infants younger than 2 years of age have been shown to pass novel implicit false belief tasks. However, the processes underlying these tasks and their relation to later-developing explicit false belief understanding, as well as to other cognitive abilities, are not yet understood. Here, we study a battery of implicit and explicit false belief tasks in 3- and 4-year-old children, relating their performance to linguistic abilities and executive functions. The present data show a significant developmental change from failing explicit false belief tasks at 3 years of age to passing them at the age of 4, while both age groups pass implicit false belief tasks. This differential developmental trajectory is reflected by the finding that explicit and implicit false belief tasks do not correlate. Further, we demonstrate that explicit false belief tasks correlate with syntactic and executive functions, whereas implicit false belief tasks do not. The study thus indicates that the processes underlying implicit false belief tasks are different from later-developing explicit false belief understanding. Moreover, our results speak for a critical role of syntactic and executive functions for passing standard explicit false belief tasks in contrast to implicit tasks.

Research highlights

- We provide comprehensive evidence for a dissociation of explicit false belief understanding and earlier-developing implicit anticipation of the actions of an agent with a false belief in 3- and 4-year-old children.
- We demonstrate that performance on standard explicit false belief tasks depends on syntax and executive functions, while implicit false belief tasks do not.
- We show that explicit false belief tasks do not correlate with implicit anticipatory looking false belief tasks.

Introduction

Theory of Mind (ToM) refers to the ability to represent the mental states of other agents – that is, their thoughts,

knowledge, and beliefs. The ability to understand others' false beliefs is considered a crucial test of ToM (Bennett, 1978; Dennett, 1978), and the standard experimental paradigm used to test this ability is the false belief task (Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983). In this paradigm, children are typically presented with a story in which a protagonist misses a piece of information and thus, unlike the child, has a false belief about the situation. The child is then asked to either make an explicit statement about the belief of the protagonist (e.g. *what does the protagonist think?*) or to predict how the protagonist is going to act (e.g. *where will the protagonist look for an object?*). Typically, children do not pass these tests before the age of 4 years (Wellman, Cross & Watson, 2001). It has therefore been argued that between the ages of 3 and 4 years there is a fundamental change in children's understanding of other

Address for correspondence: Charlotte Grosse Wiesmann, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany; e-mail: wiesmann@cbs.mpg.de

agents (Astington & Gopnik, 1991; Flavell, Green & Flavell, 1990; Perner, 1991), and that at this age children start to build representations of others' mental states, which can thus differ from reality. Whether such explicit false belief tasks are a valid measure of ToM abilities has been called into question for a number of reasons (Baillargeon, Scott & He, 2010; Bloom & German, 2000). These authors have argued that responding accurately in these tasks requires sufficient verbal and executive control abilities, which might mask false belief understanding in younger children.

In the past decade, it has been shown that already in their second year of life, infants display looking behaviors that differentiate actions of agents, depending on whether they have a true or false belief. Violation of expectation paradigms (Onishi & Baillargeon, 2005) as well as anticipatory looking tasks (Senju, Southgate, Snape, Leonard & Csibra, 2011; Southgate, Senju & Csibra, 2007) have served to argue that infants correctly anticipate the actions of agents holding a false belief, and, consequently, have an implicit understanding of others' false beliefs (Baillargeon *et al.*, 2010). This calls into question previous accounts concerning the developmental trajectory of false belief understanding. However, it has been debated whether these findings really reflect infants' access to others' beliefs or can be explained without referring to mental states, for example, by reliance on behavioral cues (Perner & Ruffman, 2005) or lower-level associations (Heyes, 2014; Ruffman, 2014). Further, even assuming that these implicit tasks reflect an access to beliefs, the relation between early implicit and later explicit false belief abilities remains unclear. Whereas some argue for developmental continuity of implicit and explicit false belief abilities (Baillargeon *et al.*, 2010), others have suggested distinct processes subserving the two: Frith and Frith (2008), for instance, suggest implicit and explicit processes, while Apperly and Butterfill (2009) argue in favor of an early efficient, but inflexible and a later more flexible, yet demanding process. A longitudinal study (Thoermer, Sodian, Vuori, Perst & Kristen, 2012) supports developmental continuity by showing that earlier performance on implicit false belief tasks predicted later explicit false belief understanding. Conversely, research on autism (Senju, Southgate, White & Frith, 2009) and a neuroimaging study on implicit and explicit false belief tasks (Schneider, Slaughter, Becker & Dux, 2014) support a dissociation of the abilities measured by the different task types. Research is therefore in need of studies investigating the relation of implicit and explicit false belief tasks in development to shed light on this debate. Studying the differential relation of implicit and explicit false belief tasks to other cognitive domains might help

to inform our understanding of the nature of the relation between the abilities measured by implicit and explicit tasks.

Relation to other cognitive domains

It is still unclear how these abilities relate to the development of other cognitive domains. Language and executive functions in particular have repeatedly been shown to correlate with explicit false belief understanding (Devine & Hughes, 2014; Milligan, Astington & Dack, 2007). There are different theoretical accounts concerning the nature of the relation between false belief tasks and other co-developing abilities, the *expression* and the *emergence account* (see e.g. Carlson, Claxton & Moses, 2015; Devine & Hughes, 2014). The *expression account* explains their correlation with the linguistic or executive control demands needed to express false belief understanding. These demands might result from *superficial task features*. In the case of executive functions, for example, standard explicit false belief tasks have been argued to require inhibition of a more prepotent response that corresponds to one's own knowledge about reality (Robinson & Mitchell, 1995; Baillargeon *et al.*, 2010). Alternatively, the correlation between false belief understanding and other cognitive domains might result from more essential *common conceptual demands* of the underlying cognitive processes; for example, inhibition might be required to handle different perspectives in order to represent others' beliefs. A correlation due to superficial task features predicts that the strength of the correlation with ToM tasks varies depending on the specific demands of the task. According to this account, standard explicit false belief tasks should thus correlate more strongly with language and executive functions than the non-verbal tasks, which have lower executive demands. In contrast, according to the common conceptual demand account, the correlation should be independent of online task demands. Finally, as opposed to the expression account, the *emergence account* assumes that executive functions or language are necessary for the emergence of ToM abilities in development; for example, inhibition might be crucial in order to notice the existence of different perspectives and thus develop an understanding of others' beliefs (Carlson *et al.*, 2015). Similar to the common conceptual demand account, the emergence account predicts that a correlation with language or executive functions does not depend on online task demands.

A large number of studies have shown a robust correlation between executive functions and explicit false belief tasks (Devine & Hughes, 2014; Perner & Lang, 1999). In contrast, the relation with implicit false belief

tasks – which make low or no online executive task demands – remains poorly understood. First studies have found contradicting results in development: Low (2010) reports no correlation of implicit false belief tasks with a dimensional change card-sorting task in preschoolers. Yet, in infants, Yott and Poulin-Dubois (2012) find a correlation using a detour-reaching task. In adults, a dual task study indicated that the abilities measured by an anticipatory looking false belief task recruited executive functions at least to some extent (Schneider, Lam, Bayliss & Dux 2012), but such findings might not apply to a developmental context. In sum, whether executive functions are necessary for the emergence of the abilities measured by implicit false belief tasks remains an open question.

In the language domain, a correlation between standard explicit false belief understanding and linguistic abilities in development is well established (Milligan *et al.*, 2007). In particular, it has been hypothesized that the mastery of complement sentences is related to false belief understanding (de Villiers & Pyers, 2002). In a complement sentence, the object is replaced by a subordinate clause. This allows for a sentence structure in which the whole sentence is true, although the main clause is false. This, in turn, is precisely the linguistic structure needed to express a false belief (e.g. Anaximander believed *that the world is flat*). In the task-related expression account, the correlation of verbal false belief tasks with complement sentences is explained by the linguistic requirements of the tasks; that is, their use of complement sentences and the need to produce these sentences to answer correctly. This account thus predicts that the correlation should vanish for non-verbal false belief tasks. The common conceptual demand account in turn states that both processing complement sentences and representing others' belief require embedding content into a higher-order context. In other words, both necessitate hierarchy processing – also needed more generally for processing syntax (Chomsky, 1956). This account therefore predicts that syntactic abilities in general, not only complement sentences, should be related to false belief understanding. Moreover, syntax should correlate with false belief reasoning more strongly than non-syntactic language abilities (e.g. as argued and shown by Astington & Jenkins, 1999).

Empirically, several studies have indeed found a relation of explicit false belief understanding with memory for complements (e.g. Cheung, Hsuan-Chih, Creed, Ng, Ping Wang *et al.*, 2014; de Villiers & Pyers, 2002; Lohmann & Tomasello, 2003; Low, 2010) as well as with more general syntactic abilities (e.g. Astington & Jenkins, 1999; Milligan *et al.*, 2007). The studies on complement

sentences, however, have not compared understanding complements with other syntactic abilities. Furthermore, the memory for complements task used so far does not allow disentangling the role of syntactic and semantic knowledge of complement sentences. In the present study, we aimed at testing specifically how syntactic knowledge about complementation related to false belief understanding. We therefore employed a repetition task of complement sentences that specifically assessed the mastery of the syntactic structure.

Furthermore, the relation of language to the novel implicit false belief tasks has received little study to date. Because these tasks are non-verbal, the task-related expression account does not predict a correlation with linguistic abilities. Indeed, a study investigating the connection between memory for complements and an anticipatory looking false belief task found no significant correlation between the two (Low, 2010). In contrast, a study by Meristo, Morgan, Geraci, Iozzi, Hjelmquist *et al.* (2012) found that deaf infants of hearing parents performed significantly worse on an anticipatory looking false belief task than hearing children. This indicates that early language input also seems to be important for the abilities underlying early non-verbal implicit false belief tests, pointing to an emergence account of the relation with language. In sum, the link between implicit false belief tests and language remains equivocal, and further research is needed to clarify their relation.

Goal of the current study

In sum, the overarching aim of the study was to obtain a better understanding of the nature and relation of implicit and explicit false belief abilities, by assessing in a single study (1) the correlation between implicit and explicit false belief tasks and (2) their respective relation with co-developing abilities, that is, in particular language and executive functions. We addressed these open issues mentioned above in an integral way, using a comprehensive battery of tests for each of the cognitive domains of interest in order to cover the different aspects of these domains. This was done with a cross-sectional approach and while controlling for general cognitive development.

Hypotheses

- 1 First, we expected to replicate the common findings of significant developmental changes in explicit false belief reasoning, syntactic abilities, and executive functions between the ages of 3 and 4 years. We did not expect age-related changes for the implicit false

- belief task, which we expected to develop earlier (Southgate *et al.*, 2007).
- 2 Based on previous empirical studies, we hypothesized to find a correlation of standard explicit false belief tasks with linguistic, particularly syntactic, abilities and different executive function tasks (Devine & Hughes, 2014; Milligan *et al.*, 2007).
 - 3 The correlation pattern with implicit false belief tasks, however, is less clear from previous literature. Different accounts of the relation of the abilities underlying implicit and explicit false belief tasks make different predictions:

a

In the *continuity account*, both implicit and explicit false belief tasks are considered to measure the same ability of representing others' mental states (e.g. Baillargeon *et al.*, 2010). This account thus predicts a correlation between implicit and explicit false belief tasks. In the case of a continuity of implicit and explicit false belief tasks, a correlation of language and executive functions with explicit false belief tasks, but not with non-verbal implicit false belief tasks, would speak in favor of superficial task features driving the correlation.

b

The lack of a correlation between the two task types, in contrast, would favor a *dual process account*. This account assumes that distinct processes underlie implicit and explicit false belief tasks (e.g. Apperly & Butterfill, 2009). The processes underlying implicit false belief understanding can, but do not need to, be related to mental states.

The correlation pattern between different types of false belief tasks as well as with their co-developing abilities can thus shed light on the relation of the processes underlying the implicit and explicit false belief tasks and might explain the gap in their developmental trajectories.

Methods

Participants

For the study, 60 normally developed monolingual German 3- and 4-year-old children were recruited from local kindergartens and the institute's database. From these, three children had to be excluded from the analyses: Two were excluded because they performed well below average (T -value < 35) in a standardized test of general language abilities (*Sprachentwicklungstest für drei- bis fünfjährige Kinder*, SETK 3–5; Grimm, 2001),

which indicates a speech development disorder. Another child was excluded because of a neurological diagnosis. This left us with 26 3-year-old children (mean age = 39.6 months, range = 36 to 43 months, 13 female) and 31 4-year-old children (mean age = 51.6 months, range = 48 to 54 months, 16 female) for the analysis. Unless stated otherwise, the reported results include this sample. Parental informed consent was obtained for all children before testing. The study was approved by the local ethics committee.

Testing procedure and tasks

The children participated in a battery of implicit and explicit false belief tasks, tests of language abilities, general cognitive abilities, and executive functions on 3 days within an average period of 14.1 days (SD : 6.4 days). The order of the tests was randomized across subjects. However, the implicit false belief task was always performed before the explicit tasks.

Implicit false belief task

An anticipatory looking task served as an implicit test of false belief understanding. In this task, children watched short film clips on the integrated monitor of a Tobii T120 eye tracker (Stockholm, Sweden) while their gaze direction was recorded. The stimuli were presented on a 17-inch monitor using Tobii Studio software. Children were seated in a car seat (Chicco, Neptune) at a distance of approximately 60 cm from the screen. Parents stood behind the children and were instructed not to interact with them.

Every child was presented with a total of 10 familiarization (FAM) trials, 12 false belief (FB) trials (two different conditions of six trials each), and six true belief (TB) trials (two different conditions of three trials each).

The film clips were compiled with Maxon Cinema 4D by the agency Form & Drang (Leipzig). They depicted a scene with a y-shaped tunnel and two boxes; one at each exit of the upper tunnel arms (see Figure 1). In each trial, a mouse entered the scene, followed by another animal (one of six different larger animals), for example a cat. The animal watched the mouse entering the tunnel, exiting it again on one of the upper two tunnel arms and hiding in the box at that exit. Continuation of the course of action depended on the condition.

In the FAM trials, the animal followed the mouse through the tunnel. After 2.5 seconds, it exited on the side where the box with the mouse was and opened it. A light illuminated the two exit areas above the tunnel arms, including the two boxes (see Figure 2), and the

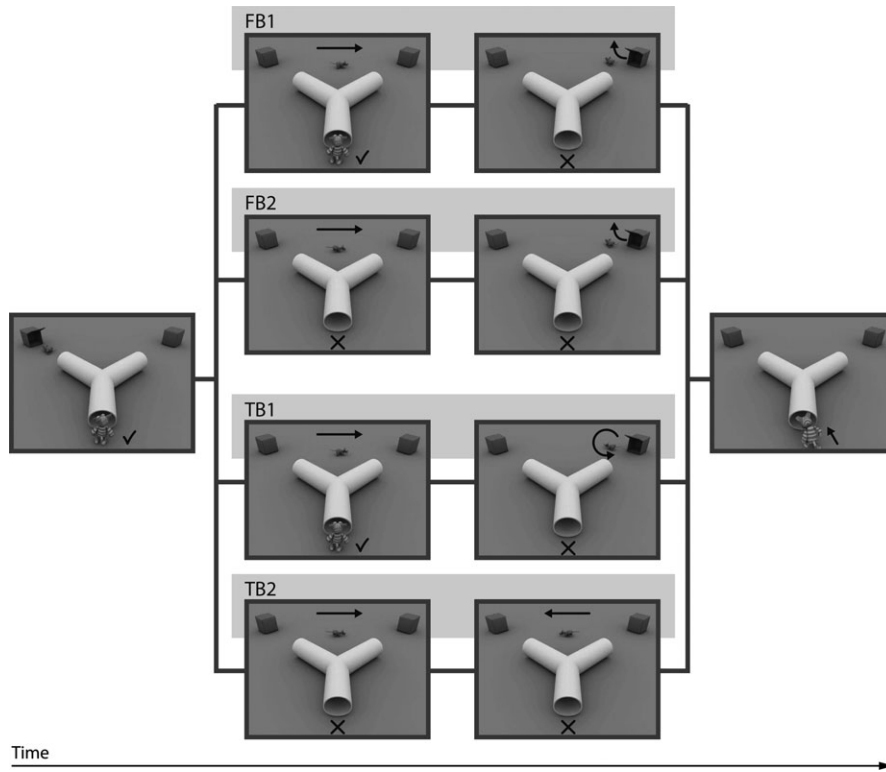


Figure 1 Anticipatory Looking False Belief Task. Selected scenes from the two false belief conditions, FB1 and FB2, as well as the two true belief conditions, TB1 and TB2, as described in the text. Arrows indicate the movement of the animal, check marks or crosses underline whether the cat can see what happens or not. The cat watches how the mouse comes out of the y-shaped tunnel and disappears into one of the boxes. Depending on the condition, the cat then leaves the scene and, while she is gone, the mouse changes her location and, in the false belief conditions, leaves the scene. When the cat returns, it has either a true or a false belief about the location of the mouse. The cat then enters the tunnel and the children’s eye-gaze is tracked.

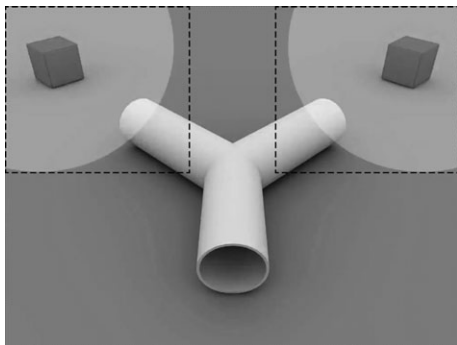


Figure 2 Region of Interest (ROI) in the Anticipatory Looking False Belief Task. The picture shows the scene with the attention lights displayed in the anticipation phase as described in the text. The dotted lines depict the ROIs in which the children’s eye-gaze was analyzed.

animal’s cry (e.g. a meow) was played, 500 milliseconds after the animal had entered the tunnel. This was intended to help the children familiarize with the animal

exiting the tunnel after the effect. In addition, the effect helped to capture the attention of the children and direct it to the areas of interest in the critical phase in which gaze direction was analyzed.

The course of action in the false belief trials followed the same structure as the anticipatory looking false belief task developed by Southgate *et al.* (2007): In condition FB1, the animal watched the mouse crossing from one box to the other. The animal then left the scene. While the animal was away, the mouse left the box and the scene. The animal came back, entered the tunnel and, as in the FAM trials, the attention light and sound occurred. However, the false belief trials then ended after 2.9 seconds without the animal exiting the tunnel. This was supposed to prevent children from learning across the false belief trials. Condition FB2 was identical, except for the fact that the animal left the scene before the mouse crossed over to the other box. Both conditions left the animal with a false belief about the location of the mouse: in FB1 the last box in which the mouse had been before leaving the scene, and in FB2 the first box to

which the mouse had gone. Using these pieces of information as clues, instead of the animal's belief, therefore led to chance-level performance. In both false belief conditions the mouse was in neither of the two boxes in the critical phase in which the gaze direction was analyzed. This prevented misinterpreting gaze directed to the mouse's location instead of to the place where the larger animal was expected to exit. Furthermore, because of the absence of the mouse, the children did not have to overcome a reality bias (Robinson & Mitchell, 1995).

The two true belief (TB) conditions were analogous to the false belief conditions. However, in TB1, instead of leaving the scene while the animal was gone, as it had done in FB1, the mouse walked in the direction of the exit but then returned to the box. Condition TB2 was analogous to FB2, but instead of leaving the scene at the end, the mouse walked back to the first box. In both TB conditions, when the animal came back, the mouse was thus in the box where the animal had seen it last. The animal, therefore, had a correct belief about the location of the mouse when she entered the tunnel. As in the FAM trials, 2 seconds after the light and sound attention effect, the animal exited the tunnel on the side where the box with the mouse was.

The order of the trials was randomized. This was done to prevent the children from being able to predict whether the larger animal would actually exit the tunnel, or whether the trial ended when the animal was still in the tunnel, as was the case for the false belief trials. However, before the first false belief trial with each of the animals, the children saw two FAM or TB trials with the same animal – one in which it exited on the right and one on the left side. The intention of these trials was to familiarize the children that the animal would go to where the mouse was. The correct side was balanced for each animal, within every condition, as well as across all trials.

Explicit false belief tasks

The children participated in two standard explicit false belief tasks: a false location (Wimmer & Perner, 1983) and a false content task (Hogrefe, Wimmer & Perner, 1986).

False Location task. In the False Location task, a hand puppet mouse was introduced to the child. The child and the mouse were shown a little bag with a gummy bear and a small, empty box. The mouse then went to sleep behind a room divider. While the mouse was away, the experimenter moved the gummy bear from the bag to the small box saying 'Ssh! The mouse can't see this, but

don't tell!' The mouse was then called back and the child was told that it liked gummy bears and was asked three probe questions: 1. *Where will it look?* 2. *Does it know where the gummy bear is?* 3. *Where does it think the gummy bear is?* Finally, the children were asked a control question about the actual location of the gummy bear.

False Content task. In the False Content task, the mouse went to sleep directly after having been introduced to the children. The children were then shown a closed *Kinder* chocolate bars box and were asked what they thought was in the box. All the children expected the box to contain chocolate. The experimenter then showed them that the box contained pencils instead of chocolate bars, while saying 'I'll show you something the mouse can't see, but ssh!' The mouse was then called back, and the children were asked three probe questions: 1. *Does the mouse know what is in the box?* 2. *What does it think is in the box?* 3. *And what did you think was in the box at first?* Finally, the children were again asked a control question about the actual content of the box.

In addition to the standard explicit false belief tasks, the children performed two additional non-standard elicited-response false belief tasks with low verbal and executive demands (Southgate, Chevallier & Csibra, 2010) reported in the Supporting Information (SI 'A1. Indirectly Elicited Response *Sefo-Task*').

Language

To test for language abilities, children performed a standardized general language test as well as a specific syntax of complement sentences test.

Test of general language abilities. The standardized test of general language abilities SETK 3–5 (Grimm, 2001) included a specific syntax subtest containing complex syntax, such as subject and object relatives and prepositional phrases, but no complement sentences. In addition, it included tests for encoding semantic relations, phonological working memory for words and non-words, as well as morphological rule building.

Repeating complement sentences. Children have been shown to repeat correct sentences more accurately than incorrect sentences and, moreover, to correct incorrect sentences (Kidd, Lieven & Tomasello, 2006; Weisenborn, Höhle, Kiefer & Cavar, 1998). This is taken to reflect their knowledge about the correct syntactic structure of these sentences (Ambridge & Lieven, 2011; Kidd *et al.*, 2006; Crain & Thornton, 2000). To test children's syntactic knowledge of complement sentences, we therefore employed a repetition task of complement

sentences with correct and incorrect word order, adapted from Weissenborn *et al.* (1998). Given that we were specifically interested in the children’s syntactic and not in semantic aspects of their mastery of complement sentences, we did not additionally test the children’s comprehension of the sentences. In the repetition task employed, children were asked to repeat complement clauses with correct or incorrect word order either with or without the German complementizer *dass* (see Table 1). In German, the use of a complementizer is optional, but verb and object take different word order depending on the presence or absence of the complementizer. Both word orders could, therefore, be correct – depending on the sentence structure. Hence, to produce complement sentences using the correct word order, children need to have knowledge of the correct sentence structure of complementation. The accuracy in repeating correct versus incorrect sentences and the tendency to correct incorrect sentences were thus taken as an indicator for mastering the correct syntax of complement sentences. Taking into account that children, who in general speak less accurately, would also repeat incorrect sentences less accurately, we considered the quotient over non-literal repetitions of all sentences. A female speaker recorded the sentences in a child-directed manner. Only correct sentences were recorded and the sentences were then cross-spliced to produce the presented stimuli. The children were asked to repeat the sentences that they were told by a duck hand puppet, in order to help a deaf mouse hand puppet. After every sentence, the mouse asked them ‘What did Anna say?’ The 3-year-old children were asked to repeat 16 sentences (i.e. four different sentences, for each of the four conditions, respectively). The 4-year-olds were asked to repeat 24 sentences (i.e. six sentences per condition). The sentences were presented in pseudo-randomized order, taking care that the order of the conditions was balanced across the sentences. The children’s answers were coded into the categories ‘complete’ or ‘incomplete’, ‘literal’ or ‘non-literal’ repetitions. For the incorrect stimuli, a category

‘corrections’ was formed for non-literal but syntactically correct repetitions. For an answer to be considered ‘incomplete literal’, it had to contain at least the object and the verb of the complement clause, as well as for the conditions with complementizer, the *dass*. The answers were coded by two independent raters who agreed on the categories in 98.1% of the cases.

Executive functions

To test the children’s executive functions, we adapted three tasks from the literature: a Reverse Categorization task (Carlson, 2005), a Go-NoGo task (Rakoczy, 2010), and the standard Delay of Gratification task (Mischel & Ebbesen, 1970). These tasks were chosen with the intention of testing cognitive flexibility (the former task) and inhibition (the latter two tasks). These were executive functions we considered to be of possible importance for false belief understanding (Devine & Hughes, 2014) and for our tasks.

Reverse Categorization task. In the Reverse Categorization task, the children were asked to sort blue cubes into a big blue box and red cubes into a small red box. Half of the cubes from each color were small and the others were big. After they had finished sorting the 20 cubes (10 cubes of each color), the rule changed, and they were asked to sort the blue cubes into the red box and the red cubes into the blue box. Then, the rule changed again and they had to sort the small cubes into the small box and the big cubes into the big box. In the final round, this rule was reversed again. After every rule change, the new rule was explained and demonstrated to the children with two cubes, and they were given feedback on the first two trials. Furthermore, after half of the trials, they were reminded of the rule. The children were given the cubes in a pseudo-randomized order. The mean performance in every round following a rule change (i.e. the last three rounds) was encoded as dependent variable.

Table 1 Stimuli the children were asked to repeat in the repeating complement sentences task: Complement clauses with correct or incorrect word order of a transitive verb and its object, with or without the German complementizer *dass*. Literal translation in squared brackets. In German, verb and object take a different order depending on the presence or absence of a complementizer. Because we used sentences with and without complementizer, both word orders could therefore be correct, depending on the sentence structure

	Correct	Incorrect
With complementizer <i>dass</i>	Anna sagt, dass der Opa Blumen _{object} krieg _{t verb} . [Anna says, that the grandpa flowers gets.] _{literal}	*Anna sagt, dass der Opa krieg _{t verb} Blumen _{object} . [Anna says, that the grandpa gets flowers.] _{literal}
Without complementizer	Anna sagt, der Opa krieg _{t verb} Blumen _{object} . [Anna says, the grandpa gets flowers.] _{literal}	*Anna sagt, der Opa Blumen _{object} krieg _{t verb} . [Anna says, the grandpa flowers gets.] _{literal}

Go-NoGo task. In the Go-NoGo task, children sat in front of a puppet theater and were told to do what a nice duck asked them to (e.g. 'Clap your hands!'), but not to do anything when the nasty crocodile asked them to make the movements. Before starting, we checked that children understood all the prompts and were able to perform the movements. They then received at least three practice trials, with the duck and with the crocodile, respectively, until they performed correctly in two consecutive trials and were corrected otherwise. If they failed to perform correctly in the crocodile condition after four trials, the experimenter held the children's hands for the fifth and sixth trial and gave them positive feedback. This was the case for three children. There were a total of 24 trials (12 duck, 12 crocodile trials) with six different hand movements (distinct from the three movements in the practice trials), presented in pseudo-randomized order.

Delay of Gratification task. In the Delay of Gratification task, children were first allowed to choose what they liked most: chocolate bars or gummy bears. Children were then taught that ringing a bell that was placed on a table in front of them would immediately bring the experimenter back after he had left the room by practicing this procedure three times. They were then asked whether they preferred a plate with a small piece of chocolate (one gummy bear) or with a whole chocolate bar (seven gummy bears). All the children, except for two, chose the bigger portion, and the other two did so after a second check. The plate with the small portion of their preferred sweets was then placed in front of them and the plate with the bigger portion was placed in a locked glass box next to it. The experimenter told the children she had to leave for a while. She explained that the children could have the small portion of sweets immediately, but if they waited until the experimenter came back without being called, they would get the big portion. If they did not want to wait any longer, they could ring the bell to summon the experimenter back, but then they would only receive the small portion. Task comprehension was checked with two control questions before the children were left alone for a maximum of 5 minutes. The waiting time was taken as a measure of inhibitory control.

General cognitive functions

As a test of general cognitive abilities, three subtests from the subscale *intellectual abilities* of the Kaufmann Assessment Battery for Children (K-ABC, German version; Melchers & Preuß, 2003) were performed: In the subtest *Magic Window*, children were asked to

identify pictures they could only see partially through a small slit on a turning disk. In the second subtest, children were asked to repeat sequences of hand movements, and the last subtest was a forward digit span test. The first two subtests tested visual working memory and selective attention as well as spatial representation abilities; assets that we considered to be of importance for our false belief tasks. The digit span accounted for acoustic working memory.

Results

The results section of this paper consists of two parts: an overview of the individual tasks, followed by the inter-correlations between the tasks. Preceding analyses revealed no effects of the order in which the tasks were performed and no gender effects if not stated otherwise.

Results of the individual tasks

In the following, unless stated otherwise, we report the mean percent of correct trials, two-tailed *p*-values. One-sample *t*-tests were performed to test performance against chance, and independent samples *t*-tests to test for age group effects.

Implicit false belief task

In the anticipatory looking task, gaze data were analyzed for a time of interest from the moment the larger animal had disappeared into the tunnel until its reappearance in the familiarization (FAM) and true belief (TB) trials (2.5 sec), or until the end of the trial in the false belief trials (2.9 sec). Two regions of interest (ROI), each covering one of the tunnel exits and the corresponding box, were defined (as depicted by the dotted lines in Figure 2). During the time of interest, the ROI that the child looked at first ('first look') as well as the ROI with the longer gaze duration ('longer look') was recorded. Trials in which children looked at neither of the two ROIs were excluded from the analysis, and the average percentage of correct trials from all valid trials is reported here. Since both measures yielded similar results (percent correct trials: first look: $M = 62.4\%$, $SD = 8.7\%$; longest look: $M = 61.9\%$, $SD = 8.9\%$, paired samples *t*-test: $t(55) = 0.45$, $p = .65$), the measures were collapsed for further analyses by taking the mean of the first and longest look in every child. The average performance in this mean value is reported in the following and is shown in Figure 3a. This measure will be used in all subsequent analyses, as well as in the correlational analyses. The reported results and patterns of correlation also hold for

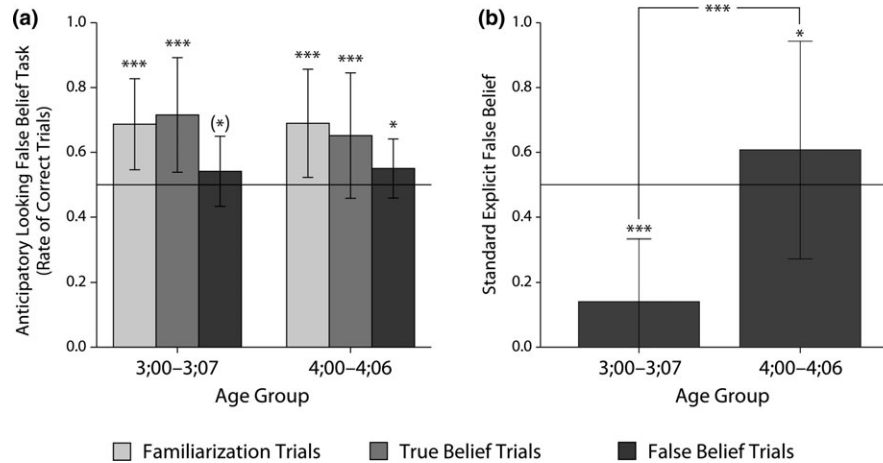


Figure 3 Results of the Implicit and Explicit False Belief Tasks. While both age groups were above chance in the implicit anticipatory looking false belief task, 3-year-olds were significantly below chance on the explicit false belief tasks. Figure (a) shows the mean rate of correct anticipations in the anticipatory looking false belief task (mean of correct first and correct longest look). Both age groups were significantly above chance in the false belief condition (one-tailed for the 3-year-olds) and performed even better in the familiarization (FAM) and TB conditions, which served as a control. Figure (b) shows the mean rate of correct answers in the two standard explicit false belief tasks (as described in the text).

the average correct first look and for the normalized difference of looking durations in the correct and the incorrect ROI (see SI 'B2. Correlation Pattern of First Look and Difference in Looking Duration'). For the mean of the correct first and longest looks, both age groups performed significantly above chance in the FAM condition (3-yos: $M = 69\%$, $SD = 14\%$, $t(25) = 7.16$, $p < .001$; 4-yos: $M = 69\%$, $SD = 17\%$, $t(30) = 6.27$, $p < .001$) and in the TB condition (3-yos: $M = 71\%$, $SD = 18\%$, $t(25) = 6.03$, $p < .001$; 4-yos: $M = 64\%$, $SD = 20\%$, $t(30) = 3.97$, $p < .001$). This result indicated that the children understood the events displayed in the film clips and showed correct anticipation when no false belief was involved. There was no significant difference between the age groups or between the two conditions. For the false belief trials, as expected, both age groups were also significantly above chance (3-yos: $M = 54\%$, $SD = 11\%$, $t(25) = 1.81$, $p = .04$ (one-tailed); 4-yos: $M = 54\%$, $SD = 11\%$, $t(30) = 2.12$, $p = .04$). Again, there was no significant age difference. However, as expected, the children performed significantly better in the FAM and TB conditions than in the false belief condition (paired samples t -test: 3-yos: $t(25) = 4.86$, $p < .001$; 4-yos: $t(30) = 3.66$, $p = .001$).

Standard explicit false belief tasks

In line with our hypotheses, the 3-year-olds were significantly below chance in both standard explicit false belief tasks (False Location: $M = 10\%$, $SD = 21\%$,

$t(25) = -9.8$, $p < .001$; False Content: $M = 18\%$, $SD = 25\%$, $t(25) = -6.4$, $p < .001$), whereas the 4-year-olds were significantly above chance in the False Location task ($M = 70\%$, $SD = 37\%$, $t(30) = 3.0$, $p = .005$) and at chance level in the False Content task ($M = 52\%$, $SD = 34\%$, $t(30) = 0.26$, $p = .80$). For both tasks, there was a significant difference between the age groups (False Location: $t(55) = -7.7$, $p < .001$; False Content: $t(55) = -4.3$, $p < .001$; see Figure 3b).

There was a strong correlation between the two standard explicit false belief tasks ($r(57) = .760$, $p < .001$).

Language

Repeating complement sentences. The children's performance for the repeating complement sentences task is shown in Figure 4. Since we considered corrections of incorrect sentences to be an even more sensitive measure for the mastery of the correct syntax of complementation than non-literal repetitions, we report the results for the quotient of corrections over total non-literal repetitions in the further analysis. However, the reported results also hold for an analyses with the other quotient (see Figure 4). Both age groups performed significantly above chance (one-sample t -test against test value 0.5: 3-yos: $M = 1.03$, $SD = 0.49$, $t(25) = 5.55$, $p < .001$; 4-yos: $M = 1.49$, $SD = 0.39$, $t(30) = 14.2$, $p < .001$) and there was a significant increase of performance with age ($t(55) = 3.9$, $p < .001$).

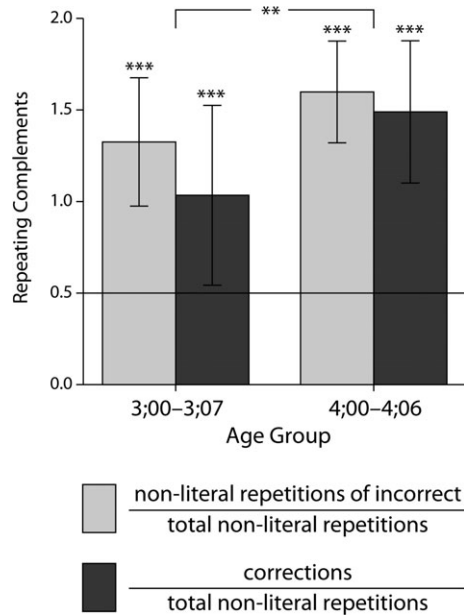


Figure 4 Results of the Repeating Complement Sentences Task. Children were asked to repeat complement sentences with correct and incorrect word order. Both age groups repeated correct sentences significantly more accurately than incorrect sentences, indicating their knowledge about the syntax of complement sentences. This can be seen from the performance on the first quotient: rate of non-literal repetitions of incorrect sentences over the rate of total non-literal repetitions. Moreover, they even corrected the incorrect sentences, as shown by the second quotient: rate of corrections of incorrect sentences over rate of total non-literal repetitions. There was a significant age difference on both quotients.

Test of general language abilities. In the standardized language test SETK 3-5, both age groups had comparable mean *T*-values (3-yos: $M = 56.7$, $SD = 6.5$; 4-yos: $M = 58.3$, $SD = 6.9$; age effect: $t(55) = 0.91$, $p = .37$).

Executive functions

Reverse Categorization task. In the Reverse Categorization task, the mean percent of correct trials on the three rounds after each of the rule changes was taken as a measure of executive function (3-yos: $M = 83\%$, $SD = 15\%$; 4-yos: $M = 93.2\%$, $SD = 7.5\%$; age effect: $t(54) = 3.1$, $p = .004$). One child had to be excluded from the test because he refused to participate in the last two rounds. All the participating children performed above 90% on the first round before any rule change had taken place.

Go-NoGo task. For the Go-NoGo task, a *d*-prime was calculated, in which the correct NoGo-trials were

interpreted as hits and the incorrect Go-trials as false alarms (3-yos: $d' = 0.74$, $SD = 0.31$; 4-yos: $d' = 0.94$, $SD = 0.10$; age effect: $t(54) = 3.1$, $p = .005$). One child had to be excluded because he refused to participate.

Delay of Gratification task. In the Delay of Gratification task, the time until the end of the trial was taken as measure of inhibitory control. The test was stopped either when the children rang the bell ($N = 15$), when they ate the small portion of sweets ($N = 1$), called the experimenter ($N = 4$), or when they left the room on their own ($N = 4$). The mean waiting time was $M = 190$ sec ($SD = 120$ sec) for the 3-year-olds ($N = 25$) and $M = 244$ sec ($SD = 96$ sec) for the 4-year-olds (age effect: $t(54) = 1.9$, $p = .03$ (one-tailed)). One child had to be excluded because she had to go to the bathroom during the trial. There was a main effect of gender on this task (boys: $M = 190$ sec, $SD = 120$ sec, girls: $M = 253$ sec, $SD = 91$ sec, independent samples *t*-test: $t(54) = 2.4$, $p = .02$).

Total executive function score. The executive function tasks correlated with each other (Reverse Categorization and Go-NoGo: $r(55) = .396$, $p = .003$; Reverse Categorization and Delay of Gratification: $r(55) = .406$, $p = .002$), except for the Go-NoGo and the Delay of Gratification task, which only showed a trend ($r(55) = .206$, $p = .065$ (one-tailed)). For the further analyses, we therefore aggregated the *z*-scores of the three tasks to a total executive function *z*-score (3-yos: $M = -0.6$, $SD = 1.1$; 4-yos: $M = 0.46$, $SD = 0.55$; age effect: $t(55) = 4.1$, $p < .001$). All the children who participated in at least two out of the three tests were included in the aggregated score, which left us with the full sample again.

General cognitive functions

In the selected subtests of the K-ABC, our sample was comparable to the norm sample (scale values: 3-yos: $M = 10.1$, $SD = 1.4$; 4-yos: $M = 10.4$, $SD = 1.4$) with no significant difference between the age groups ($t(55) = -0.69$, $p = .49$).

Correlations between the tasks

The intercorrelations between the false belief tasks are shown in Table 2. The two standard explicit false belief tasks were strongly correlated with each other; however, they did not correlate with the implicit anticipatory looking false belief task.

The correlations of explicit and implicit false belief tasks with the other cognitive domains are shown in

Table 2 Correlations between explicit false belief (FB) tasks (1. & 2.) and the implicit anticipatory looking false belief task (3.)

FB tasks	1.	2.
1. False Location	–	
2. False Content	.760***	–
3. Implicit FB	–.108	.044

N = 57. **p* < .05; ****p* < .001.

Table 3 Correlations of explicit false belief tasks (1. & 2.) and implicit false belief (FB) task (3.) with the aggregated executive function score (EF), the quotient of correcting complement clauses (COMPs), the syntax part of the standardized test of general language abilities (SETK-Syn), the part of the SETK testing non-syntactic abilities (SETK-noS), and the test of general cognitive abilities (K-ABC)

FB tasks	EF	COMPs	SETK-Syn	SETK-noS	K-ABC
1. False Location	.496***	.434**	.233(*)	.040	.146
2. False Content	.320*	.334*	.365**	.174	–.017
3. Implicit FB	.057	–.187	.239	.046	–.142

N = 57. (*) < .10; **p* < .05; ***p* < .01; ****p* < .001.

Table 3. We found that the standard explicit false belief tasks correlated with syntactic abilities and executive functions, whereas the implicit task did not. To compare these correlations, we aggregated the two highly inter-correlated standard explicit false belief tasks by forming the mean of their respective *z*-scores. The correlation of the complement sentences test with the standard explicit false belief score obtained differed significantly from the correlation with the implicit false belief task (Williams’ *t*-test: $t(54) = -3.39$, $p = .001$) (Williams, 1959). The same pattern was present for executive functions: The correlation with the standard explicit false belief score differed significantly from the correlation with the implicit task (Williams’ *t*-test: $t(54) = -2.41$, $p = .019$). The correlations were similar for each of the three executive function tasks, as shown in Table 4.

Table 3 also shows that the correlation of explicit false belief with repeating complements was not found to be significantly stronger than with other syntactic abilities tested in the SETK (Williams’ *t*-test: $t(54) = 0.57$, $p = .57$). We therefore aggregated the two tasks in order to form a total syntax score (mean of the *z*-scores) and found that the standard explicit false belief score correlated with the total syntax score ($r(57) = .477$,

Table 4 Correlations between different executive function tasks and explicit and implicit false belief (FB) tasks: While standard explicit false belief tasks showed a robust correlation with all three different executive function tasks, neither of the two implicit false belief tasks showed a correlation with any of the executive function tasks

Tasks	Rev. Cat.	Go-NoGo	Delay of Grat.
Explicit FB	.333*	.338*	.305*
Implicit FB	.014	.117	–.031

N = 57. **p* < .05.

$p < .001$), but not with the non-syntactic abilities tested in the SETK ($r(57) = .114$, $p = .40$). Again, these correlations differed significantly (Williams’ *t*-test: $t(54) = 2.73$, $p = .009$).

Discussion

The present study aimed to clarify the nature and relation of the processes underlying the implicit and explicit false belief tasks as well as their relation to other cognitive domains in development. For explicit false belief understanding, we found a critical developmental change between our two age groups, in which 3-year-olds performed significantly below chance and 4-year-olds significantly above chance – in line with previous literature (Wellman *et al.*, 2001). This development was paralleled by significant age differences on syntactic abilities and executive functions between the ages of 3 and 4 years. In contrast, for the implicit anticipatory looking false belief task, both age groups performed above chance and there was no significant age difference. Moreover, we found no correlation between the explicit and the implicit anticipatory looking false belief tasks. Finally, studying the relation of the different false belief tasks with other cognitive domains revealed the following pattern: While standard explicit false belief tasks correlated with syntactic abilities and executive functions, the implicit anticipatory looking false belief task did not.

Age effects

While we observed a critical age development for explicit false belief understanding between the ages of 3 and 4 years, performance on the implicit false belief task was equally above chance in both age groups. This is in line with previous studies, which showed that correct anticipation of actions of an agent with a false belief has already developed by the age of 2 years (Senju *et al.*,

2011; Southgate *et al.*, 2007). Moreover, similar anticipatory looking false belief tasks with adults have shown that even adults do not perform at ceiling on these tasks (Schneider *et al.*, 2014; Senju *et al.*, 2009). In Senju *et al.* (2009), for example, neurotypical adults showed similar performance levels (76% correct first looks) to infants for the same task (Southgate *et al.*, 2007). Together with these findings, our results suggest that the abilities measured by such implicit false belief tasks might already be in place in infancy and might not improve considerably thereafter. Indeed, even in adulthood performance that is far from ceiling on these tasks might be due to the spontaneous and automatic nature of the responses.

Correlations of implicit and explicit false belief tasks

As expected, the standard explicit false belief tasks correlated strongly with each other. However, the explicit tasks did not correlate with the implicit anticipatory looking false belief task. Our data thus suggest that distinct cognitive processes underlie explicit false belief reasoning and earlier-developing spontaneous anticipation of the actions of an agent with a false belief. The results are therefore compatible with a dual process view of implicit and explicit ToM. This account suggests an automatic, cognitively efficient possibly unconscious belief-tracking system already present in infancy, and an explicit more flexible but cognitively more demanding belief processing system, which develops later (Apperly & Butterfill, 2009; Frith & Frith, 2008; Schneider *et al.*, 2014). Alternatively, the abilities underlying implicit spontaneous looking behavior might also be lower-level processes unrelated to mental-state processing, as argued for example by Heyes (2014) and Ruffman (2014).

Relation with other cognitive domains

The relation of the different false belief tasks with other cognitive domains also informs the relation between implicit and explicit false belief abilities. While standard explicit false belief tasks correlated with syntactic abilities and executive functions, the implicit false belief tasks did not, with a significant difference between the correlations. The lack of a correlation between implicit and explicit false belief tasks pointed to distinct processes underlying these tasks. This correlation pattern therefore seems to be neutral with respect to whether syntactic knowledge and executive functions are required for the *expression* or for the *emergence* of explicit false belief understanding. However, the results

do not give any support for the need of linguistic abilities and executive functions for passing implicit false belief tasks.

False belief understanding and language

A closer look at the relation of explicit false belief understanding with language revealed that the standard false belief tasks correlated with syntactic abilities, but not with the non-syntactic measures we acquired – that is, semantics and phonological working memory. These correlations differed significantly from one another. However, the mastery of complement sentences correlated with explicit false belief understanding no more markedly than understanding other complex hierarchical syntactic structures. These results are compatible with the view that the correlation of explicit false belief reasoning and syntactic abilities is driven by common conceptual demands on processing complex embedded structures (Frye, Zelazo & Palfai, 1995). In this view, the meta-representation of others' beliefs as well as mastering syntactic hierarchy both require hierarchy processing, driving a correlation between the two. The results are also compatible with an account in which syntactic abilities (but not other language abilities) are important for the emergence of false belief understanding (Astington & Jenkins, 1999; de Villiers & de Villiers, 2014), a view also supported by studies on language-delayed deaf children who show delayed false belief understanding on low-verbal tasks (Schick, de Villiers, de Villiers & Hoffmeister, 2007). Non- or low-verbal false belief tasks that show a correlation with the standard explicit false belief tasks might help to clarify whether the correlation of syntax and ToM is also driven by superficial task demands (see SI A1 and e.g. Newton & de Villiers, 2007).

False belief understanding and executive functions

Similar to the correlation patterns with syntax, executive functions also correlated with the standard explicit false belief tasks, but not with the implicit anticipatory looking task. This pattern of correlation held for all three acquired tasks of executive functions, measuring inhibition, re-description of task stimuli, and the use of conflicting and conditional rules respectively (Table 4). This suggests that the correlation between executive functions and explicit false belief tasks is not merely driven by specific common processes, such as reasoning about embedded conditional rules (Frye *et al.*, 1995), or by specific task requirements, such as response inhibition. Instead, it suggests that executive functions are

needed more generally for the emergence or expression of false belief understanding.

The absence of a correlation of any of the executive function tasks with the implicit anticipatory looking false belief task indicates that executive functions might not be required for spontaneous anticipation of the actions of an agent with a false belief. This is in line with a study by Low (2010) employing a dimensional change card-sorting task. The results, however, contrast with a study by Yott and Poulin-Dubois (2012), which found that infants' performance on an implicit violation of expectation false belief task correlated with a detour reaching task taken as a measure of inhibition. Furthermore, they contrast with the reported interference of a 2-back task and an anticipatory looking false belief task in a dual task study with adults (Schneider *et al.*, 2012). Interestingly, in these studies the object that the agent falsely believed to be in a different location remained present in the scene in the test phase in which the looking response was measured. This might have introduced the need of inhibitory control to suppress a looking response to the actual object location (as argued for explicit tasks in Robinson & Mitchell, 1995). In our task, in contrast, the object was entirely removed from the scene before the response phase started, thus possibly reducing executive function demands. This suggestion needs to be tested in future studies which examine the correlation of different executive function tasks (including inhibition and other executive function measures) with implicit looking-behavior false belief tasks that vary the presence or absence of the object in the scene.

Performance in the implicit false belief tasks

Although significantly above chance in both age groups, the performance in the anticipatory looking false belief task with on average 54% correct anticipatory looks was somewhat lower than in previous studies of anticipatory looking, where infants under the age of 2 years showed between 77% and 85% of correct looks (Senju *et al.*, 2011; Southgate *et al.*, 2007). Importantly, however, the percentage of correct anticipatory looking in the familiarization and true belief control trials in our study was also lower than in previous studies, although very clearly significantly above chance robustly across all trials (see SI 'B1. Time Course of the Performance across Trials'). This indicates that our task reliably measured anticipatory looking and that the lower absolute performance in anticipatory looking in all conditions was probably because of the visual setup of our task. In our task, the scene was more center-oriented than in the task by Southgate *et al.* (2007), and gaze direction could

frequently be observed along the arms of the tunnel and not only in the ROIs that covered the tunnel exits and boxes. Moreover, the top-view perspective of the scene was more complex than in previous tasks. This might have led to greater variance and a lower percentage of correct anticipation. Another possible reason for the differences in performance is that in previous anticipatory looking false belief studies a relatively high number of participants had to be excluded due to fussiness or similar reasons (around 45% in Southgate *et al.*, 2007, and 25% in Senju *et al.*, 2011). This might have led to a selection bias for the more attentive and thus mature children, possibly with better ToM abilities. Such a potential bias was reduced in the present study: because of the larger number of trials, only single trials had to be excluded, but all children were included in the analysis.

Conclusions

Using a comprehensive task battery of implicit and explicit false belief tasks, the present study finds critical developmental changes on explicit false belief tasks between the ages of 3 and 4 years, but not on the implicit anticipatory looking task. This developmental breakthrough is paralleled and seems to be fostered by important improvements in executive functions and syntactic, but not other linguistic abilities. In contrast to the standard explicit false belief tasks, the implicit false belief task does not depend on executive functions or language. Moreover, later-developing explicit false belief understanding appears to be independent of earlier implicit action anticipation of an agent with a false belief. These results support the view that distinct cognitive processes underlie implicit and explicit false belief tasks.

Acknowledgements

The research was supported by a grant from the German National Academic Foundation to CGW and a grant from the European Research Council (ERC-2010-360 AdG 20100407) awarded to ADF. NS was kindly supported by a Jacobs Foundation Early Career Research Grant. We would like to thank the children and parents who volunteered to participate in the study, as well as Christiane Attig, Christoph von Borell, Elisabeth Mueche, and Hung Nguyen Trong, who contributed to preparing the stimuli and acquiring and coding the data.

References

- Ambridge, B., & Lieven, E.V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.
- Apperly, I.A., & Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116** (4), 953–970.
- Astington, J.W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology*, **9** (1), 7–31.
- Astington, J.W., & Jenkins, J.M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, **35** (5), 1311–1320.
- Baillargeon, R., Scott, R.M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, **14** (3), 110–118.
- Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, **21** (1), 37–46.
- Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, **1**, 557–560.
- Bloom, P., & German, T.P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, **77** (1), B25–B31.
- Carlson, S.M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, **28** (2), 595–616.
- Carlson, S.M., Claxton, L.J., & Moses, L.J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development*, **16** (1), 186–197.
- Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Ping Wang, S., et al. (2014). Relative roles of general and complementation language in theory-of-mind development: evidence from Cantonese and English. *Child Development*, **75** (4), 1155–1170.
- Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, **2**, 113–124.
- Crain, S., & Thornton, R. (2000). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- Dennett, D.C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, **1**, 568–570.
- de Villiers, J.G., & de Villiers, P.A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, **34** (4), 313–328.
- de Villiers, J.G., & Pyers, J.E. (2002). Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, **17**, 1037–1060.
- Devine, R.T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Development*, **85** (5), 1777–1794.
- Flavell, J.H., Green, F.L., & Flavell, E.R. (1990). Developmental changes in young children's knowledge about the mind. *Cognitive Development*, **5** (1), 1–27.
- Frith, C.D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, **60** (3), 503–510.
- Frye, D., Zelazo, P.D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, **10** (4), 483–527.
- Grimm, H. (2001). *Sprachentwicklungstest für drei-bis fünfjährige Kinder: SETK 3–5*. Diagnose von Sprachverarbeitungs-fähigkeiten und auditiven Gedächtnisleistungen: Hogrefe, Verlag für Psychologie.
- Heyes, C.M. (2014). False belief in infancy: a fresh look. *Developmental Science*, **17** (5), 647–659.
- Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: a developmental lag in attribution of epistemic states. *Child Development*, **57** (3), 567–582.
- Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, **21** (2), 93–107.
- 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.
- Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, **81** (2), 597–615.
- Melchers, P., & Preuß, U. (2003). *K-ABC Kaufman Assessment Battery for Children, German version, Individualtest zur Messung von Intelligenz und Fertigkeit bei Kindern (6. teilweise ergänzte Auflage)*. Göttingen: Hogrefe.
- Meristo, M., Morgan, G., Geraci, A., Iozzi, L., Hjelmquist, E., et al. (2012). Belief attribution in deaf and hearing infants. *Developmental Science*, **15** (5), 633–640.
- Milligan, K., Astington, J.W., & Dack, L.A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, **78** (2), 622–646.
- Mischel, W., & Ebbesen, E.B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, **16** (2), 329–337.
- Newton, A.M., & de Villiers, J.G. (2007). Thinking while talking: adults fail nonverbal false-belief reasoning. *Psychological Science*, **18**, 574–579.
- Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, **308** (5719), 255–258.
- Perner, J. (1991). *Understanding the representational mind: Learning, development, and conceptual change*. Cambridge, MA: MIT Press.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, **3** (9), 337–344.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science*, **308** (5719), 214–216.
- Rakoczy, H. (2010). Executive function and the development of belief-desire psychology. *Developmental Science*, **13** (4), 648–661.
- Robinson, E.J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: backwards explanation versus prediction. *Child Development*, **66** (4), 1022–1039.

- Ruffman, T. (2014). To belief or not belief: children's theory of mind. *Developmental Review*, **34**, 265–293.
- Schick, B., De Villiers, P., De Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: a study of deaf children. *Child Development*, **78** (2), 376–396.
- Schneider, D., Lam, R., Bayliss, A.P., & Dux, P.E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, **23** (8), 842–847.
- Schneider, D., Slaughter, V.P., Becker, S.I., & Dux, P.E. (2014). Implicit false-belief processing in the human brain. *NeuroImage*, **101**, 268–275.
- Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, **22** (7), 878–880.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, **325** (5942), 883–885.
- Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, **13** (6), 907–912.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, **18** (7), 587–592.
- Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, **30** (1), 172–187.
- Weissenborn, J., Höhle, B., Kiefer, D., & Cavar, D. (1998). Children's sensitivity to word-order violations in German: evidence for very early parameter setting. In A. Greenhill (Ed.), *BULCD 22 Proceedings*, 756–767.
- Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, **72** (3), 655–684.
- Williams, E.J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society*, **21** (2), 396–399.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–128.
- Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, **30**, 156–171.

Received: 27 April 2015

Accepted: 18 March 2016

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Implicit and Explicit false belief development.