Theory of Mind: Four-year-revolution revisited

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Behavior and Cognition (BeCog)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Neşe Oktay-Gür

aus Türkeli/Sinop/Türkei

Göttingen, 2017

Betreuungsausschuss

Prof. Dr. Hannes Rakoczy, Abteilung kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Dr. Tanya Behne, Abteilung kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Markus Steinbach, Seminar für Deutsche Philologie, Universität Göttingen

Prof. Dr. Michael Waldmann, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Hannes Rakoczy, Abteilung kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Korreferent: Dr. Tanya Behne, Abteilung kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Markus Steinbach, Seminar für Deutsche Philologie, Universität Göttingen

Prof. Dr. Michael Waldmann, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. York Hagmayer, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Annekathrin Schacht, Abteilung Affektive Neurowissenschaft und Psychophysiologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Tag der mündlichen Prüfung: 16.05.2017

Preliminary Note

This dissertation is not a cumulative dissertation. However, the Experiments 1 and 2 in Study 2 describe here have been published (Oktay-Gür & Rakoczy, 2017). Some parts of this paper are used in Chapter 7 to describe these studies and their results (with permission from Elsevier).

Thank you…

… Hannes Rakoczy, for great supervision, cooperation, guidance and discussion of research ideas.

… Tanya Behne and Markus Steinbach, for valuable feedback and input in the thesis committee meetings.

… all my colleagues from the Kindsköpfe lab, Regina, Marlen, Konstanze, Hannes, Tanya, Jonas, Marina, Steffi, Lisa, Ella, Ben, Anika, Nadja, Louisa and Franziska, for interesting discussions and a joyful every day.

… Regina, Marlen, Konstanze and Lisa for their great support in organizing the studies.

… the students and research assistants who helped conducting these and other studies, Lisa, Kira, Mareike, Alex, Delia, Frederike, Lydia, Michaela, Natalia, Victoria, Tamara, Nina, Fabienne, Linda, Larissa, Josefin, Isabell, Isabel, Theresa, Manuel, Margarita, Carina, Sandra, and Simon.

…Damaris, Simon, Anika and Franziska for helpful comments on earlier versions of this dissertation.

…Müslim, my husband, who took on many burdens in order to make this dissertation possible for me.

… my parents, my grandparents, my aunt Ayla and her family, for always being enthusiastic about my work.

…, last but not least, to all the children who have participated in my studies and their parents.

Table of Contents

0. Abstract

The *standard picture of Theory of Mind development* is this: Children begin to explicitly ascribe beliefs and other propositional attitudes to themselves and others around age four. Therefore, this has been considered as the age at which a fully-fledged Theory of Mind (ToM) is acquired. This picture is caused by numerous studies consistently showing that children master verbal false belief (FB) tasks reliably from age four on, while younger children fail to do so. The standard picture, though, has been attacked by two sorts of critique. On the one hand, studies using implicit measures show early competences in children much younger than four years of age, leading to an *underestimation claim*. Given children's newly discovered implicit competences the standard picture seems to underestimate children's real competence. On the other hand, studies on the scope of the explicit competence acquired at age four reveal limitations. This challenges the assumption that this competence is fully-fledged and therefore, raises the claim, that the standard picture may overestimate children's real competence. This kind of critique has been called the *overestimation claim*. This dissertation focusses on the latter claim. This claim is borne by studies using two different kinds of tasks, tasks involving aspectuality understanding and tasks on true belief (TB) ascription. For both kinds of tasks, several studies have shown that only children much older than four years of age are able to pass them. Such findings cast doubt on the standard picture of ToM development, especially on the unity of the explicit competence. They bring into discussion the possibility that the competence children classically show around age four is limited or even based on the usage of simpler strategies rather than proper belief reasoning.

In three studies, I investigate the validity of such findings of incompetence. For both, aspectuality understanding and TB ascription, I contrast competence-based with performance-based explanations. According to competence-based accounts children's failure in aspectuality and TB tasks reflect a lack of competence. According to performance-based accounts, however, children's failure is caused by performance problems and the standard picture of ToM development is true. Study 1, following a study by Rakoczy et al. (Rakoczy, Bergfeld, Schwarz, & Fizke, 2015), investigates 4- to 6-year-olds performance on a newly designed aspectuality task with reduced extraneous task demands. Study 1 shows that once aspectuality tasks are suitably modified children are able to solve them as soon as they master classical ToM tasks.

The results of Study 1 therefore support a performance-based explanation of children's former failure. Study 2 investigates TB competences of children older than four years of age. Experiment 1 in this study replicates and even extends findings of incompetence. However, Experiment 2 shows that once the procedure's superficial qualities are changed, children are able to attribute TBs. This again, clearly speaks in favour of a performance-based explanation of children's incompetence in former studies. Study 3 presents and empirically underpins a performance-based framework for the initial problem's emergence. Taken together, these three studies suggest that previous findings of limitations of children's explicit ToM constitute false negatives and that the classical picture of ToM development is justified.

1. Introduction

Social interaction is an indispensable quality of our everyday life. We interact with others on a daily base and devote (at least some) cognitive resources on understanding and predicting their actions. In this process, we understand others and ourselves as minded and assume that observable actions are driven by invisible mental states like intentions, believes, desires, thoughts and emotions. The cognitive capacity enabling us to reflect about these mental states, the so-called *"theory of mind" (ToM)* (Premack & Woodruff, 1978), has been investigated intensively for the last three decades. Developmental psychological research mainly used explicit *false belief (FB)* ascription as the standard measure to investigate ToM development. This research consistently reveals the following pattern. Children younger than four years of age are not able to explicitly ascribe FBs to themselves and others. Whereas older children succeed in explicit belief ascription and related tasks *("standard picture of ToM development")*. This picture led to the assumption of a deep cognitive revolution around age four that equips the child with the necessary cognitive foundation of belief ascription *("four-year-revolution")* (Perner, 1991). Though this claim had already been attacked by different kinds of critique when it was first raised, it has survived. However, different lines of research are questioning its validity again. On the one hand, studies using implicit measures show that even much younger (e.g. 15-month-old) children are able to form expectations about an agent's behaviour based on her (false) beliefs. On the other hand, children older than four years of age seem to fail in two very basic kinds of belief ascription; tasks requiring an understanding of one of the critical characteristics of mental states, namely their aspectuality, and tasks requiring the ascription of a *true belief (TB)* to an agent. These kinds of tasks are found to be mastered much later (around 6 years of age). Therefore, two contradictory questions are raised. Does the assumption of a cognitive revolution around age four underestimate children's competences? In the light of early implicit competences in toddlers this seems to be the case. This is referred to as the *underestimation claim*. In addition, at the same time, does this assumption overestimate children's competences? This seems to be true in the light of incompetence in children older than four years of age. This is referred to as the *overestimation claim*.

The aim of this thesis is to examine the different challenges the four-year-revolution is facing in order to shed light on this recent issue of ToM development.

While I will especially focus on the overestimation claim, I will also consider the underestimation claim. I will start by first providing an introduction to ToM in general (Chapter 2). A closer examination of the underestimation claim will follow (Chapter 3). I will present findings of early implicit competences and their interpretation. Additionally, I will suggest that this critique on the standard picture of ToM development can be resolved by a two-systems account of mindreading (Apperly & Butterfill, 2009). This will be followed by a closer look on the overestimation claim (Chapter 4). I will introduce the critical tasks and theoretical positions predicting competence or incompetence in those tasks. After clarifying the aim of this dissertation (Chapter 5), I will present three studies I conducted within my PhD in which I tested these predictions (Chapters 6-8). The first study addresses children's understanding of aspectuality. This study shows that children are able to solve aspectuality tasks earlier once extraneous task demands are adopted. Moreover, children are able to solve them as soon as they master classical FB tasks. This supports a performance-based interpretation of the initial findings of children's incompetence. The second study explores children's ability to ascribe TBs and shows that children's problems with TB ascription are mere performance than genuine competence problems. Children are able to ascribe TBs if the task is changed in a way that sets TBs in a more meaningful context. The third study aims to provide a framework for children's former poor TB performance and empirically investigates different candidate causes. Finally (Chapter 9) I will discuss my findings and relate them to the standard picture of ToM development and the overestimation claim. I will also provide an outlook on what these findings imply with respect to the underestimation claim.

2. The Standard Picture of ToM Development: The four-year-revolution

ToM is considered as one of the key issues of cognitive development (Wellman & Liu, 2004). From a developmental point of view, ToM offers a variety of interesting research questions. Before these questions can be tackled, it is useful to clarify in more detail what is meant by the expression “Theory of Mind”. The next step is to discuss how this ability is measured and why this special kind of measurement is used. To set the stage for the most important question “How does ToM competence develop?” This chapter aims to shed light on these relevant questions by drawing the standard picture of ToM development that resulted from more than a quarter century of research.

As it was mentioned before, in our everyday life, we perceive ourselves and others as driven by mental states. We thereby assume that observable behaviour is caused by (unobservable) mental states. To predict or explain people’s behaviour within our folk psychology, we need an ability to attribute mental states to others and ourselves. This special ability has been named “Theory of Mind” (Premack & Woodruff, 1978) and it was first investigated in chimpanzees in the context of an action prediction task that measures the ability to ascribe intentions/ goals. Sarah, a twelve-year-old chimpanzee was presented with several videotaped problems, e.g. an actor trying to escape from a locked room, and offered pictures of different solutions to the problem, e.g. an undamaged, a crooked or a damaged key. Sarah was able to reliable select the solution matching the presented actor’s aims. This has been interpreted as showing that the chimpanzee was able to attribute goals to the agent.

For action prediction might be solved in other ways (not necessarily mental; e.g. associations) this study caused a debate on what competence must be shown by an individual in order to be judged as holding a ToM. Dennett's (1978) comment on Premack and Woodruff's (1978) paper suggests that only the ability to ascribe FBs can be considered as a marker for a real belief ascription ability. The reason is that the attribution of a FB covers one of the important characteristics of mental states, namely that mental states do not need to match reality (Wellman, Cross, & Watson, 2001). Wellman and Bartsch (1988) introduced ToM as a research area into developmental research by asking whether children have a ToM. These considerations led to the development of the classical ToM tasks, the *“unexpected content”* and the *“Maxi and the chocolate”* task.

In the unexpected content task by Gopnik and Astington (1988) a child is shown a smarties box and asked what she thinks is in the box. After the child responds by typically saying, that she thinks the box contains smarties, it is revealed that the box actually contains pens. Then the child is asked what she thought was inside the box when she first saw it and what another person, e.g. her best friend, would think is in the box when she sees the box for the first time. A person holding a ToM should be able to predict that (a) she herself had a FB about the content when she first saw the container (1st person) and (b) her friend would falsely belief the container to be filled with smarties (3rd person).

In the Maxi and the chocolate task (also called location change or standard FB task) by Wimmer and Perner (1983) the protagonist, Maxi, puts his chocolate into a green cupboard and goes out to play. In his absence, Maxi’s mother takes the chocolate out of the green cupboard and puts it into a blue cupboard. In this task, children are asked to predict where Maxi will look for his chocolate when he comes back. A person holding a ToM should be able to predict that Maxi will look for his chocolate in the green cupboard because he believes it to be there.

These tasks and several modifications reveal a clear developmental pattern. Children younger than four years of age fail to ascribe FBs, while at the same time succeeding on TB control tasks. TBs are comparable in different aspects (like memory demands) to FBs and are used in order to ensure that children do understand the scenario’s general pattern. Children younger than four years of age did not respond differently in FB and TB. In both belief condition they predict that Maxi will search for his chocolate in the blue cupboard where it really is. For they were not able to pass the FB task, it was concluded that they lack a ToM. Only children older than four years of age consistently succeed on FB tasks (location change and unexpected content tasks) by taking into account the agent’s FB. This developmental milestone has been named “four-year-revolution” and it seems to be robust. Children’s performance is not influenced by different superficial parameters like, (a) the nature of the protagonist the belief is ascribed to (e.g. puppet, picture of a person and so on), (b) the nature of the target object the belief is about (e.g. a real object or a toy), (c) the type of question used in the task (e.g. “Where will the protagonist look for the object?” or “What does the protagonist think?”) or (d) the type of task used (e.g. smarties or maxi and the chocolate). However, the performance of children who are around age four can be boosted by factors like explicit deception, active participation of the child, highlighting the salience of mental states or reducing the salience of the reality-belief contrast (Wellman et al., 2001). Additionally, it has been shown that first and third person belief ascriptions develop at the same time. This developmental pattern is consistent across cultures (Callaghan et al., 2005). Furthermore, ToM is an ability that is uniquely human[1]. Moreover, the emergence of an explicit ToM around age four goes along with the emergence of competences on other tasks that require an understanding of representations (e.g. understanding of identity expressions). Taken together these findings support the assumption of a four-year-revolution by showing its far-reaching validity.

What is it that develops around age four? There are different possibilities. Nativist accounts supporting an early presence of the ToM competence suggest that children may have these competencies from early on (or even innately) and that these competences are masked by task demands, information processing limitations or confusion (Baillargeon, Scott, & He, 2010; Carruthers, 2013; Fodor, 1992; Leslie, 2005). What develops around age four, then, is not ToM itself but more likely something enabling them to overcome these extraneous limitations and to show their real competence. These nativist accounts clearly contrast with conceptual change accounts suggesting that there is indeed an elementary change in how children think about mental states. There are different possibilities for the nature of this change. Modularity theories (which are kind of a hybrid account) suggests that the (innate) module for belief ascription maturates (e.g. neurologically) (Leslie, 1994). Simulation theories in contrast suggest another mechanism. They assume that we ascribe mental states to others by simulating our own inner life in their situation (Goldman 1992; Gordon 1986; Harris 1991). Children's ToM competence therefore improves the more experienced children get with the perspectives of others'. Theory-theories, that assume that children use theories to understand other's minds, in contrast suggest that what changes is the theory children use to explain behaviour. Within theory-theories again there are several possibilities. The theory children use may change from a desire to a belief-desire naïve psychology (Wellman & Woolley, 1990), from a connectionist to a

[1] Note that there is one recent study suggesting that chimpanzees are at least implicitly able to ascribe false beliefs. However, this contradicts with a longer tradition of studies in which non-human primates consistently fail to do so.

representation-based understanding or from a situation-based to a representation-based theory.

One theory on ToM development will be given special consideration in this dissertation, the representational ToM development suggested by Perner (1991). In this theoretical framework, the critical ability that is acquired around age four is a concept of representations. This concept is a critical condition in order to form a representational ToM that allows an understanding of beliefs and other propositional attitudes. Representations represent a referent; they are mental and therefore aspectual (represent the referent under one description); and they allow for nonexistence (the referent does not necessarily have to exist in the real world, e.g. one can have a representation of unicorns) and misrepresentation (they are not unconditionally true, e.g. FBs). Children around age four become able to represent the minds of others using representations and therefore are able to handle potential non-existence of the referent, misrepresentation and aspectuality. Therefore, their concept of beliefs is full-blown. Before this age, children use a theory of behaviour, that only operates with relations (in contrast to representations). This theory of behaviour is not capable of non-existence of the referent, misrepresentation and aspectuality and is later replaced by the new representational ToM.

Taken together, the standard picture of ToM development seems to be the following. Children younger than four years of age are not able to ascribe FBs. Their incompetence may be caused by extraneous task demands or more likely by the lack of an explicit understanding of mental states. Around age four children consistently form the cognitive foundation of belief ascription and gain a full-blown conception of belief and other propositional attitudes. However, there are different accounts on the question what it is that emerges around that age. I favour the representational ToM theory by Perner (1991) described above, which suggests that a former relational theory of behaviour is replaced by a representational ToM. Even though this standard picture seems to be consistent and coherent, it has been criticized from two contradictory directions right from early on. For one, it has been considered as overestimating children’s ToM competence, mainly because it assumes the acquired ability to be full-blown. Secondly, it has been considered as underestimating young children’s competences like it is suggested by nativist accounts

In the next chapter, I will present different underestimation accounts, supportive empirical evidence and a theoretical possibility to integrate these findings into the standard picture of ToM development.

3. Underestimation Claim

This chapter will focus on critique of the four-year-revolution from an early competence point of view. Early competence accounts claim that the conclusion that children acquire a full-blown ToM around age four underestimates young children's competences. First, I will present findings of early competences. Then, I will relate these to different theoretical positions of ToM development. Finally, I will present the two-systems account of mind reading proposed by Apperly and Butterfill (2009) as a solution to the tension caused by these findings.

3.1. Findings of early competences in children

Postulating an age limit for the acquisition of a full-blown ToM (four-year-revolution) raises questions on what is going before that age. How do younger children operate with mental states before age four? One option that appears likely is that younger children cannot handle mental states as such at all. This position is taken by the representational ToM theory by Perner (1991). This theory assumes that children younger than four years are not able to handle representations. Understanding beliefs requires an understanding of representations, therefore children younger than four years of age cannot handle beliefs. However, children younger than four years of age do show behaviour that can be interpreted as signs of an early mind reading ability. For instance, they show eye movement behaviour in accordance with the belief of an agent. From the representational ToM theory point of view, children younger than four years of age are able to show such behaviour that looks like belief ascription because they have a relational theory of behaviour. This relational theory of behaviour allows children to implicitly predict behaviour using behavioural rules without any belief reasoning. But the four-year-revolution clearly underestimates children's real competence if children's early competent behaviour (e.g. correct eye-movement) does indicate a real ToM competence that uses the ascription of beliefs in any way. The idea that this might be the case has been present since the very first days of the assumption (Perner, 1991).

Previous research has shown some indicators for early belief ascription (Clements & Perner, 1994). Children younger than four years of age (in this case 3-year olds), who failed to explicitly ascribe FBs in a location change task (Maxi and the Chocolate) showed correct anticipatory looking behaviour when mistaken Maxi reentered the situation to search for the chocolate. Whereas children claimed that mistaken Maxi would search for the chocolate where it really is (wrong answer), children did look at the location where a person with a FB would search for the chocolate (correct answer). This observation indicates an early understanding which children are merely not able to express explicitly and it fits perfectly into a nativist view (e.g. Carruthers, 2013).

Recent research that uses different and partially new methods, has provided even stronger evidence. The underestimation claim gained momentum when studies using eye tracking in *violation of expectation (VoE)* tasks were conducted. In these tasks, children were presented with an actor who either was mistaken about an object's location or not. 15-month-olds looked longer at a given scene, when the actor showed behaviour that conflicted with her belief. Children were surprised when both, a mistaken agent reached for an object where it really was and when an agent with a TB showed the opposite behaviour (Onishi & Baillargeon, 2005). Several other studies confirmed these findings (Kovacs, Teglas, & Endress, 2010; Song & Baillargeon, 2008; Surian, Caldi, & Sperber, 2007; for an overview see Baillargeon et al., 2010).

Similarly, in *anticipatory looking (AL)* studies, children expected a protagonist to act in accordance to her FB. In the Southgate-Senju paradigm (Southgate, Senju, & Csibra, 2007; also used in adults, see Senju, Southgate, White, & Frith, 2009) 25-months-olds were presented with a movie showing an agent behind a divider with two openings. Right under these two openings, two boxes were placed (from the child's perspective). Children first saw several familiarization trials, in which an object is hidden in one of the two boxes, the windows are illuminated, a tone sounds and the agent reaches through the window behind the box containing the object. After three trials, participants showed anticipatory looking behaviour (looking at the right box after the illumination and the sound). In the following test trials, the agent sees how the object is placed in one of the boxes. Thereafter she is distracted by a phone ringing in the background that causes her to look in a different direction. To introduce a FB of the agent, the object is then removed from the scene. When the agent turns her head back to the scene, the illumination and the sound follow. Similar to the familiarization trials, participants again showed anticipatory looking behaviour. They expected the agent to search for the object were she mistakenly believed it to be. Critically, adults with Asperger's syndrome were also tested in this task. Individuals with Asperger's syndrome are known to have difficulties with explicit ToM tasks. Therefore they are a critical sample in order to investigate whether these implicit competence really show ToM competence. In this study, adults with Asperger's syndrome did not show a sensitivity to the agent's belief in their AL behaviour. This reinforces the assumption that these tasks measure some sort of ToM competence.

Even less implicit methods provide evidence for an early belief understanding. For example in some helping scenarios children were demonstrated to show a sensitivity to the beliefs of others (Knudsen & Liszkowski, 2012; Southgate, Chevallier, & Csibra, 2010). Buttelmann and colleagues (Buttelmann, Carpenter, & Tomasello, 2009) confronted 15- month- olds with an interaction situation in which two boxes were present and children were shown how to open the lock mechanism of these boxes. An actor, who did not know how to open the boxes, was then introduced. The actor placed an object in one of the two boxes and left the situation. Either in his absence or in his presence the location of the object was changed, resulting in a FB or a TB of the agent about the object's location. Upon his return, the actor always tried to open the empty box. In the FB condition he tried to open the box he believed the object to be, in the TB condition he tried to open the box he initially had put his object in but knew to be empty. Children's helping behaviour differed depending on the agent's mental state. When the agent had a TB about the object's location and tried to open an empty box, children helped him to do so. However, when the actor had a FB about the object's location and tried to open the box he falsely believed to contain the object, children showed a different behaviour. In this case, they helped the actor to open the other box containing the object, the one not targeted by the agent.

Comparably, in referential communication tasks, children take into account the (false) belief of an agent when they infer the name of an object (Southgate et al., 2010). In this task, children saw how an agent placed two different objects in two boxes. Either in the agent's presence (TB) or the absence (FB) another experimenter exchanged the objects. Upon her return, the experimenter pointed at one box, claiming that there is a "Sefo" in the box and asking the child to give her the "Sefo". In the TB condition, children took the object from the targeted box in order to give it to the agent, whereas in the FB condition they chose the object from the non-targeted box. This indicates that children were able to take into account the protagonist's belief when she provided the label for the object.

Taken together these findings suggest that there are early competences in children who are yet not able to pass explicit ToM tasks. This contradicts conceptual revolution accounts on ToM development and is therefore a problem for the assumption of a four-year-revolution.

3.2. Relation of these findings to different theoretical positions

In the following, I will present different ways of integrating findings of early implicit competences in children younger than four years of age into the standard picture of ToM development. These possibilities differ in their impact on the standard picture; some of them (e.g. nativist accounts) are longing for a revision, while others are compatible (e.g. behavioural rule account).

Behavioural rule accounts have been used to explain young children's performance on implicit ToM tasks without ascribing them real ToM competences. Like it has been described above in Perner's (1991) theory of ToM development, for example, he postulates, that children younger than four years of age do not have a ToM but a "theory of behaviour" (also called "situation theory"). Perner (1991) suggests that, e.g. when children in the smarties task are asked, what they thought was inside the box when they first saw the box, this question makes them theoretically reconstruct the former situation. Critically, the reconstruct the former situation in accordance with the given information at the current point of time. For the child now knows that the box contained pencils, it cannot construct a situation that does not match this reality. Therefore, the child answers that it thought that there were pencils inside the box. This is an example for the blind spot of situation theory. The situation theory, that children use, is not suitable for misrepresentation. Situation theory cannot be used to construct a situation that misrepresents the current state of the world. Therefore, this cannot explain children's success on implicit FB tasks because an understanding of misrepresentation to solve these tasks. Perner and colleagues (Clements & Perner, 1994; Perner & Ruffman, 2005; Perner & Roessler, 2012) suggest that children with a situation theory use behavioural rules, so called "implicit social knowledge of lawful regularities" (p.519, Perner & Roessler, 2012), to form an expectation about the behaviour of an agent. These behavioural rules would enable children to expect, e.g. that an actor will search for an object where he last saw it. This behavioural rule can explain findings showing a sensitivity of children's eye movement for the beliefs of an agent. When an agent is searching for an object, according to this rule, it is useful to expect him to look for it where he last saw it. Therefore, this behavioural rule predicts that an agent with a FB will look at the wrong location and makes children look at the right location while explicitly giving wrong answers. It is important to note that the usage of behavioural rules can explain children's success without ascribing them belief reasoning abilities.

In a similar way, Heyes (2014) suggests that domain general cognitive mechanisms can explain findings from VoE and AL studies. Such an explanation again does not need belief ascription. For the findings from the Senju-Southgate paradigm ((Southgate et al., 2007), for example, she suggests a distraction-based explanation. Children's looking behaviour, which was interpreted as indicating a belief ascription, can be explained as own belief disclosure. This account suggest that when the phone rings in the background and the agent turns her head, the participant herself is distracted because she looks at the head and the area the agent pays attention to. This makes her pay less attention to the change of location taking place at the same time. This results in a FB of the participant herself about the object's location. Therefore, the AL behaviour shown right after the illumination and the sound shows the participants' and not the agent's FB about the object's location. This alternative explanation also holds for the non-existence of this pattern in adults with Asperger's syndrome. Individuals with Asperger's syndrome are known to show less joint attention behaviour (Charman et al., 1997) and therefore, would pay less attention to the agent's head movement and be less distracted. This would then results in a TB of the participant that the object is removed from the scene and they would therefore not show any AL behaviour.

However, these and similar alternative explanations are limited in several ways. First, they cannot count for less implicit competences shown in children like their helping behaviour and their ability to take into account a communicator's belief in a referential communication task. Another critical point is that alternative explanations can only explain local findings. There is no alternative explanation, which can explain away all kinds of evidence in favour of implicit ToM competences. Alternative explanations that only account for local findings do not seem to be satisfying in the light of the growing body of studies that use different paradigms and that support the existence of an implicit ToM.

Another possibility is to have a nativist view on the presented data. From a nativist point of view, implicit ToM findings show what nativists already knew. Once task demands are reduced, children's real competence is revealed. Children are already handle mental states but their competence is masked in explicit ToM tasks (Leslie, 2005; Leslie, German, & Polizzi, 2005; Luo & Baillargeon, 2007; Onishi & Baillargeon, 2005; Scott & Baillargeon, 2008; Surian et al., 2007). Therefore, previous findings of incompetence of children younger than four years of age are interpreted as false negatives that are caused by extraneous task demands such as general processing capacities, language abilities, executive functions, and working memory (Leslie, 2005; Carruthers, 2013). This possibility would then call for a complete revision of the standard picture of ToM development. Instead of assuming that ToM is acquired around age four, from a nativist point of view it would be more interesting to investigate if ToM is present from birth on or if it develops later (earlier than four years of age).

A third possibility to explain children's early competences is the two-systems account of mindreading by Apperly and Butterfill (2009). This account suggests the existence of two systems of belief ascription.[2] One early developing (probably even innate) and cognitively efficient System 1 and a later developing and cognitively demanding System 2. System 1 is, therefore, responsible for findings of early competences. While System 2 is needed to explicitly ascribe mental states. The distinction between the systems can additionally account for findings of automatic belief ascription in adults (Kovacs et al., 2010; Newton & Villiers, 2007; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). Despite the implicit- explicit distinction, Apperly and Butterfill (2009) suggest that the systems operate with different kinds of mental states. While the more sophisticated System 2 operates with propositional attitudes (i.e. full-blown understanding of belief), these kinds of representations would be too demanding for an implicit and automatic system like System 1. Therefore, Apperly and Butterfill (2009) postulate that System 1 does not operate with a fully-fledged understanding of beliefs but with relational forms of representations, namely with registrations. Registrations represent the relation

[2] Note that the most important reason to consider a two-systems account is the following. The assumption of two systems that operate independently, can explain situations in which participants show two contradicting behaviour. In belief reasoning this is, e.g. the case when children explicitly fail to attribute a FB, but implicitly show eye-movement that takes into account the belief of another agent. With two systems, this pattern can easily be explained. While System 1 is responsible for the implicit correct eye-movement behaviour, System 2 (or its absence) is responsible for the incorrect explicit answer.

between a real object in the world, its properties and its location in the world. This allows for the computation of location change FB tasks resulting in a (looking) behaviour that looks like belief understanding. The mechanism to solve FB tasks with System 1 is the following. When things change in reality, e.g. the location of an object is changed, and an agent cannot update his/her registration, then he/she has a false registration of the object. For children can represent this registration, they can act in a way that seems like belief attribution.

However, registration lacks some fundamental features of propositional attitudes. Propositional attitudes like beliefs represent objects always under some aspect or description but not under others, making them aspectual (Anscombe, 1957; Searle, 1983; McKay & Nelson, 2014). The following example illustrates the problem. Lois Lane thinks that Superman can fly. Clark Kent is Superman. But in the comics and the movies, Lois Lane does not know that Clark Kent is Superman (despite the fact that Clark Kent looks exactly like Superman with glasses). In this case, it is crucial to notice the aspectuality of Lois' belief. When she thinks about Superman and Clark, she is thinking about two different persons while in reality they are the same person. When Lois is given the information that Clark Kent is at the bar, she is not able to conclude that Superman is at the bar. When representing Lois' mind, its aspectuality must be represented, too. The two-systems account of mindreading makes clear and testable predictions about the ability of Systems 1 and 2 to handle aspectual contexts like this. While System 2 is able to understand aspectual FBs (Lois does not know that Clark Kent is Superman), it is a signature limit to System 1. Whether or not children can implicitly solve tasks involving aspectuality is important because (as mentioned before) fully-fledged belief understanding includes aspectuality understanding. If aspectuality is a limitation to children's early belief ascription ability, this indicates that findings of an implicit competence are not necessarily caused by a full-blown ability to ascribe mental states. Such a finding would be compatible with the standard picture of ToM development.

There are several studies that investigate early mindreading competences in situations that are supposed to contain aspectuality. They show that children are able to implicitly take into account FBs about identity (Scott & Baillargeon, 2009; Scott, Baillargeon, Song, & Leslie, 2010; Scott, He, Baillargeon, & Cummins, 2012). The most convincing study from this set of studies is the one by Scott and Baillargeon (2009). In this study, 18-month-olds' looking time was measured. Infants observed scenes containing two indistinguishable objects, a one-piece penguin and a two-piece penguin. In the familiarization phase, the protagonist demonstrated her aim to hide a key inside the two-piece penguin, which was always presented disassembled. In the test phase, the protagonist was presented with a complete penguin in a transparent box. The protagonist either knew (TB) or did not know (FB) that the visible object was the desired penguin. In both belief conditions, children looked longer when the agent behaved inconsistently with her belief. These results show success on an implicit task involving aspectuality and therefore, clearly conflict with the prediction of the two-systems account. However, the realization aspectuality in these studies has been criticized. Butterfill and Apperly (2013) suggest that the tasks used in these studies are about properties of object, not their identities. However, they must be about identity in order to be aspectual. Otherwise, they can be solved in easier ways that do not include aspectuality. Two sets of studies have investigated implicit ToM by comparing beliefs about location and beliefs about identity more comprehensively and more straightforwardly. The first set of studies by Low and colleagues (Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013) presented 3- and 4-year-olds and adults with a location change and an identity task and measured implicit AL and explicit answers to test questions. In the identity task an object with two aspects was presented (it was blue on the one side and red on the other side). For the protagonist did not know about this dual identity he at some point arrived at a FB about the location of a preferred object. While all age groups showed correct anticipatory looking behaviour in the location change control tasks, only 4-year-olds and adults explicitly solved these. In the identity task, however, none of the age groups showed correct anticipatory looking behaviour while 4-year-olds and adults again were able to solve the explicit task. These findings have been interpreted as confirming aspectuality as a signature limit to implicit ToM competences. However, this interpretation has caused some debate (Carruthers, 2013, 2016, 2017). From a nativist point of view, one alternative explanation for children's failure in this implicit identity task is that this task is more demanding than the location change control task and therefore, children's competence is masked.

In order to overcome these limitations, a second set of studies by Fizke and colleagues (Fizke, Butterfill, van der Loo, & Rakoczy, 2014) used a modified version of the classical helping paradigm (Buttelmann et al., 2009) to implicitly measure children's understanding of FBs about location. Additionally, they also designed a new version that aimed to test FBs about identity. In the identity version soft toys (e.g. identity A: bunny) were used that could be turned inside out and thereby transformed into something else (e.g. identity B: carrot). The object was hidden under its identity A in one box. Either in the protagonist's presence (TB) or her absence (FB) the object was transferred into its identity B and put back into the initial box. The object was then, in the presence of the protagonist (both conditions), put on the floor under its identity B. In both belief conditions the protagonist now tried to open the (empty) box. Fizke et al. (2014) replicated the findings of Buttelmann et al. (2009) for the location change task. Children's helping behaviour differed between FB and TB. In the identity task, however, no such sensitivity to the agent's belief was found. This again demonstrates that aspectuality is a signature limit to children's early competences and clearly supports a two-systems view on belief ascription.

3.3. Summary

Early belief ascription competence seem to be very robust. However, they have clear signature limits such as aspectuality. Therefore, the two-systems account seems to be suitable to the spectrum of findings. The two-systems account is also a way to integrate findings of early competences in children into the standard picture of ToM development. In this case, findings of early competences would not challenge the standard assumption. Children do acquire a full-blown (meta-) representational conception of beliefs and other propositional attitudes around age four (System 2). Before that age they already have an understanding for relations between persons and objects, enabling them to show behaviour in implicit tasks that can be interpreted as an implicit belief tracking competence (System 1). However, this system is qualitatively different from System 2. It is not just the same system that operates implicitly. The main support for this assumption comes from studies that show that not all kinds of tasks can be solved by System 1, even if they are implicit.

Taken together the two-systems account is a possibility to defeat the underestimation claim and to maintain assuming a four-year-revolution. However, as already mentioned, the standard picture is also challenged by the opposite claim, namely that it overestimates children's competences by ascribing them full-blown ToM competences. In Chapter 4, I will present studies that raise doubt on the "full-blown" aspect of the ToM competence that is acquired around age four.

4. Overestimation Claim

In this chapter, I will focus on the overestimation claim. This claim says that the four-year-revolution overestimates children's real ToM competence given at age four. The basic idea is that when children reliably pass explicit FB tasks, they still fail to master some sorts of belief ascription tasks. Two kinds of tasks have been discussed in this area, tasks requiring an understanding of the aspectuality of beliefs and TB tasks. Children between four and six years of age have been shown to fail these tasks. However, part of a full-blown ToM is a fully-fledged understanding of beliefs and other propositional attitudes. This also includes to understand the aspectuality of beliefs and the fact that beliefs can be true. Therefore, findings of such failure raise doubts whether children who master classical explicit FB tasks really have a full-blown understanding of beliefs. However, it remains open whether children's failure is caused by competence limitation or mere performance problems. For both kinds of tasks (aspectuality and TB), I will present data supporting a competence limitation view, relate these to competence limitation accounts and finally suggest a performance-based alternative explanation for children's poor performance.

4.1. Understanding the Aspectuality of Beliefs

As it was mentioned before, one of the fundamental features of propositional attitudes, e.g. beliefs, is their aspectuality. They represent entities of the real world under some aspect but not under others. When we talk about propositional attitudes we use intensional expressions like "Lois believes that *p*". In such intensional expression co-referential terms (e.g. Superman/ Clark Kent) should not be exchanged (in contrast to extensional contexts, which describe the world outside a mind). For example, in the sentence "Lois believes that Superman can fly", one cannot exchange "Superman" and "Clark Kent" without the sentence losing its validity. It is not true to say "Lois believes that Clark can fly", for she does not know that Clark is Superman. Having a full-blown ToM and thus understanding the aspectuality of propositional attitudes therefore should prevent one from doing such, so-called, extensional errors in intensional contexts.

Russell (1987) investigated 5- to 7-year-olds competence to judge the permissibility of such extensional errors (replacing co-referential expressions) in intensional contexts. Children were presented with a story in which an agent, George, had less information than the child about a thief, who stole his watch. The additional piece of information the child was given was that the thief had curly red hair. Now children were asked “Can we say that George was thinking: ‘I must find the thief who stole my watch’?” The correct answer here is of course “yes”. Nevertheless, additionally children were asked “Can we say that George was thinking: ‘I must find the man with the curly red hair who stole my watch’?” This exchange of co-referential terms is not permissible because George did not know that “the thief” was “the man with the curly red hair”. However, 5- to 7-year olds did not deny the validity of the latter sentence. Therefore, children made extensional errors in intensional contexts. This was interpreted as indicating a lack of understanding the aspectuality of beliefs and therefore supporting the assumption that the ToM ability children acquire at age four is not full-blown.

One critical aspect of this study is that it is linguistically demanding. Children have to judge the permissibility of sentences which have a complex grammatical structure. Therefore, Apperly and Robinson (1998) have argued that it remains unclear if Russell’s findings show an incompetence to understand aspectuality or the lack of critical linguistic knowledge. To overcome the limitation of this task Apperly and Robinson (1998)) designed a procedure with reduced linguistic demands. In their tasks partial knowledge about an object caused the aspectuality of a belief (also see Sprung, Perner, & Mitchell, 2007). They used the following procedure. P is looking for an eraser.

(1) There is an eraser in Box 1
(2) There is a dice in Box 2
(3) The dice is also an eraser
(4) P knows (1) and (2) but doesn’t know (3)
Test Question: where is P going to look for an eraser?

Given (4), the answer is clear. P is going to look for an eraser in Box 1 because P doesn’t know that the dice in Box 2 is also an eraser. Only children older than six years of age were able to reliably solve this task by giving the presented answer. This has again been interpreted as showing a real competence deficit in 4- to 6-year-olds. They are able to succeed in FB tasks about the location of an object but yet are not able to take into account the aspectuality of belief when handling mental states. If this is the case, to assume that children acquire a full-blown ToM competence at age four again seems to clearly overestimate the extent of this ToM competence.

However, even this simplified version of an aspectuality task has been criticized as imposing some additional demands on children (Rakoczy et al., 2015). One possible problem is still a linguistic one. The child has to engage in reference resolution to understand the question. Both, the eraser as well as the dice, are erasers, which means that “eraser” refers to both of them. The child now has to select the intended referent, namely the obvious eraser. This is an additional demand that is not related to the real aim of the task. Therefore Rakoczy and colleagues (Rakoczy et al., 2015) designed a task with even more reduced extraneous demands that is parallel to a location change task. In their aspectuality task to ascribe a belief about an object’s location to an agent, one needs to take into account the aspectuality of the agent’s belief (note, that this procedure is similar to the one used by Fizke et al., 2014). The procedure’s basic logic is the following. An object A, e.g. a pen is put in Box 1 in the presence of a protagonist. In the absence of the protagonist it is revealed that A (the pen) is also a B (rattle). Upon the protagonist’s return the object is transferred from Box 1 to Box 2 under its (hidden) identity B (e.g. the experimenter covers the pen with her hands and rattles it on the way to Box 2). The child is asked where the protagonist will look for the A. The protagonist falsely believes the A to be still in Box 1 because she does not know that the object she saw being transferred to Box 2 was the A. In this study 4- to 6-year-olds were able to solve this aspectuality task. Moreover, performance on this task was correlated to their performance on the standard location change task. These results indicate that children can consider the aspectuality of beliefs as soon as they can attribute FBs once performance factors are suitably reduced. However, the way in which aspectuality was realized in this experiment can be criticized. Although it is unlikely, it is possible that children did not perceive the target object’s different identities as different identities. It is also possible to imagine the pen/rattle as a pen that rattles or a rattle you can use to write. Therefore, more research has to be done in order to investigate the validity of this finding.

4.2. Attributing true beliefs: insights from a former control task

As it was mentioned before, studies on 4- to 6-year-olds’ ability to attribute TBs are another kind of studies that challenge the standard picture of ToM development. In TB tasks a child is asked to attribute a belief which is true and therefore shared by the child herself. This kinds of tasks have classically been used as mere control tasks with children who are not yet able to explicitly ascribe FBs. This was done in order to ensure that they do understand the structure of a given task. Two kinds of TB tasks play a role in the overestimation debate, classical TB tasks and aspectuality TB tasks. Children between four and six years of age have been found to struggle in both kinds of TB tasks. What do these findings show? One possibility is that they show true incompetence of children of this age group to ascribe TBs. Another possibility, however, is that this pattern is caused by a performance problem. But first, a closer look on the empirical findings is needed.

The TB version of a location change task is parallel to the FB, with the following exception. The protagonist does witness the object's transfer object from Box 1 to Box 2. To ensure parallelism between the tasks in a lot of TB control conditions, the protagonist also leaves the situation, misses some action, like the object being taken out of a box and being put back in to the very same box, but does not end up holding a FB. Some recent studies using TB versions of unexpected content and location change tasks have revealed a surprising finding. Children from age four to six, who are able to pass FB tasks, fail to attribute TBs and only children older than six years of age succeed in both tasks.

Fabricius and colleagues (Fabricius, Boyer, Weimer, & Carroll, 2010) tested 3.5, 4.5-, 5.5- and 6.5-year-olds in TB versions of the classical unexpected content and location change tasks. In the unexpected content TB task the child was presented with a Smarties container and asked what she thinks is inside the box. Like in the classical version, it was then revealed that the container was actually filled with pencils. Other than in the classical version, the container was then emptied and filled with smarties. Children were now asked what another child would think is in the box when he sees the box. Since the box does contain Smarties, the belief to be attributed is true. In a similar way, minimal adoptions lead to a TB version of the location change task. After Maxis chocolate had been placed in one cupboard and Maxi had left, his sister took the chocolate out of the cupboard. She thought about putting it into another cupboard, but finally decided to put it back into the initial place. Now the child was asked where Maxi will search for the chocolate. For the object's location had not been changed, Maxi's belief was true. Additionally, in both tasks, children were asked to justify their answers. The results were surprising. Most 3.5-year-olds succeeded on both tasks, whereas most 4.5- and 5.5-year-olds failed this alleged easy tasks. At 6.5 years of age again most children succeeded. This means, that a u-shaped age-related development is observed. The age groups also differed with respect to the justifications. 3.5-year-olds referred to reality (reality reasoning); most 4.5-year-olds and 5.5-year-olds gave justifications related to the protagonist's perceptual access, like "he did not see" (perceptual access reasoning), and most of the 6.5-year-olds referred to the belief of the protagonist in order to justify their answer (belief reasoning).

These findings led Fabricius and colleagues (Fabricius et al., 2010) to the following suggestion of the developmental change children would undergo. Children younger than four years of age are "reality reasoners". When they are asked about the beliefs of another person, like "What does the protagonist think where the object is?" they just reason about reality ("Where is the object really?") and give an answer based on this. In TB this answer is correct, whereas in FB this answer is incorrect. Children between age four and six, however, are in a different developmental stage. They use perceptual access reasoning following two rules in order to answer belief related questions. (1) Seeing leads to knowing and (2) knowing leads to acting correctly. Likewise, (1') not seeing leads to not knowing and (2') not knowing leads to acting wrongly. Hence, when 4- to 6-year-olds are asked where the protagonist believes the object to be, they check whether the protagonist saw everything. When the protagonist leaves the situation in FBs, his perceptual access to the situation is interrupted. Therefore, he will get things wrong. The only way to get a FB wrong is to search at the false location. This answer is correct in FB tasks and children succeed. However, the very same mechanism leads to the TB failure. For the protagonist did not have full perceptual access (missed how the object was taken out and put back again) children expect the agent to act wrongly. Only older children who are able to use belief reasoning would be able to ascribe both kinds of beliefs correctly. Thus far, this theory is the only one on ToM development that can explain these findings.

However, another theory on ToM development makes a very similar prediction for another subset of TBs, namely aspectual TBs. The Mental File Card Theory (MFCT) by Perner and colleagues (Perner, Huemer, & Leahy, 2015) is a formal theory on belief reasoning that uses mental file cards (Recanati, 2012) (see Figure 1). The basic idea of this theory is that entities of the real word are represented using mental files. These files are representational structures that individuate their referent (Superman) (see Figure 1, (1)). They also include predicative information like "can fly". When new objects are encountered in discourse or thought, a new file is formed, i.e. when I think about Superman a file is formed, when I think about Clark Kent an additional file is formed that individuates Clark (see Figure 1, (2a)). Around age four children become able to understand identity statements like “Clark Kent is Superman” (Perner, Mauer, & Hildenbrand, 2011). This is enabled by learning to form horizontal links between mental files representing the same referent, like File A: “Clark Kent” and File B “Superman” (see Figure 1, (2b)). These horizontal links allow the flow of predicative information. When I know that Superman can fly and I know that Superman is Clark Kent, I now know that Clark can fly. Around the same age children additionally learn to represent the content of other’s minds using, so-called, vicarious files. Vicarious files are vertically linked to one’s own file of a referent and represent some else’s representation of the same referent (see Figure 1, (3), dotted lines show vertical linking, files in thinking bubbles show the vicarious files). When I think about what Lois thinks about Clark, I link my own File A to a vicarious File A’ representing what Lois thinks about Clark (vertical linking). This architecture allows children to ascribe aspectual FBs. When Clark Kent is first at the office of the “Daily Planet”, the newspaper he is working for, and then Lois and a reasoner see Superman flying onto the top of the building, reasoners would update their File B “Superman” and Lois File B’ “Superman”. For Files A “Clark Kent” and B “Superman” are horizontally linked a reasoner would know that Clark is now at the top of the building across the street. But a reasoner would also be able to represent that Lois’ File A’ is not updated because her Files A’ and B’ are not horizontally connected (because she does not know that A=B).

More critically, the MFCT postulates, that children around age 4 are not yet able to coordinate both kinds of linking (horizontal and vertical) at the same time. To represent aspectual TBs, however, a reasoner needs to coordinate horizontal linking in vicarious files vertically linked to the reasoner’s own files. This assumption leads to an interesting prediction: if children were not able to coordinate horizontal and vertical linking, they would not be able to represent TBs about identity. What does this imply with respect to a reasoner’s representation of Lois’ representation of Clark Kent and Superman? When Lois finds out that Superman is Clark Kent, now the vicarious File A’ and B’ representing Lois representations of Clark Kent and Superman must be horizontally linked. The MFCT predicts that children between four and six years of age would fail aspectual TBs by behaving as if Lois did not know that Clark is Superman because they are unable to handle this complex built up of horizontal and vertical links.

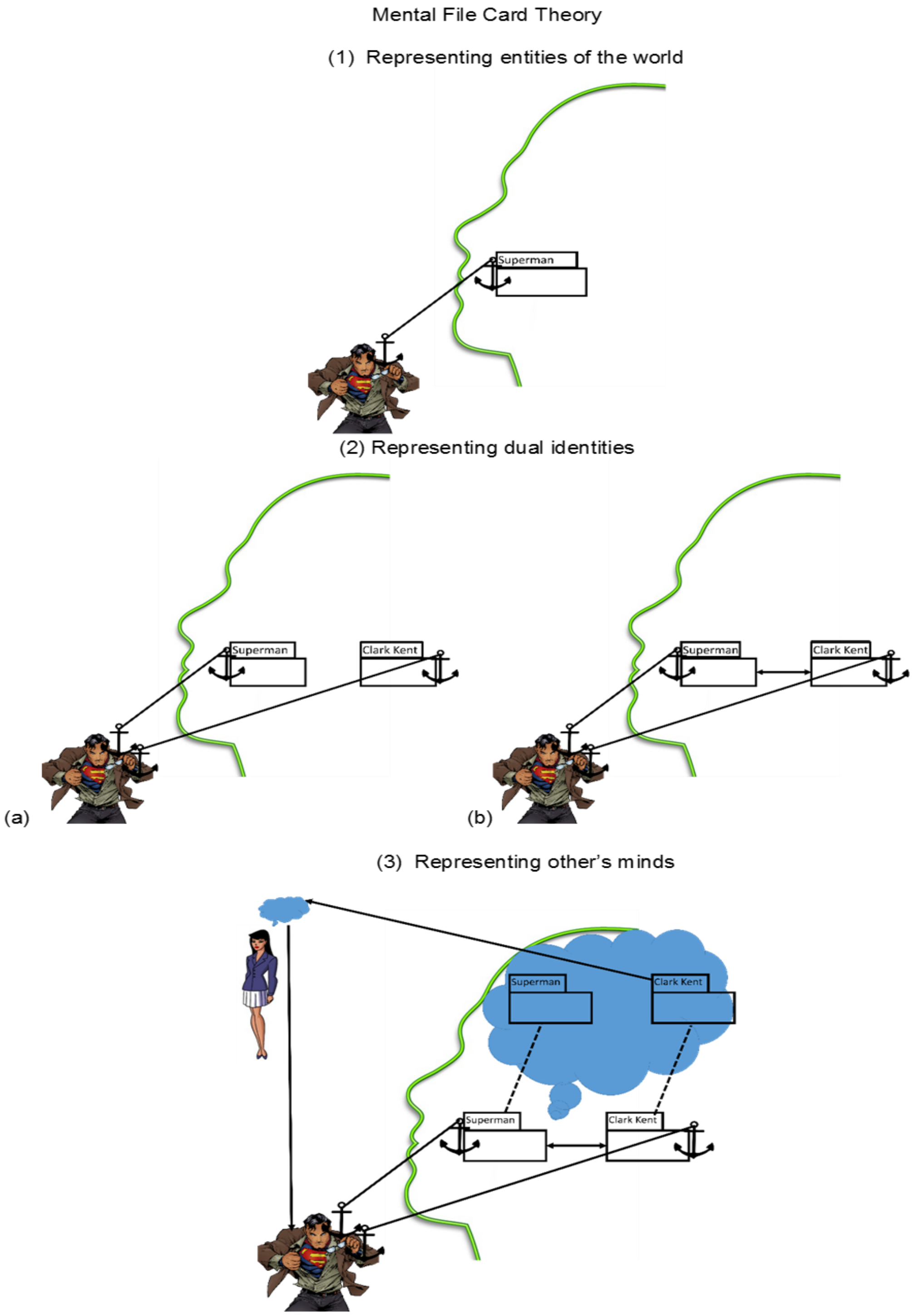


Figure 1. Mental File Card Theory. (1) Entities of the world are represented using mental file cards. These contain individuating information (“Superman”) and descriptive information (“can fly”). (2) Entities with two identities are represented using two mental file cards (a). Around age four, children learn to represent dual identities by connecting files that represent the same referent with horizontal linking (b). (3) Minds of others' are represented using vicarious files that are linked vertically to one's own files. The figure contains a correct representation of Lois' representation of Clark Kent. The files representing her representation of Clark Kent and Superman are not linked because she does not know that Superman is Clark Kent.

Perner and Colleagues (Perner et al., 2015) tested this surprising prediction of incompetence in TB tasks between four and six years of age in a recent study. Children were presented with aspectual tasks after the tasks used by Rakoczy and colleagues (Rakoczy et al., 2015). An object with a dual identity, e.g. a pen that was also a rattle, was first placed in Box 1 and then transferred to Box 2 under its non-overt identity (rattle). This was always done in the presence of a protagonist. What differed between FB and TB was whether the protagonist knew (TB) or did not know (FB) the object's dual identity. As in the original study by Rakoczy and colleges, (Rakoczy et al., 2014) 4- to 6-year olds were able to solve the false belief version of the task. They answered that the protagonist would think that the pen is in Box 1 because he did not know that the object he had seen being rattled and put into the other box was the pen. However, in TB, they again claimed that the protagonist had a FB about the location of the pen even though it was clear that he knew the pen was the rattle. Moreover, children's TB performance in this aspectuality task followed an age-related u-shaped curve comparable to the TB performance in location change and unexpected content tasks.

Taken together these findings of incompetence in standard and aspectual TB tasks challenge the assumption of a full-blown explicit ToM. The PAR account by Fabricius et al. (2010) calls into question whether the processes children use are related to belief reasoning at all. The findings of Perner and colleagues (Perner et al., 2015), in contrast, call into question whether the belief representation competence children acquire around age four is really full-blown.

Despite the empirical underpinnings of both positions, showing very similar u-shaped developmental patterns, both theories are restricted in their scope of applicability. The PAR account can only explain u-shaped findings in location change and unexpected content tasks, while the MFCT only applies to TBs involving aspectuality. Hence, these theories only offer local explanations of children's u-shaped performances in the different tasks.

In contrast to this limitation of the presented competence accounts, a performance-based account could explain the pattern in both fields. If these findings can be explained performance-based, TB-based overestimation attacks against the standard picture of ToM development can be fend off. The basic idea is similar to what Rakoczy et al. (2015) suggest for aspectuality FB tasks. TB tasks are not more complicated that standard FB tasks. I suggest that once TB tasks are suitably modified children are able to solve them.

There are several candidate factors that may explain children's poor performance on TB tasks. One prominent possibility is that TB tasks are pragmatically confusing due to extraneous factors of the task structure or their format. Pragmatic factors have been shown to play a performance limiting role in different developmental fields. Many classical Piagetian pre-operational failures seem to originate in a lack of understanding the test question and related pragmatics rather than real competence deficits (Siegal & Beattie, 1991). In a similar vein, nativist accounts used pragmatic deficits in order to explain younger children's failure in explicit FB tasks (for the most recent work along such lines, see Helming, Strickland, & Jacob, 2014; Westra, 2016; Westra & Carruthers, 2016).

One possible pragmatic-factors account reads as follows. When children are asked (i) trivial questions (ii) about beliefs (iii) without any obvious reason that explains the question's the triviality, they get confused. This pattern may even be correlated to ToM competence. The older and therefore more competent I get in ToM, the more I wonder why the experimenter is asking me such trivial questions (i) about the protagonist's the mental states (ii). If I do not have an extraneous explanation for that (iii), I conclude that there must be something significant about the protagonist's belief. The only way in which this belief can be significant is by being false. Therefore, I assume that the protagonist has a FB. This process can be intensified by children's awareness of the pragmatic fact that the main point of belief talk is to refer or at least highlight the possibility of their falsity (Papafragou, Cassidy, & Gleitman, 2007).

Another possible pragmatic factor related to the question's triviality may be the usage of a test question. Children are able to differentiate between genuine and test questions from a very early age on (Grosse & Tomasello, 2012). Therefore, the usage of a trivial question as a test-question may mislead children even more. It is clear that the experimenter is not asking about the protagonist's mental states in order to gain new information. The experimenter even knows exactly what is going on. Therefore, her usage of a test question may be interpreted as highlighting the significance of the protagonist's belief. And again the only way for the belief to be significant is by being false.

Furthermore, a non-pragmatic performance factor may be related to the salience of the protagonist’s belief. At the beginning of FB scenarios, the experimenter, the child and the protagonist have a common basis of shared knowledge. At one point of time, namely when the location of the object is changed in the protagonist’s absence or the dual identity is revealed in his absence, the protagonist’s (false) belief becomes salient. The child now has a reason to ascribe a belief to the agent. In TB scenarios, however, there is no point in time at which the protagonist’s belief becomes salient until the child is explicitly asked for the belief. This may irritate children and hinder them from showing their real competence.

Taken together the presented performance-based accounts make clear and testable predictions. These predictions can be used to differentiate between competence- and performance-based explanations of the phenomenon. Competence-based accounts do not predict any effect of the suggested extraneous factors. If studies manipulating these factors show an improvement of children’s TB competence, this clearly speaks in favour of performance accounts. This then would clearly support the classical picture of ToM competence by showing that the empirical underpinning of the overestimation claim is based on performance problems.

4.3. Summary

In this chapter I have presented two lines of research that have been the root of the overestimation claim. For both, aspectuality understanding and TB ascription, there is a notable amount of findings of incompetence in children older than four years of age. These findings have caused a debate on whether the standard picture of ToM development is overestimating the real competence children acquire around age four. In contrast to the full-blown (meta-) representational conception of ToM, these studies depict the explicit ToM competence as limited. Moreover, they suggest an age-related development different from the four-year-revolution. Children do acquire basic ToM competences around age four as shown by their performance in classical FB tasks. However, they are only able to fully understand beliefs and their fundamental qualities around age six (like the aspectuality of beliefs and the fact that they can not only be misrepresentations, but also sometimes true). One theory goes even beyond this limitation, by claiming that what children acquire around age four is not belief reasoning at all (PAR by Fabricius and colleagues (2010)) but more likely the usage of a heuristic that produces patterns that are falsely interpreted as indicating true ToM competences. In contrast to such competence accounts, performance accounts offer an alternative interpretation that is compatible with the standard picture of ToM development. From a performance-based point of view, findings of limitations are false negatives. Children's real competence is masked by different extraneous task demands, which are not necessarily related to belief ascription competence. In the case of aspectuality understanding there is already one study showing that reduced task demands have a positive effect on children's performance (Rakoczy et al., 2014). In a similar vein, I suggested that once task demands are modified in TB testing procedures, children will reveal their real competence.

5. Aim of Dissertation

The aim of this dissertation is to investigate the validity of the overestimation claim which states that assuming that children would acquire a full-blown ToM around age four overestimates their real competence given at that age.

That children at that age fail to handle the aspectuality of beliefs has been one root of overestimation claims. Therefore, the first study of this dissertation focusses on 4- to 6-year olds' ability to understand the aspectuality of beliefs. While different studies have shown that this age group struggles with these tasks, Rakoczy et al. (2015) have already demonstrated that children are able to solve aspectuality tasks once they are suitable modified. My first study aims to investigate children's aspectuality understanding in a novel, comprehensive and stringent design. Therefore, structurally analogous aspectual and non-aspectual FB tasks were designed that are equivalent in their task demands. My study uses one object, which is represented under different identities that are not connected to its appearance. This is different from the dual-objects used by Rakoczy et al. (2015). Additionally, given that aspectuality plays a role in the scope of both- explicit and early implicit ToM competences- this design also has the potential to be easily transferred into a helping paradigm in order to investigate implicit aspectuality understanding in younger children. There is already one study using a helping paradigm version of this task (Schulz, Oktay-Guer, & Rakoczy, 2016).

The second and third study of this dissertation focus on the TB competence of children older than four years of age. At the end of the last chapter I presented competence- and performance-based explanations for children's TB problems. As a first step, Study 2 aims to clear which of the two general possibilities (competence vs performance) is correct. The results of Study 2 clearly support a performance based explanation. Study 3 goes beyond this by further investigating the factors that cause children's poor TB performance.

The superior aim of this dissertation is to integrate critical findings of limitations in explicit ToM competences into the standard picture of ToM development, namely that children acquire a full-blown ToM around age four.

6. Study 1

6.1. Introduction

The existing research on children's aspectuality understanding has revealed conflicting results. On the one hand, previous studies suggest that understanding aspectuality emerges later than general FB competence and therefore questions the full-blown conception of the ToM competence acquired around age four. On the other hand, one recent study by Rakoczy and colleagues (2015) suggests that children are able to solve aspectuality tasks as soon as they can solve classical FB tasks. In order to shed light on this conflicting picture, the aim of this study is to further investigate children's aspectuality understanding in a novel, comprehensive design. Therefore, I designed structurally analogous aspectual and non-aspectual tasks, in which the protagonist at some point arrives at a FB about the number of objects in a box. In the non-aspectual version, this belief occurs because the protagonist missed the change of location of one object. In the aspectual version, the protagonist arrives at the same FB but in a different way. She fails to witness the object's transfer as the transfer of this very object. While she in reality sees the same object being transferred twice, she is not aware of the fact that the second object she saw being transferred was exactly the same she saw being transferred before. In Experiment 1 children are tested on a classical location change task, the newly designed aspectuality task and a parallel numerical location change task. In the aspectuality and the numerical location change task, the test questions refer to the agent's numerical belief ("How many object does he belief to be in that box?"). Experiment 2 tests even more directly for aspectuality understanding by directly asking children for the procedure's critical facet procedure making the task aspectual.

6.2. Experiment 1

6.2.1. Method

6.2.1.1. Participants. Fifty 3- to 6- year-olds (twelve 3-year olds, seventeen 4-year-olds, eighteen 5-year-olds and three 6-year-olds; range: 38-72 months; M=57, SD=9, 7; 22 female) from mixed socioeconomic backgrounds were included in the final sample. Three further children were tested but excluded from the analysis because they were uncooperative. Participants were recruited from a database of children whose parents had previously given permission to experiment participation. Children were tested by a female experimenter (E) either in an appropriate room in their daily childcare or in the lab.

6.2.1.2. Design and Procedure. In a within-subjects design, children were tested in three different tasks (two trials of each). Task order and the sides of the boxes was counterbalanced (see Appendix A for details).

6.2.1.2.1. Verbal Ability. Verbal ability (for use as a covariate in control analyses) was measured at the beginning of the session with the vocabulary subscale of the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1999; Kaufman, A., & Kaufman, N.).

6.2.1.2.2. Standard Location Change Task (SLT). Each child was tested in two trials of a standard location change FB task (after Wimmer & Perner, 1983). The child and the protagonist, a puppet, were shown two boxes and an object. The object was hidden in one of two boxes (box 1) and the puppet left the scene. The experimenter suggested to play a trick[3] on the protagonist who was absent and transferred the object to the other box (box 2). When the protagonist returned, control questions (CQ1: "Where did we put the [object] in the beginning?" and CQ2: "Where is it now?") and the test question (TQ: "When the puppet wants the [object], where will he look for it?" [Correct answer:" box 1"]) were asked. The order of the objects and the location of box 1 were counterbalanced.

6.2.1.2.3. Aspectuality Task (AT). In addition, children were tested in a new aspectuality task where the identity of an object caused a FB of the protagonist about the number of objects hidden in a box (see Fig. 1; for details see Appendix A). The child and the puppet were presented with two boxes per trial. One of the boxes (box 1) was empty, the other one (box 2) contained a multitude of qualitatively identical objects (e.g. blue toy blocks). The child was asked to take two exemplars of the object out of box 2 for a game the child, the experimenter and the puppet were going to play together.

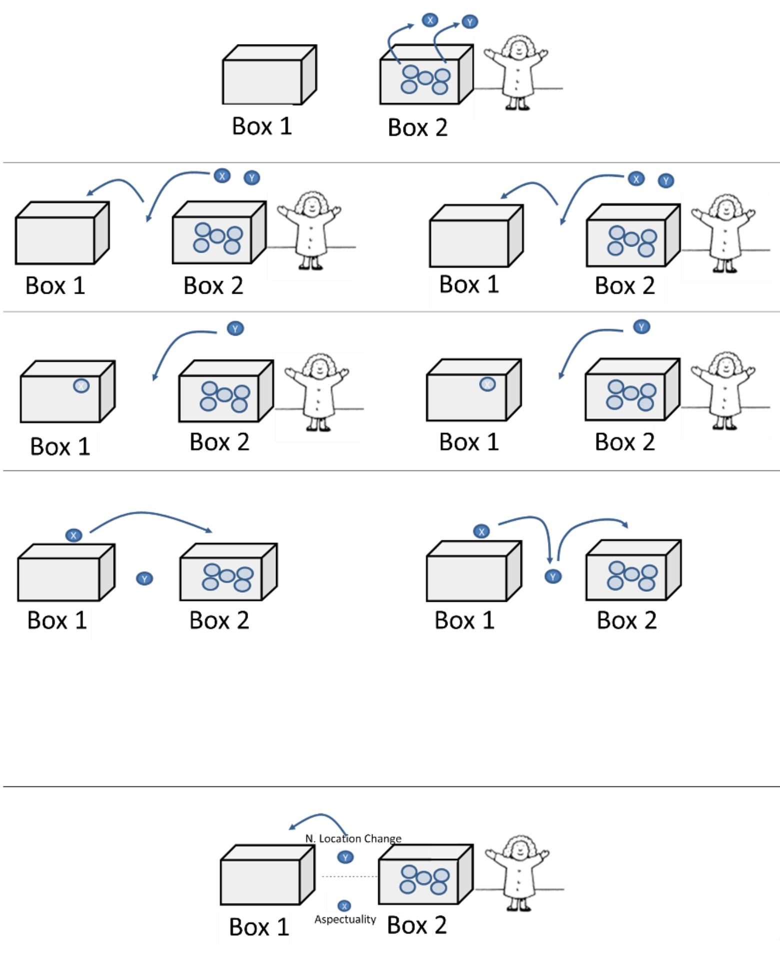
Introducing the game's rules. The experimenter explained that the object was first put in the middle (between the two boxes) and then in box 1. After doing so, the child was asked how many objects were in the box. When the child gave the right

[3] This was implemented since acting out the transfer of the object in change-of-location false belief tasks in ostensive deceptive ways has been shown to be helpful to younger children in some studies (Wellman et al., 2001).

answer [“one”], the game continued and E put the second object in the middle and then in box 1. The experimenter again asked how many objects were in box 1 [correct answer: “two”]. If the child gave any incorrect answer at any time, E opened box 1 and allowed the child to count the content. After the child answered the questions correctly, both objects were taken out of box 1 and the game started again. When the child gave correct answers to all questions in a run, the test trial began.

Critical test trial. The test trial started as described above, but before E could ask for the number of objects in the box, the puppet announced she had to leave. E put the second object in the middle, the protagonist left and they waited for the protagonist to come back. In the protagonist’s absence, E suggested the following trick. The object from the middle was put back to box 2. The object from box 1 was taken out and replaced the former object from the middle. The first control question (“Does the [protagonist] know that we took the object from the middle and put it back in box 2 and took the one out of box 1 and placed it in the middle?” (Q1) (Right answer: “no”) was asked. Upon the puppet’s return, this manipulation resulted in a FB of the protagonist about the identity of the object in the middle (protagonist thinks the object she sees in the middle is different from the one she saw being put in the box before she left). Then the game continued and the very same object as in the first scene was put in box 1 again. Now E asked the remaining control questions (“Does the puppet know that we exchanged the objects when he was absent? (Q2; repetition of Q1) [correct answer: “no”]” and “How many objects are in box 1? [correct answer: “1”]” (Q3)) and the test question (“How many objects does the [protagonist] think are in the box?” (TQ) [correct answer:”2”]). Children were asked again (at most twice) and corrected if they answered any control question incorrectly.

6.2.1.2.4. Numerical Location Change Task (NLT). This task was designed to test for children’s performance in a structurally analogous non-aspectual FB task. This task was closely matched in terms of complexity and task demands to the AT but differed in the one crucial respect (see Figure 2, for details see Appendix A). In both tasks, the protagonist arrived at a FB about the number of objects in a box, but in order to understand how this belief came about, the AT did, whereas the NLT did not require an understanding of the aspectuality of the protagonist’s belief. In the crucial step four in this task, in the absence of the protagonist, the experimenter removed the object from box 1, but did not swap it for the other object in the middle, but rather put it directly into box 2. There was thus no FB on the part of the protagonist upon her return concerning the identity of the object in the middle, but simple a FB about the location of the object formerly in box 2 and thus a false numerical belief about the content of box 1.



Conditions		Remarks and Test Questions	
Numerical Location Change	Aspectuality	Experiment 1	Experiment 2
		E1 shows that box1 is empty and box 2 contains indistinguishable objects	E1 shows that box1 is empty and the box 2 contains indistinguishable objects
			Object X is assigned to child and Y to Protagonist; Protagonist present
			Test Question 1 "What does the puppet think… - … where your object is now?" (Location Change) / -… whose object is the one in the middle?" (Aspectuality)
		Test Question: "What does the puppet think? How many balls are in this box (box 2)?"	Test Question 2: "What does the puppet think? How many balls are in this box (box 2)?"

* Note that box 1 contains collection of qualitatively identical objects that cannot be perceptually distinguished; "X" and "Y" are only marked here for the reader but were perceptually indistinguishable to children and the protagonist.

* Note that box 1 contains collection of qualitatively identical objects that cannot be perceptually distinguished; "X" and "Y" are only marked here for the reader but were perceptually indistinguishable to children and the protagonist.

Figure 2. Procedures of the Aspectuality and Numerical Location Change tasks used in Study 1.

6.2.2. Results

6.2.2.1. Control Questions. Children answered 89% of all control questions correctly and about 74% of the children (N=37; 33% of the 3-year olds (n=4), 77% of the 4-year olds (n=13), 94% of the 5-year olds (n=17) and all 6-year olds (n=3)) answered all control questions correctly on the first request. Table 1 depicts the percentages of children solving the different types of control questions.

Table 1
Percentage of correctly answered control questions in Experiment 1 in Study 1.

	Standard Location Change Task		**Aspectuality Task**			**Numerical Location Change Task**		
# trials correct	CQ1: Location 1	CQ2: Location 2	CQ1: Knowledge 1	CQ2: Knowledge 2	CQ3: Real Number	CQ1: Knowledge 1	CQ2: Knowledge 2	CQ3: Real Number
2	92% (N=46)	100% (N=50)	82% (N=41)	90% (N=45)	98% (N=49)	84% (N=42)	86% (N=43)	100% (N=50)
1	2% (N=1)	-	10% (N=5)	4% (N=2)	2% (N=1)	8% (N=4)	10% (N=5)	-
0	6% (N=3)	-	4% (N=2)	6% (N=3)	-	8% (N=4)	4% (N=2)	-

6.2.2.2. Main analysis (whole sample). The consistency in performance of children over trials 1 and 2 of each task were high (Φ=.87 in the Location Change task; Φ=.85 in the Aspectuality Task and Φ=.60 in the Numerical Control Task). Therefore, sum scores of trials solved correctly per task [0-2] were computed for further analyses. The mean values of these sum scores in the different tasks are depicted in Figure 3. First, in order to test whether the tasks differed in difficulty, a univariate ANOVA with task as factor was conducted but did not reveal any effect, (F(2,98) = .34 , p = .71).

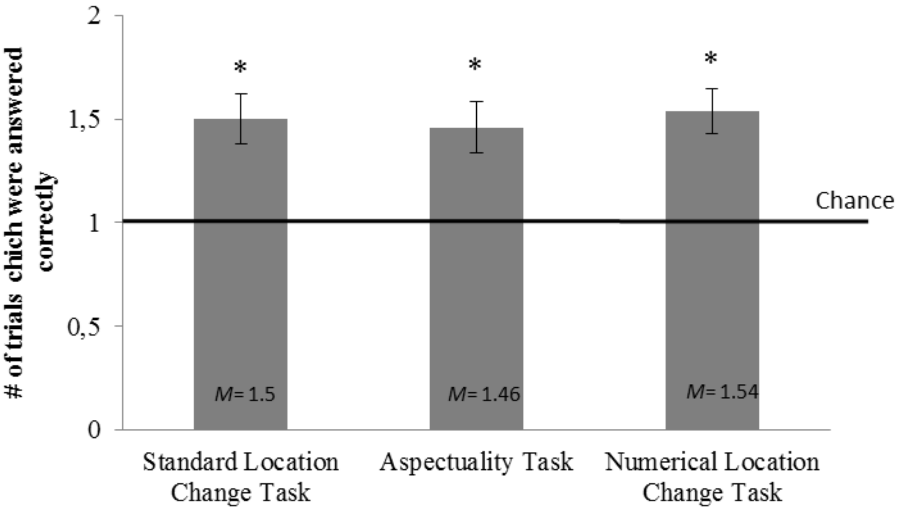


Figure 3. Mean number of trials answered correctly as a function of task in Experiment 1 in Study 1 (* p < .05).

Second, comparisons against chance performance showed that children gave the correct answer significantly more often than expected by chance in all tasks (Standard Location Change Task, *t*(49) = 4.21 , *p* < .001, *d* = .60; Aspectuality Task, *t*(49) = 3.78, *p* < .001, *d* = .53 and Numerical Location Change Task, *t*(49)= 5.07, *p* < .001, *d* = .71). Third, raw and partial correlations (correcting for age and verbal ability) of the sum scores between the different tasks were computed and showed that the two new tasks were strongly related to each other and to the standard FB task (see Table 2) (for supplementary control analyses that take into account performance in control questions, see Appendix B).

Table 2
Correlations (and partial correlations correcting for age and language ability) between the different tasks in Study 1.

	Aspectuality Task	Numerical Location Change Task
Standard Location Change Task	.58** (.40*)	.50** (.25+)
Aspectuality Task		.88** (.85**)

+*p* < .10; * *p* < .01; ** *p* < .000

6.2.3. Discussion

Experiment 1 had three main results. First, children performed above chance in both the Aspectuality as well as the Numerical Location Change Task. Second, all tasks (AT, NLT, SLT) were equally difficult. Third, children's performances on the different tasks were strongly correlated.

These findings taken together thus seem to speak for the unity and convergence in explicit ToM competence. However, it may be objected, perhaps this does not show that children can solve aspectual and non-aspectual FB tasks in analogous ways. It may rather suggest that they solved the supposedly aspectual FB tasks in ways that did not require an understanding of aspectuality after all. Adults would typically solve the aspectual FB task in the following way. They would reason about how the protagonist had perceived the objects, appreciating that she had seen what was in fact the very same object under different aspects at different times ("this [object]" at time 1, and "another [object]" at time 2) and thus arrived at a false numerical belief ("there are two different balls in the box"). However, perhaps children here arrived at the correct solution in much simpler ways.

One particular sceptical concern along such lines is the following: children may not have paid attention to the object's identity at all. Rather they may have simply kept track of the number of events of putting objects in box 1 and removing them from there that the protagonist witnessed, engaging in some kind of belief-bookkeeping (along the following lines: he witnessed "+ 1", he did not witness the "-1", but then did witness the second "+1" again, therefore his numerical belief is "+2").

Experiment 2, therefore, was designed to address this concern. If children solve the Aspectuality Task in such simpler, non-aspectual ways, then they should be unable to explicitly ascribe to the protagonist beliefs about the numerical identity of the object in the middle. If, however, they do solve the task in aspectual ways, they should be able to ascribe such beliefs, and their ascription of such beliefs and their general performance in the task should strongly converge and correlate.

6.3. Experiment 2

In this Experiment, therefore, the same closely matched aspectual (AT) and non-aspectual (NLT) numerical belief ascription tasks as in Experiment 1 were used. With one crucial modification. In addition to the test question concerning the protagonist's numerical belief at the end, another test question concerning the protagonist's belief about identity/location of the object in the middle was added beforehand (see Fig. 3).

6.3.1. Method

6.3.1.1. Participants. Thirty-four 4- to 6- year-olds (range: 51-80 months; *M* = 62; 17 female) from mixed socioeconomic backgrounds were tested. Two additional children were tested but excluded from data analysis because it turned out that they could not reliably count up to 2 (N = 1) or they were uncooperative (N = 1). Participants were recruited from a database of children whose parents had previously given permission to experiment participation. No child from Experiment 1 participated in Experiment 2. Children were tested by a female experimenter in an appropriate room in their daily childcare or in the lab.

6.3.1.2. Design and Procedure. In a within-subjects design, children were tested in the Aspectuality and the Numerical Location Change tasks and received two trials of each task. Task order as well as the sides of the relevant boxes were counterbalanced across subjects.

6.3.1.2.1. Verbal Ability. Children completed a vocabulary test (subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999) at the beginning of the session.

6.3.1.2.2. Aspectuality Task (AT). This AT was the very same as in the first experiment with the following modifications (see Fig. 1 and Appendix A).

(1) The objects the game was played with were assigned to the child and the protagonist.

(2) The child always started by putting her object in the middle and then in box 1.

(3) The protagonist left the scene after placing her object in the middle, in her absence her object was put in box 1 and it was replaced by the child's object from box 1.

(4) Still in the absence of the protagonist the child was asked the first control question (CQ1: "Whose object is the one in the middle?" [Right answer: "the child's"])

(5) Upon the protagonist's return, the first test question (*identity questio*n) was asked (T1: "What does the puppet think whose object is the one in the middle?" [Right

answer: “the protagonist’s”]) and the game went on, with the protagonist moving the object from the middle to box 1.

(6) Finally, the same control questions as in Experiment 1 and the numerical test question were asked.

6.3.1.2.3. Numerical Location Change Task (NLT). This task was a modification of the task used in Experiment 1, with control and test questions equivalent to the ones used in the AT in Experiment 2 (see Fig. 1 and see Appendix A). In the protagonist’s absence, the child’s object from box 1 was moved to box 2 and the first control question was asked. Upon the protagonist’s return the first so-called *location test question* (TQ1: “What does the puppet think where your object is?” [Right answer: “in box 1”]) was asked and the game went on analogous to the AT.

Children were directly corrected if they answered the second control question (CQ2), asking for the protagonist’s knowledge about the manipulation, incorrectly. Each child received two trials per task. Task order and side of boxes was counterbalanced between subjects.

6.3.2. Results

6.3.2.1. Control questions. Table 3 depicts the percentages of children solving the different kinds of different control questions on the first trial. Overall, children answered 67% of the control questions correctly, with about 56 % of the children (N = 19) consistently answering all control questions correctly. As can be seen from Table 3, most incorrect responses pertained to CQ2, with 94% of the children consistently answering the first and the third control question correctly in all of the trials.

Table 3
Control questions answered correctly (%) in Experiment 2 in Study 1.

	Aspectuality Task			Numerical Location Change Task		
# trials correct	CQ1: Whose object?	CQ2: Knowledge	CQ3: Real Number	CQ1: Where is object?	CQ2: Knowledge	CQ3: Real Number
2	100% (N=34)	62% (N=21)	94% (N=32)	97% (N=33)	62% (N=21)	94% (N=32)
1	-	15% (N=5)	6% (N=2)	3% (N=1)	23% (N=8)	6% (N=2)
0	-	23% (N=8)	-	-	15% (N=5)	-

6.3.2.2. Main analyses. The consistency in performance of children over trials 1 and 2 of each test question was high (AT identity question Φ = .81 and number question Φ = .82; NLT location question Φ = .77 and number question Φ = .90). Therefore, trials 1 and 2 per test questions were combined to yield sum scores [0-2]. In addition, within each trial I computed an *aggregate score* that took into account whether children solved both the identity/location and the number question. A given trial received the *aggregate score* “correct” only if children answered both questions correctly (with a chance level of guessing correctly of 1/4). The mean sum scores for the different tests questions as well as the mean sum of aggregate scores across trials 1 and 2 of a given type of task are depicted in Figure 4 as a function of conditions.

First, in order to test whether there were differences between tasks or test questions, a 2 (task: AT vs. NLT) x 2 (question: identity/location vs. number) ANOVA was conducted on the mean sum of correct trials. This analysis yielded no main effect of task (AT vs. NLT, *F*(1,33) = 0, *p* = 1), a main effect of test questions (such that the number question was easier than the identity/location question, (*F*(1,33) = 5.38, *p* < .05)), and no interaction effect (*F*(1,33) = 1.00, *p* = .33) between the factors.

Second, comparisons against chance performance showed that children gave the correct answer significantly more often than expected by chance in all tasks and test questions (AT identity question, *t*(33) = 2.51, *p* < .05, *d* = .43; and number question, *t*(33) = 5.14 , *p* < .001, *d* = .88; NLT location question, *t*(33) = 3.53 , *p* < .01, *d*= .61 and number question, *t*(33) = 5.14, *p* < .001, *d* = .88). With regard to the aggregate score, children’s performance was also significantly different from chance in the AT (*t*(33) = 5.11, *p* < .001, *d* = .88) and NLT (*t*(33) = 3.44, *p* < .01, *d* = .60).

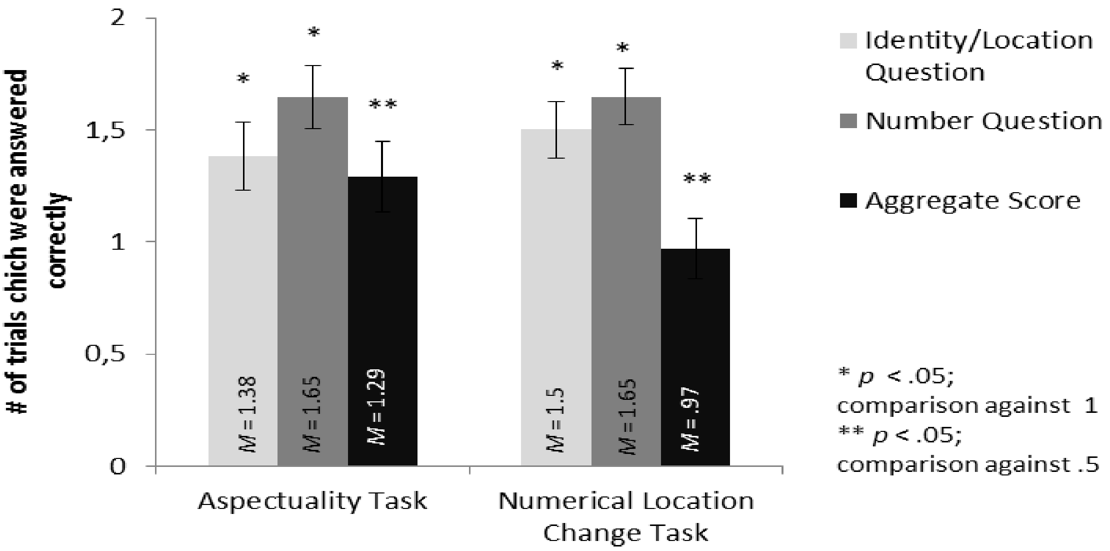


Figure 4. Mean number of trials answered correctly as a function of task and question type in Experiment 2 in Study 1.

Third, in order to analyse convergence in performance, correlations between the different test questions within a task and between tasks were computed. Performance on the different test questions within a task was highly correlated both for the AT (Identity and Number Question, *r* = .68, *p* < .001; partial correlation, controlling for age and verbal ability, *r* = .61, *p* < .001) and for the NLT (Location and Number Question, *r* = .60, *p* < .001; partial correlation, controlling for age and verbal ability, *r* = .37, *p* < .05). Performance on a given test question, and on both test questions per trial combined, also correlated substantially across the different tasks (see Table 4; for supplementary control analyses that take into account performance in control questions, see Appendix C).

Table 4
Correlations (and partial correlations correcting for age and language ability) of performance in a given questions type and the aggregate scores between Location Change and Aspectuality tasks in Experiment 2 in Study 1.

Identity/ Location Questions	Number Questions	Aggregate Scores
.60**	.89**	.60**
(.51*)	(.86**)	(.53*)

*p<.01, **p<.001

6.3.3. Discussion

Experiment 2 replicated the main findings of Experiment 1. AT and the NLT did not differ in difficulty, children performed competently in both, and performance was strongly correlated across tasks. However, Experiment 2 also extended the results of Experiment 1 in crucial ways. Children's performance showed convergence and unity even with explicitly aspectual questions. This clearly speaks against more parsimonious strategies of solving the AT without understanding aspectuality. Taken together Experiment 1 and 2 thus supply converging evidence for unity and convergence in performance across various explicit ToM tasks.

7. Study 2[4]

7.1. Introduction

The aim of this set of experiments was to investigate the development of children's patterns of TB performance systematically, and to test whether these patterns can be best explained by competence- or by performance-limitation accounts. To do so, I first investigated the development of FB and TB performance in a comprehensive design with different kinds of ToM tasks (standard location change and aspectuality) across a wide age range (from age 3 to adulthood). This was done to see whether TB-performance generally and robustly yields a u-shaped curve while FB performance simply increases with age. Secondly, I derived and tested competing predictions of competence- vs. performance- limitation accounts. Competence limitation accounts predict u-shaped curves in TB tasks only in specific cases under limited circumstances. The Mental File Card Theory, predicts a U-shaped curve for TB performance only in the specific sub-class of aspectual TB tasks (but no such pattern for standard change-of-location TB tasks). The other competence-limitation approach, the PAR account, predicts a u-shaped curve only for the sub-class of TB tasks in which the protagonist has "comparable lack of perceptual access" relative to FB tasks (Hedger & Fabricius, 2011, p .432).

Performance limitation accounts in terms of extraneous task factors surrounding TB tasks (such as salience/relevance, and/or pragmatics), in contrast, would predict u-shaped curves in TB tasks to be a much more general phenomenon. First of all, the pattern of TB and FB performance should be analogous over different types of tasks (standard change-of-location and aspectual), including those for which either the PAR account or the MFCT account do not even apply. Second, some performance factor accounts would assume that the sensitivity to the crucial performance factors (that make the TB tasks difficult) depends on ToM (which in turns is tapped in FB tasks); and they would thus predict an inverse relation between FB and TB performance. For both change-of-location and aspectual tasks, children's FB and TB performance should be negatively correlated[5]. Such a prediction follows clearly from pragmatic

[4] Note that the experiments reported in this section are Reprinted from Cogntition, , 166, Oktay-Gür, N. & Rakoczy, H., Children's difficulty with true belief tasks: Competence deficit or performance problem?, 28-41, Copyright (2017), with permission from Elsevier

[5] One important qualification is in order here. Clearly, pragmatic performance factor accounts assume that pragmatically based failure in TB tasks is a transient phenomenon (after all, older children and adults finally do master TB tasks again). Presumably, at some point, children's pragmatic capacities

performance factor accounts. Sensitivity to pragmatics is known to depend developmentally on ToM (e.g. Happé, 1994; Winner & Gardner, 2012), and thus increase in ToM (indicated in FB performance) should go along with increase in pragmatic competence, and thus with pragmatic confusion in TB tasks, and thus in general with decrease in TB performance.

Thirdly, once the critical performance factors have been removed or alleviated, children's difficulty with TB tasks (and the negative correlations between TB and FB) should vanish.

These predictions were tested in 2 experiments against those of the competence-limitation accounts. Experiment 1 investigates the development of performance in standard and aspectual TB and FB tasks from early childhood to adulthood. The results revealed analogous patterns of u-curves in TB in standard change-of-location and aspectual tasks, increase in FB tasks, and negative correlations of FB and TB between ages three and six. In Experiment 2, new FB/TB tasks were devised that removed potential performance factors (such as the pragmatic oddity and the relevance and salience of the agents' beliefs), and children from age 4 now performed competently on both FB and TB trials.

7.2. Experiment 1

7.2.1. Method

7.2.1.1. Participants

171 subjects were included in the final sample (3-year olds[6], 37-41 months, *M* = 39, n = 14; 3.5-year olds, 42-47 months, *M* = 44, n = 26; 4-year olds, 48-59, *M* = 54, n = 26; 5-year olds, 61-70 months, *M* = 66, n = 20; 6-year olds, 72-85 months, *M* = 79, n = 25; 8-year olds, 96 to 107 months, *M* = 102, n = 20; 10-year olds, 122-143, *M* = 127, n = 22; adults, 21-38 years; *M* = 26 years; n = 18). Participants came from mixed

have developed to a higher level at which they now understand why people may engage in trivial and seemingly pointless test questions. At this stage, then, children should be able to apply their belief concept in all kinds of pragmatic situations and thus perform equally competently in FB and TB tasks (with positive correlations between TB and FB). This means that the predicted negative correlation should only be expected in the intermediate period in which children have acquired a concept of belief, are capable of applying it in the FB tasks, and yet are pragmatically still vulnerable in the TB tasks. When exactly this period ends is an empirical question (previous research suggests perhaps around age 6, whereas the current findings point to a much more protracted development; see below).

[6] I included two separate groups of younger and older 3-year-olds since in previous studies and pilot work in our lab, a considerable proportion of older 3-year-olds already passed FB tasks. I thus targeted young 3-year-olds specifically since I wanted to make sure that the youngest age group performs close to floor in FB.

socioeconomic backgrounds and were recruited from a databank of children whose parents had previously given consent to experimental participation (children) or via recruiting in a teaching class (adults). Fourteen additional children were tested but excluded from data analysis because they were uncooperative (n = 2), due to insufficient linguistic abilities (n = 2), or due to experimental error (n = 10). Children were tested by either a male or a female experimenter in their day care or in the lab. Adults were tested in the lab and received chocolate for participation.

7.2.1.2. Design and Procedure. Children's performance was investigated in a 2 (task type: standard vs. aspectuality; between subjects factor) x 2 (belief type: FB/TB; within subjects factor) design. Each participant received two trials per condition, four in total (order of FB/TB tasks counterbalanced across subjects). The procedure in the standard and aspectuality tasks is described below. Children (except for the 10-year olds, for whom the task was not age-appropriate anymore) received the same task of verbal ability (K-ABC) as in Study 1. When children failed to answer the control questions correctly the experimenter repeated the test question. If children insisted on their wrong answer in this experiment they were corrected (but, conservatively, their first answer to the control question was used for further analysis and coded as "incorrect").

7.2.1.2.1. Standard Task. Two trials of standard change-of-location tasks with different stimuli were administered per child, either in TB or in FB versions (Wimmer & Perner, 1983). The protagonist and the child were introduced to an object X [e.g. a plastic duck]. The object was then placed in one of two boxes (box1) before the protagonist left. Either in her absence (FB condition) or after her return (TB condition), the object was moved to the other box (box2) and the following control and test questions were asked:

- Control Question 1: Where did we put the [object] in the beginning? [correct answer: box1]
- Control Question 2: Where is the [the object] now? [correct answer: box2])
- Test question: Where will the protagonist look for the [object]? [correct answer: box 2 (TB)/box 1(FB)]

What is crucial about the TB version of the standard task is that neither the PAR nor the MFCT would predict that it should be difficult. The latter only applies to aspectual TB tasks, and the former does not apply because the protagonist leaves before the crucial events unfold (transfer of the object) and thus has no lack of relevant perceptual access.

7.2.1.2.2. Aspectuality Task. Two trials of aspectual FB/TB tasks with different stimuli (dual-function and dual-identity) were administered per child, either in TB (1 dual function/1 dual identity) or in FB versions (1 dual function/1 dual identity). The basic logic of these tasks (closely modelled after Study 3 of Rakoczy et al., 2015) is depicted in in Figure 5. In the presence of a protagonist an object was put into a box (box 1) under aspect A [e.g. pen]. In the protagonist's presence (TB) or absence (FB) it was revealed that the object had another identity B [e.g. rattle] and it was stored in the same box again. In the presence of the protagonist the object was now transferred to box 2 under its identity B, like in the following example: The experimenter covered the object with her hands while taking it out of its initial box, rattled with it and then moved it to the other box such that the A-identity (pen) remained invisible throughout and only the B-identity (rattle) could be heard. In both belief conditions (TB and FB) the protagonist witnessed the object's transfer. The critical difference between the conditions was that in FB the protagonist did not know that the objects she saw at different time points as A and B were identical. The following control and test questions were asked.

- Control Question 1: Does the protagonist know that the A (e.g. pen) is the B (e.g. rattle)? [correct answer: yes (TB)/no(FB)]
- Control Question 2: Where did we put the A [e.g. pen] in the beginning? [correct answer: box1]
- Control Question 3: Where is the A now? [correct answer: box2]
- Test question: Where will the protagonist look for the A? [correct answer: box 2 (TB)/box 1(FB)]

When children failed to answer the control questions correctly, the experimenter repeated the question. If children insisted on their wrong answer they were not corrected.

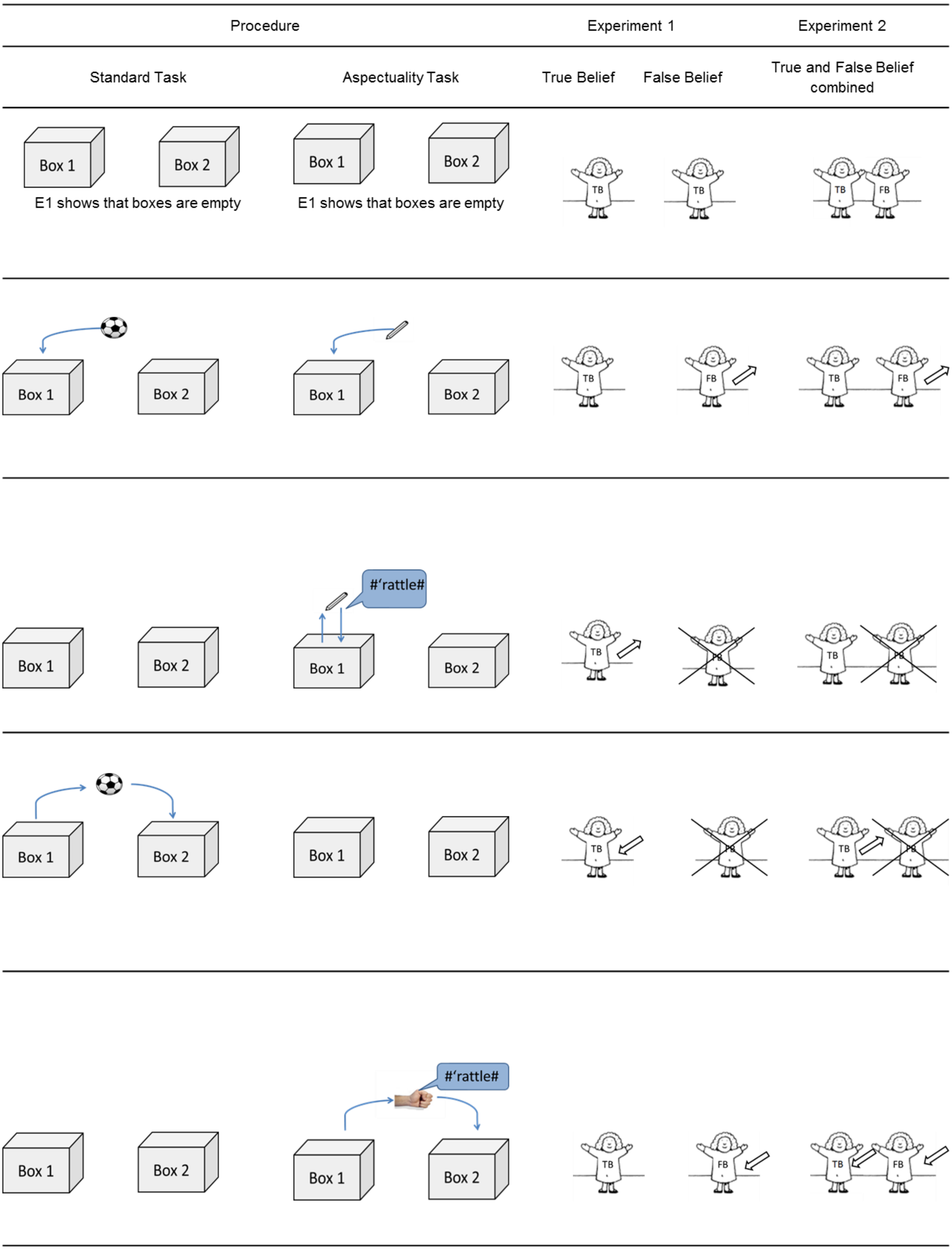


Figure 5. Procedure of the different tasks used in the experiments of Study 2.

Concerning the TB version of the aspectuality task, only one competence-limitation account, the MFCT, predicts that it should be difficult, whereas the PAR account does not even apply – again, because the protagonist leaves before the crucial events and thus has no lack of relevant perceptual access.

7.2.2. Results

Children answered control questions correctly in the following percentages of given trials. Standard FB/TB: Control question 1: 93 % correct; Control question 2: 99 %; Aspectuality FB/TB: Control question 1: 82 %; Control question 2: 95 %; Control question 3: 99 %. Overall, 81% (n = 118) children answered all control questions correctly while all adults answered all control questions correctly. A closer analysis of control question performance as a function of age revealed the following patterns: 3-year olds (N = 14) performed moderately on Standard FB/TB control questions (success in at least 61% of the trials), competently on Aspectual TB control questions (success in more than 90% of the trials) but poorly on control question 1 in Aspectuality FB (10% correct) while performing moderately on control question 2 and 3 in Aspectuality FB (success in at least 70% of the trials). 3.5 year olds (N = 26) solved control questions in at least 86% of the trials, except for control question 1 in Aspectuality TB (56% correct). 4-year-olds (N = 26) also performed worst on Aspectuality TB control question 1 (67% correct), while they solved all other control questions in at least 89% of the trials. 5-year-olds (N = 20) showed a similar pattern, solving all but Control question 1 in Aspectuality TB (45%) in at least 95% of the trials. All other age groups (5-, 6-, 8- and 10-year-olds) solved all control questions in at least 90% of the trials.

7.2.2.1. Consistencies across trials. The consistency in performance of children over trials 1 and 2 of the same type of task was very high for all conditions (Φs >.48). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

7.2.2.2 Performance as a function of condition. The mean sum of trials answered correctly as a function of conditions is depicted in Figure 6. As it can be seen from the figure, both for standard change-of-location and for aspectuality tasks, the development of TB performance marks a clear U-shaped curve whereas FB performance shows increase with age. Since adults performed at ceiling with no variance whatsoever, they serve as a validation or reference group but cannot be entered into any inference-statistical analyses. These analyses thus focus on the remaining seven age groups. Since preliminary analyses (a 2 (belief: FB/TB) x 2 (task: standard/aspectuality) x 7 (age groups) x 2 (order) ANOVA on the mean sum of trials correct) failed to find any main or interaction effects for the order (FB-TB vs. TB-FB) of test blocks (all *p*s > .18), this factor was skipped from further analyses.

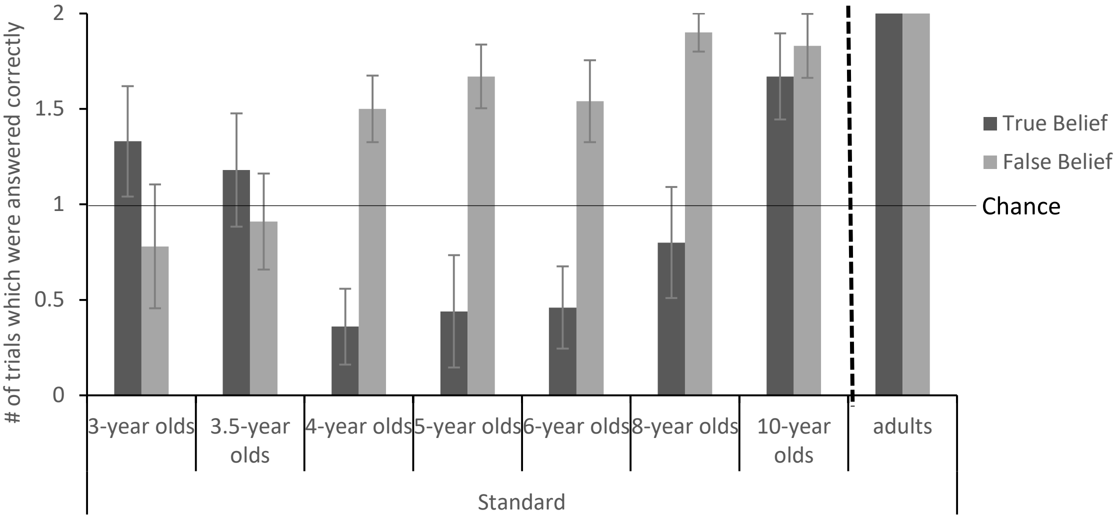
A 2 (FB/TB) x 2 (standard/aspectuality) x 7 (age groups: 3-/3.5-/4-/5-/6-/8- and 10-year-olds) ANOVA on the mean sum of trials answered correctly yielded a main effect of belief type ($F(1,139) = 15.96$, $p< .001$, $\eta p^2 = .10$), a main effect of age ($F(6,139) = 11.07$, $p < .001$, $\eta p^2 =.32$) and no effect of task type (standard/aspectuality) ($F(1,139) = .10$, $p = .74$). Crucially, there was an interaction effect of belief type (FB/TB) and age ($F(6,139) = 7.02$, $p < .001$, $\eta p^2 = .23$) and no other interaction effect.

7.2.2.3 Performance as a function of age. To test for a potential age related development I conducted age-related regression analyses for FB and TB. These analyses revealed that children's performance increased with age and that the age-related FB development is best fitted by a linear model ($F(1,77) = 15.00$, $p < .01$). In TB, in contrast, children's performance followed a U-shaped curve and age-related development was best fitted by a quadratic model ($F(2,77) = 10.09$, $p < .01$).

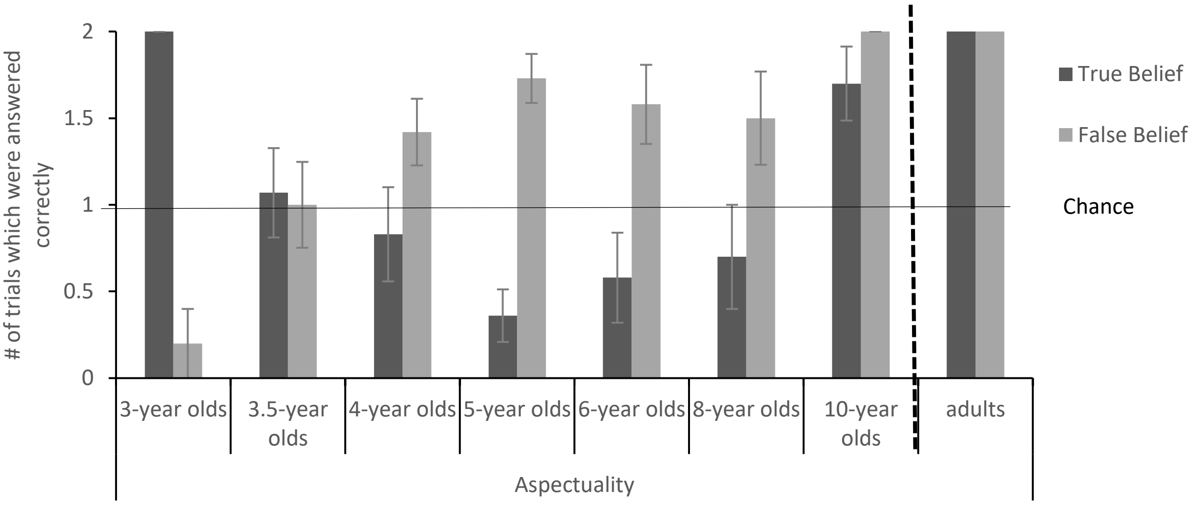
To test for children's performance in FB and TB as function of age in more fine-grained ways, post-hoc follow-up tests against chance in FB and TB tasks were computed separately for the different age groups. These analyses yielded the following results. For FB tasks, only 3-and 3.5-year olds did not perform above chance (3-year olds, $t(13) = -1.88$, $p = .08$; 3.5-year olds, $t(25) = 0$, $p = 1$), while all other age groups did so (4-year olds, $t(25) = 3.64$, $p < .01$, $d = .71$; 5-year olds, $t(19) = 6.66$, $p < .01$, $d = 1.49$; 6-year olds, $t(24) = 3.65$, $p < .01$, $d = .73$, 8-year olds, $t(19) = 4.77$, $p < .001$, $d = 1.07$ and 10-year olds, $t(22) = 10.00$, $p < .001$, $d = 2.13$).

TB performance, in contrast, revealed a rather different (u-shaped) pattern. 3- and 10-year olds performed significantly above chance (3-year olds, $t(13) = 2.83$, $p < .05$, $d = .61$; 10-year olds, $t(21) = 4.47$, $p < .001$, $d = .95$), 3.5- and 8-year olds at chance (3.5-year olds, $t(25) = .40$, $p = .69$ and 8-year olds, $t(19) = -1.23$, $p = .23$), and 4-, 5- and 6-year olds performed below chance (4-year-olds, $t(25) = -2.52$, $p < .05$, $d = -.50$; 5-year olds, $t(19) = -3.94$, $p <.01$, $d = -.88$ and 6-year-olds, $t(24) = -2.92$, $p < .01$, $d = -.58$).

7.2.2.4 Correlations between tasks. Across all age groups, TB and FB tasks were negatively correlated – both in terms of raw and partial correlations (see Table 5). Separate analyses as a function of age groups suggests that these correlations were mainly driven by the 3- to 6-year-olds (with the exception of the 5-year olds).



(a)



(b)

Figure 6. Mean number of trials answered correctly in (a) the standard and (b) the aspectuality FB/TB tasks as a function of age group in Experiment 1 in Study 2.

Table 5

Correlations (and partial correlations correcting for age and language ability) between TB and FB overall and as a function of age group and task type in Experiment 1 in Study 2.

	Correlations TB – FB		
	overall	Standard	Aspectuality
All children	- .42 ** (-.54)**	-.40 ** (-.50)**	-.45 ** (-.58)**
3-year olds	-.55* (-.89)*	-.50 (-.80)*	n/c
3.5-year olds	-.82** (-.81)**	-.71* (-.77)*	-.89** (-.87)**
4-year olds	-.43** (-.34)	-.71* (-.58)	-.17 (-.04)
5-year olds	-.10 (-.11)	.19 (-.21)	.04 (.00)
6-year olds	-.74** (-.70)**	-.72* (-.69)*	-.78* (-.65)*
8-year olds	.04 (.06)	.31 (.06)	-.07 (-.07)
10-year olds	-.10 (-.06)^{x}	-.14 (-.39)^{x}	n/c

* *p*<.05; ** *p*<.001; n/c- not computable due to at least one constant variable; ^{x}- only controlled for age

7.2.3. Discussion

The main findings of Experiment 1 were the following. First, children performed on comparable levels, on standard change-of-location and aspectuality FB tasks, and the same was true for the two types of TB tasks. Second, TB performance (both for standard and for aspectuality tasks) followed a u-shaped curve such that 3-year-olds and children from age 10 performed competently, with children in between failing. In FB tasks (both for standard and for aspectuality tasks), in contrast, performance increased with age such that children younger than four failed while children from four on passed. Third, FB and TB performance was negatively correlated until the age of eight to ten (when children began to master both types of tasks).

These results are very much in line with the predictions of performance-limitations accounts (and not readily explainable by either of the two competence-limitation accounts). Taken by themselves, however, they remain somewhat indirect. More direct evidence would be desirable from experiments that manipulate the alleged performance factors, showing that children's failure in TB tasks (and the negative TB-FB correlations) can be alleviated once the relevant task demands have been removed. Experiment 2 was designed to test for such evidence.

7.3. Experiment 2

The rationale of Experiment 2 was to test for children's TB and FB performance in novel tasks in which the TB versions are less affected by potential performance factors. One prime candidate for the unnecessary complexity of the TB tasks used previously and in Experiment 1, is the lack of relevance or salience of the protagonist's TB (nothing belief-relevant happens in these scenarios, so why should one pay attention to or care about the protagonist's epistemic situation?). Another one is pragmatic oddity (why would one ask such trivial test questions about an agent's beliefs and actions if there is no point in talking about beliefs since the possibility of mistake has not even been raised?).

In order to remove or at least reduce these potential performance factors, I devised tasks (both standard change-of-location and aspectuality) with two protagonists one of whom failed to witness a crucial event and thus had a FB while the other one had full perceptual access and thus TBs. The basic idea is that in this context, the contrast between one agent's FB and the other one's TB makes the TB much more salient and relevant. From a pragmatic point of view, asking about the TB of an agent –given the contrast to the other agent's FB and the fact that this other agent brings into play the possibility of mistake and thus a motivation for belief-talk- is now much less trivial and thus confusing (for similar preliminary findings that adding a second protagonist may help to make FB-ascription more salient and relevant, see (Lewis, Hacquard, & Lidz, 2012; Pham, Bonawitz, & Gopnik, 2012).

The underlying reasoning and prediction from the point of view of the performance-limitation account is the following. If children from around age four have the meta-representational capacity to ascribe beliefs (including aspectual beliefs), both true and false, to agents, and if this competence is masked in some TB tasks by pragmatic or other performance factors, then removing these factors (by making the tasks less pragmatically confusing, more relevant etc.) should have a positive effect. Children from around age four should now master different versions of FB and TB tasks in much the same way; that is, performance in FB should be as proficient as in Experiment 1; but performance in TB should be significantly better than in Experiment 1; negative correlations should disappear, and the tasks should be positively correlated instead. For 3-year-old children who do not yet have the competence to operate with fully-fledged belief concepts, though, these manipulations will have little effect (they will continue to solve TB and fail FB tasks, and the tasks will remain negatively correlated).

7.3.1. Method

7.3.2.1. Participants. One-hundred-and-one children were included in the final sample (3-year olds, age = 37-47 months, *M* = 44, n = 20; 4-year-olds, age = 48-59 months, *M* = 53, n = 41 and 6-year olds, age = 73-83, *M* = 78, n = 40). Children came from mixed socioeconomic backgrounds and were recruited from a databank of children whose parents had previously given consent to experimental participation. Four additional 4-year-olds were tested but excluded from data analyses because they were uncooperative (n = 3) or due to experimental error (n = 1). Children were tested by a female experimenter either in a quiet room of their day care or in the laboratory.

7.3.2.2. Design and Procedure. The basic design was a 2 (belief: TB-FB) X 2 (condition: standard- aspectuality) within-subjects design. Each child received four trials in total, two trials of standard change-of-location tasks and two trials of aspectuality tasks, with each trial containing TB and FB questions. The order of the tasks as well as which protagonist was holding the TB/FB was counterbalanced across subjects. Again, the vocabulary subscale of the Kaufmann Assessment Battery for Children was used to measure children's verbal ability. The same tasks as in Experiment 1 (standard and aspectuality) were used, with the following modification.

Instead of one protagonist per trial holding either a FB of a TB, I introduced two protagonists per trial, of which one was holding a FB and the other one holding a TB. This was realized as described below.

7.3.2.2.1. Standard Task. The standard task differed from the task used in Experiment 1 in the following ways (see Figure 5). Instead of one protagonist, I introduced two protagonists [e.g. ape and horse]. In the presence of both protagonists an object [e.g. ball] was put in a box [box1]. Then one of the protagonists [the ape] left the situation. In her absence and the presence of the other protagonist [the horse] the object was then transferred to the other box [box2] and the horse left the situation, too. In the absence of both protagonists the first and second control questions were asked (in cases in which children did not answer correctly, the experimenter explained the relevant part of the story to them again and corrected them).

- Control Question 1: Who was present when we transferred the object from box 1 to box 2? [correct answer: the horse]
- Control Question 2: What about the other one? Was she present? [correct answer: no]

Then both protagonists returned and the following control and test questions were asked.

- Control Question 3: Where did we put the object in the beginning? [correct answer: box 1]
- Control Question 4: Where is the object now? [correct answer: box 2]
- Test Question 1: What does the horse think where the object is? [depending on her belief]
- Test Question 2: What does the ape think where the object is? [depending on her belief][7]

7.3.2.2.2. Aspectuality Task. The basic logic of the different aspectuality tasks (FB/TB) did not differ from the ones used in Experiment 1. The difference was again that I used two protagonists within a trial, who left the scene at different points in time. The one holding the FB left the scene before the object’s dual identity was revealed;

[7] Note that test question 1 and 2 remained always the same. It was counterbalanced whether the horse or the ape was holding the FB.

the one holding the TB left the scene after learning the dual identity. The test and control questions in the aspectuality task were the same as in the standard task with the following differences in the first and second control question.

- Control Question 1: Who knows that the A [e.g. pen] is also a B [e.g. rattle]? [correct answer: depending on who was holding the TB]
- Control Question 2: What about the other one? Does she know? [correct answer: no]

7.3.2. Results

7.3.3.1. Control questions. The percentages of children who spontaneously answered the different kinds of control questions correctly and thus needed no feedback is depicted in Table 6.

Table 6
Children's performance on the control questions as a function of questions and age group in Experiment 2 in Study 2.

% trials correct	Standard Task				Aspectuality Task			
	CQ1: Presence	CQ2: Other one?	CQ3: Location 1	CQ4: Location 2	CQ1: Knowledge	CQ2: Other one?	CQ3: Location 1	CQ4: Location 2
3-year-olds	63%	90%	55%	55%	53%	63%	93%	60%
4-year-olds	99%	96%	88%	91%	89%	93%	96%	96%
6-year-olds	100%	100%	98%	98%	100%	100%	100%	100%

62% of all children (N = 63) answered all control questions correctly. While 95% of the 6-year olds (N = 38) and 57% of the 4-year-olds (N = 24) did so, only one 3-year old answered all control questions correctly.

7.3.3.2 Main Analyses (whole sample)

7.3.3.2.1 Consistency across trials. The consistency in performance of children over trials 1 and 2 of the same type of task and belief was high for all conditions (Φs > .34). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

7.3.3.2.2 Performance as a function of condition. The mean sum scores of trials in which children answered TB questions correctly and the sum scores of trials in which they answered FB questions correctly as a function of age and task type are depicted in Figure 7.

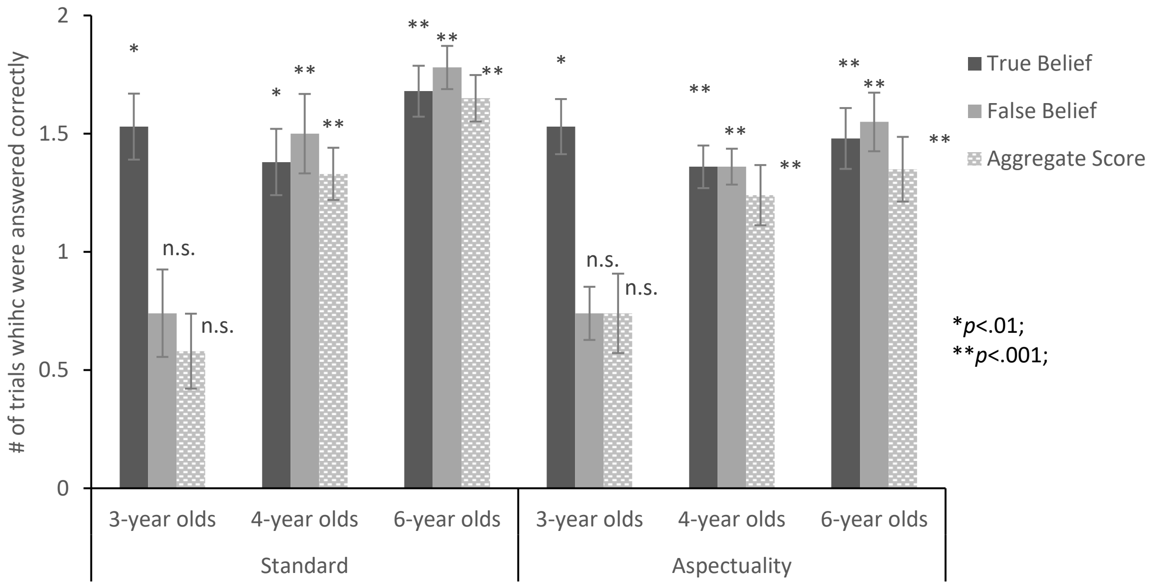


Figure 7. Mean number of trials in which TB and FB questions were answered correctly and aggregate scores as a function of age and task type in Experiment 2 in Study 2. [note that the chance level of guessing correctly differed between TB/FB (chance level = 50%, i.e. 1) and the aggregate score combining both measure (chance level = 25%, i.e. 0.5)]

A 2 (belief type: TB/FB) x 2 (task type: standard change-of-location/aspectuality) x 3 (age) mixed-factors ANOVA on these mean sum scores of correct trials yielded no main effect of task type ($F(1,98) = 1.76$, $p =$ .10), a main-effect of belief type ($F(1,98) = 8.23$, $p < .01$, $\eta p^2 = .08$) and a main effect of age group ($F(1,98) = 9.05$, $p < .001$, $\eta p^2 = .16$. Furthermore, there was an interaction effect between belief type and age group ($F(2,98) = 12.17$, $p < .001$, $\eta p^2 = .20$).

To test for children's competence as a function of task type and age, separate planned comparisons against chance were conducted. These analyses revealed that all age groups performed significantly above chance on all TBs (3-year olds: standard TB, *M* = 1.45, *t*(19) = 2.93, *p* < .01, *d* = .65 and aspectuality TB, *M* = 1.50, *t*(19) = 3.68, *p* < .01, *d* = .82 ; 4-year olds: standard TB, *M* = 1.41, *t*(40) = 3.96, *p* < .001, *d* = .61 and aspectuality TB, *M* = 1.37, *t*(40) = 3.06, *p* < .01, *d* = .48 and 6-year olds: standard TB, *M* = 1.67, *t*(39) = 7.46, *p* < .001, *d* = 1.17 and aspectuality TB, *M* = 1.47, *t*(39) = 3.68, *p* < .001, *d* = .58).

On FB, however, 3-year olds did not perform above chance in both task types (standard FB, *M* = .80, *t*(19) = -1,07, *p* = .30 and aspectuality FB, *M* = .75, *t*(19) = -1,56, *p* = .14) while 4- and 6-year-olds did so (4-year-olds: standard FB, *M* = 1.49, *t*(40) = 5.23, *p* < .001, *d* = .82 and aspectuality FB, *M* = 1.37, *t*(39) = 3.57, *p* < .01, *d* = .50; 6-year-olds: standard FB, *M* = 1.78, *t*(39) = 10.22, *p* < .001, *d* = 1.63 and aspectuality FB, *M* = 1.55, *t*(39) = 4.44, *p* < .001, *d* = .70).

7.3.3.2.3 Correlations. The correlations of TB and FB for the different tasks and age groups are depicted in Table 7. As can be seen from the table, FB and TB performance was highly correlated for the 4- and 6-year-olds (*r*s > .54) but not for 3-year olds (rs < .3).

Table 7
Correlations (and partial correlations correcting for age and language ability) between TB and FB versions of a given task in Study 3.

	Standard FB/TB	Aspectuality FB/TB
All children	.47 * (.43)*	.50 * (.53)*
3-year-olds	.18 (.10)	-.05 (.10)
4-year-olds	.59* (.56)*	.78* (.56)*
6-year-olds	.75* (.73)*	.54* (.73)*

**p*<.001

7.3.3.2.4 Aggregate Scores Analyses. In a second analysis, I computed aggregate scores that took into account whether children solved both, TB and FB, within a given trial. A trial only received an aggregate score “correct” if children answered both TB and FB in this trial correctly (with a chance level of guessing correctly of 1/4). The sum aggregate scores as a function of condition and age group are depicted in Figure 7.

A 2 (task type: standard/aspectuality) x 3 (age group: 3-, 4- and 6-year olds) ANOVA on these mean aggregate scores revealed that there was no main effect of task type (*F*(1,98) = .71, *p* = .40), but a main effect of age (*F*(2,98) = 12.34, *p* < .001). Post-hoc Tuckey-B tests revealed that this was due to the fact that 3-year-olds performed worse than 4-year-olds (*p* < .01) and 6-year olds (*p* < .001), while 4- and 6-year-olds’ performance did not differ (*p* = .26).

Post-hoc tests against chance showed that 3-year olds did not perform above chance (standard, *M* = .55, *t*(19) = .33, *p* = .75 and aspectuality, *M* = .70, *t*(19) = 1.22, *p* = .24) while 4- and 6-year-olds did so in both the standard and aspectuality task (4-year olds: standard, *M* = 1.37, *t*(40) = 7.94, *p* < .001, *d* = 1.25 and aspectuality, *M* = 1.27, *t*(40) = 6.10, *p* < .001, *d* = .95 and 6-year-olds: standard, *M* = 1.65, *t*(39) = 11.69, *p* < .001, *d* = 1.85 and aspectuality, *M* = 1.35, *t*(39) = 6.22, *p* < .001, *d* = .98).

Aggregate scores for the standard and the aspectuality tasks were correlated (*r* = .40) even if controlled for age and verbal ability (*r* = .32).

7.3.3.2.5 Control Analyses (only the sub-sample with correct control questions). Since the present task was rather taxing in terms of memory demands, in particular for the younger children, a substantial number of 4-year-olds, and even the majority of 3-year-olds answered at least one control question incorrectly. Therefore, in a secondary more conservative control analysis, these children were removed from the analyses. These control analyses on the remaining sub-sample of children answering all control questions correctly (one 3-year-old, 24 4-year-olds, and 38 6-year-olds) largely replicated the results of the main analyses for the 4- and 6-year-olds (given the sample size of n = 1 of the remaining 3-year-olds, this age group could not be included in the control analyses) (for details, see Appendix D).

7.3.3.2.6 Complementary Analysis: Comparison between Experiment 1 and 2. In order to test whether the removal of the potential performance factors in Experiment

2 made a crucial difference to TB (but not to FB) performance, I compared the FB and TB performance of the 4-and 6-year-olds between Experiment 2 and Experiment 1. These analyses revealed that the 4-year-olds in Experiment 2 outperformed those in Experiment 1 in TB ($t(53) = 4.86$, $p <.001$; $d = 1.46$), but not in FB ($t(53) = .19$, $p = .85$). The same was true for 6-year olds who performed better in Experiment 2 than in Experiment 1 in TB ($t(51) = 6.01$, $p< .001$, $d = 1.78$ but did not differ in their FB performance ($t(51) = 1.31$, $p = .20$).

7.3.3. Discussion

The main results of Experiment 2 were the following. First, the modified TB version was much easier than the previous versions. 4- and 6-year-olds performed competently on the present TB tasks and significantly better than they did in Experiment 1. Second, children’s performances on FB and TB of the different tasks were now positively correlated, with strong convergence between tasks. Third, 3-year olds performance remained largely unchanged (although these findings remain somewhat difficult to interpret given the poor performance on control questions), in the sense that they performed competently on TB but still failed FB tasks. Taken together, these findings are thus clearly in line with the predictions of performance limitation accounts.

8. Study 3

8.1. Introduction

Study 3 aimed to further investigate the idea that children's initial problems with TB tasks were driven by performance factors. The rationale of this study is to test for different potential factors that may have caused the problem. In Experiment 1 I removed belief as the content of the test by using false and true photo tasks that were perfectly parallel to location change TB and FB tasks. Children did not have any problems in this analogous tasks that did not involve beliefs. Therefore, the problem seems to be special to belief discourse. Experiment 2 tested for the effect of direct questions by replacing direct belief questions with a non-verbal helping paradigm. Again children did not show any problems with TBs. Additionally, Experiments 3 to 5 investigated the role of the usage of test questions by exchanging them with genuine questions. These experiments do not show an improvement in children's TB performance. However, it remains open whether the absence of an effect speaks against a relevance of this factor or if it was due to deficits in the realization of genuine questions. In Experiment 6 I designed a new procedure in which the salience of TBs was increased by making them belief updates. This manipulation did not have a positive effect on children's TB performance. Finally, in Experiment 7 I tested the idea that children's performance increases when they have the possibility to resolve their confusion caused by the trivial question. Therefore, I introduced different task complexity levels that could explain why TBs were so easy. Children's TB performance benefited from this. Taken together, the experiments in Study 3 aim to support the performance-limitation account by serving a detailed framework for the problem's occurrence.

8.2. Experiment 1

In this experiment, I investigated whether the TB performance is specific to belief discourse by testing children in a parallel task that was analogous but did not involve belief attribution.

8.2.1. Method

8.2.1.1. Participants. Thirty-one 4- to 5-year-olds (49- 72 months, *M* = 60, *SD* = 8) from mixes socioeconomic background were tested. Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter in the laboratory.

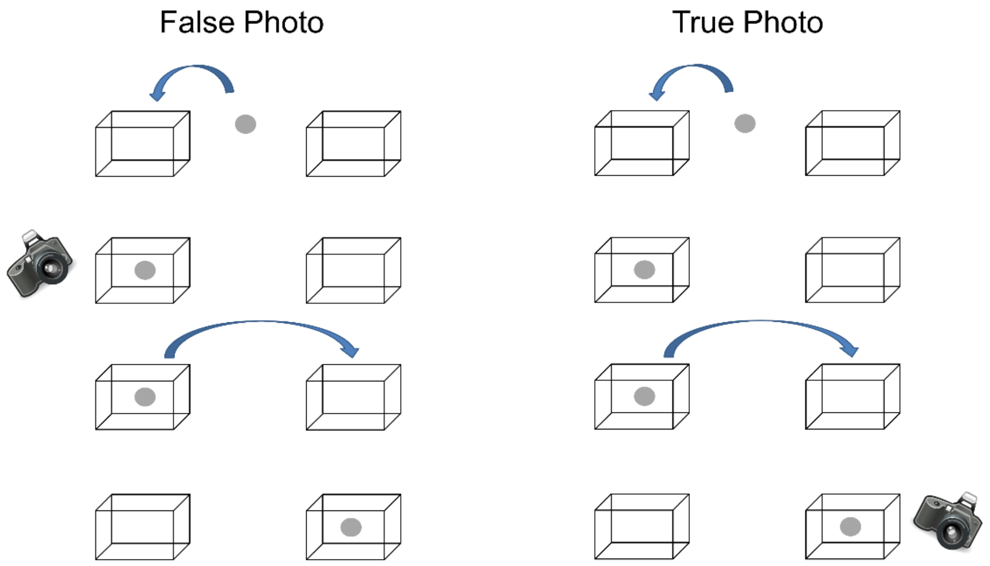
8.2.1.2. Design and Procedure. The basic design was a 2 (Task: belief/photo) x 2 (Condition: false/true) within subjects design. Children received two trials of each task in each condition resulting in eight trials per child. The order of the tasks was counterbalanced as well as the order of the conditions within the task blocks.

8.2.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.2.1.2.2. Belief (Standard Location Change) Task. Four trials of standard change-of-location tasks with different stimuli were administered per child, 2 in TB and 2 in FB versions (Wimmer & Perner, 1983). The protagonist and the child were introduced to an object X. The object was then placed in one of two boxes (box1) before the protagonist left. Either in her absence (FB condition) or after her return (TB condition), the object was moved to the other box (box2) and the following control and test questions were asked.

- Control Question 1: Where did we put the X in the beginning? [correct answer: box1]
- Control Question 2: Where is the X now? [correct answer: box2])
- Test question: Where will the protagonist look for the X? [correct answer: box 2 (TB)/box 1(FB)]

8.2.1.2.3. Photo Task. Each child received four trials of a Photo Task (modelled after Zaitchik, 1990). The child was introduced to a digital camera and was allowed to take a picture of something (e.g. a drawing she draw in the warm-up phase before the testing). The child and the experimenter then together looked at the digital photo on the camera. In the next step the experimenter introduced a protagonist and two rooms of her flat (two transparent boxes, one designed as a living room and one as a bedroom). The protagonist then placed an object first in one of the rooms (opened box, placed object and closed box again). Either before (true) or after (false) a photo of the scene was made by the experimenter the protagonist moved the object to the other box (see Figure 8).


False Photo

True Photo

Figure 8. Procedure of the photo tasks used in Experiment 2 in Study 3.

8.2.2. Results

Children answered the control questions correctly in 87% of the FB trials, 98% of the TB Trials and in 100% of False and True Photo trials.

8.2.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same task and condition were high (Φs > .38, *p*s < .05), despite of False Photo (Φ = .02, *p* = .90) (see Table 8). The percentage of children who showed the same performance in both trials of a task was 77% in FB, 90% in TB, 71% in False Photo and 74% in True Photo. Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 8
Consistencies in children's performances over trials 1 and 2 of the different tasks in Experiment 2 in Study 3

		Trial 2							
		Belief Task				Photo Task			
		False		True		False		True	
		0	1	0	1	0	1	0	1
Trial 1	0	4	2	23	3	1	6	5	5
	1	5	20	0	5	3	21	3	18

8.2.2.2. Main Analyses. The mean sums of trials answered correctly as a function of task and condition are depicted in Figure 9. Children performed significantly above chance on FB ($t(30) = 3.72$, $p < .01$, $d = .66$), False Photo ($t(30) = 6.52$, $p < .001$, $d = 1.17$) and True Photo ($t(30) = 3.05$, $p < .01$, $d = .55$), while significantly below chance on TB ($t(30) = -3.77$, $p < .01$, $d = -.68$).

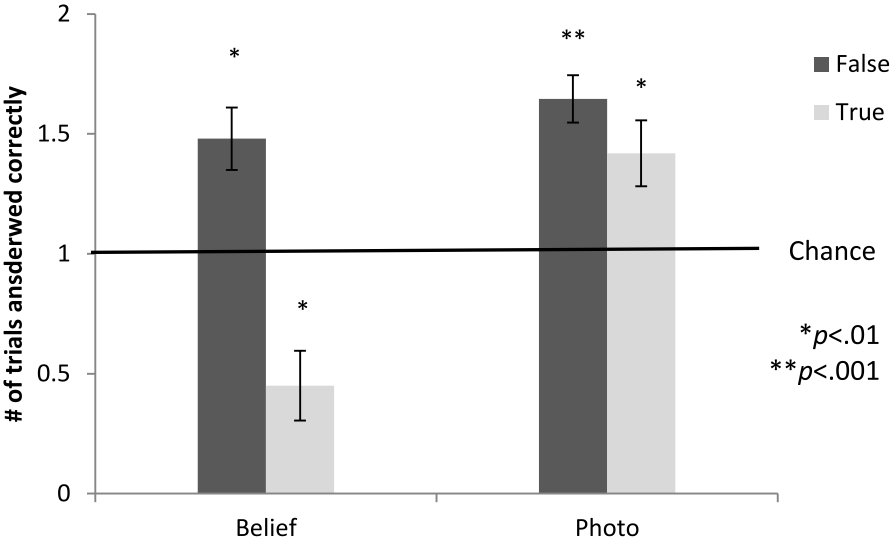


Figure 9. Children's performance on the different tasks in Experiment 2 in Study 3.

A 2 (Task: belief/photo) x 2 (Condition: false/true) x 2 (Order: False-True/True-False) repeated measures ANOVA revealed the following. A main effect of task ($F(1,29) = 35.77$, $p < .001$, $\eta p^2 =.55$), a main effect of condition ($F(1,29) = 17.89$, $p < .001$, $\eta p^2 = .38$) and a main effect of order ($F(1,29) = 5.92$, $p < .05$, $\eta p^2 = .17$, such that children starting with False performed better). Furthermore, an interaction of task and order ($F(1,29) = 7.33$, $p < .05$, $\eta p^2 =.20$; such that children starting with False performed especially better on Photo), no interaction of condition and order ($F(1,29) = .09$, $p = .76$) and no interaction of task, condition and order ($F(1,29) = .01$, $p = .91$).

Correlations of the different tasks and conditions is depicted in Table 9. FB and TB were highly significantly negatively correlated (even corrected for age and verbal ability) while False and True Photo were positively correlated when corrected for age and verbal ability.

Table 9
Correlations of the different tasks and conditions in Experiment 2

	True Belief	False Photo	True Photo
False Belief	-.61**	.03	-.02
	(-.58**)	(.05)	(-.03)
True Belief		.07	.01
		(.06)	(-.10)
False Photo			.29
			(.37*)

**p*<.05; ***p*<.01

8.2.3. Discussion

Experiment 1 has two main findings. (i) The classical performance pattern on FB and TB could be replicated (high performance on FB and low performance on TB and negative correlation); and (ii) no such pattern could be found for False and True Photo. This supports the hypothesis that TB performance problems are specific to belief discourse.

8.3. Experiment 2

In this experiment I investigated the role of direct test questions by replacing them by indirect measures.

8.3.1. Method

8.3.1.1. Participants. Seventy- five 4- to 7-year-olds (4-year-olds, N = 20, 49-59 months, *M* = 55; 5-year-olds, N = 20, 62- 73 months, *M* = 66; 6-year-olds, N = 20, 74-83 months, *M* = 77; 7-year olds, N = 15, 86-95 months, *M* = 90) from mixes socioeconomic background were included in the final sample. One additional 6-year old was tested but excluded from data analyses because she obviously guessed (using a counting-out game). Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter in the laboratory.

8.3.1.2. Design and Procedure. Children were tested in two trials of FB and TB in a randomized order.

8.3.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.3.1.2.2. Sticker Game. Children were introduced to a sticker game in which they could win stickers they were allowed to select from a box containing several stickers. A protagonist who joined the game to help the child was introduced. Each trial began by the child choosing the sticker. The sticker was then hidden in one of two boxes behind an enclosure invisible to the child while the protagonist was behind the enclosure and could see the sticker’s location. After the enclosure was removed, the protagonist gave an advice by saying “I think it is in here” and pointing to one of the two boxes. The experimenter then moved the boxes towards the child who was allowed to choose freely in which of the boxes she wanted to look. If the child found the sticker she won it, if she looked in the wrong box or tried to cheat, the sticker was lost and placed in a savings box being no more available (see Figure 10).

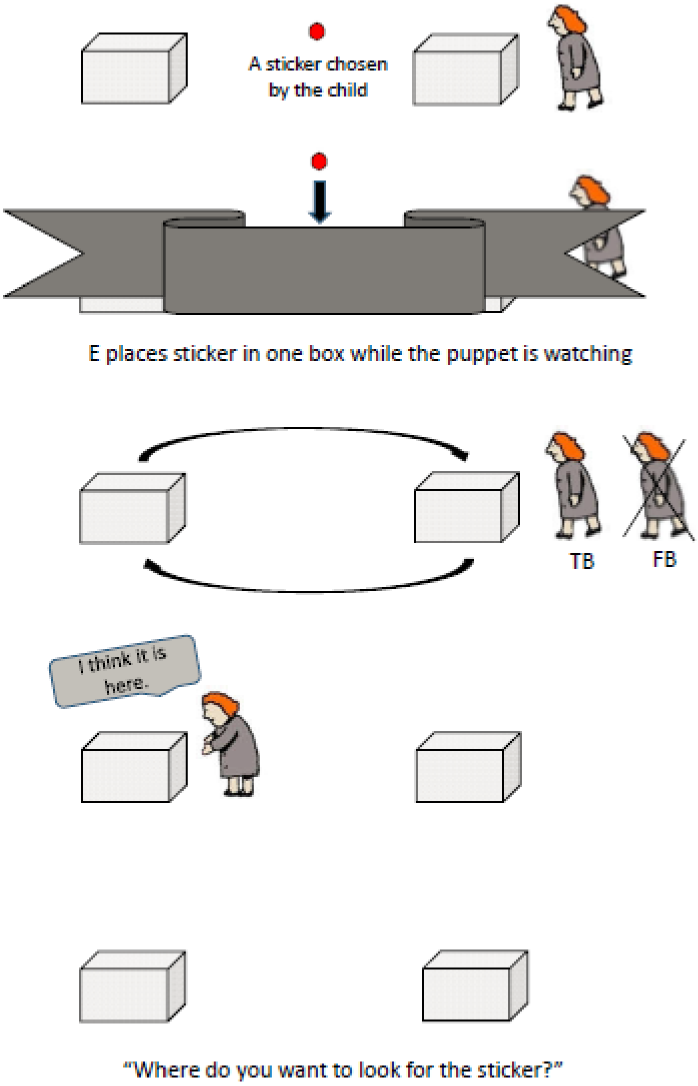

A sticker chosen by the child
E places sticker in one box while the puppet is watching
TB
FB
I think it is here.
“Where do you want to look for the sticker?”

Figure 10. Procedure of the sticker game used in Experiment 1.

Warm-up trials. Warm-up trials were nearly modelled after Call and Tomasello, (1999), Apperly and colleagues (Apperly, Samson, Chiavarino, & Humphreys, 2004) and Fizke et al. (2014). Children received two trials of a control in which the puppet saw the sicker being hidden and immediately after the removal of the enclosure gave his advice (“Control Trial”). In the following two trials (“Invisible Displacement”) after the protagonist gave his advice the boxes were exchanged visible to the child and then the child was allowed to search. In the next two trials (“Ignore Communicator”), after the removal of the enclosure the protagonist claimed that he had to leave. In the protagonist’s absence the sticker was moved from one to the other box by visibly (to the child) transferring it to the other box. After the puppet’s return, the puppet gave a (wrong) advice in accordance to his belief. And the child was again allowed to search. Before the test trials began again two Control Trials were played. These warm-up trials aimed to make clear that the child was allowed to disobey the protagonist’s hint.

False Belief. After the sticker was hidden and the enclosure was removed, the protagonist left the scene. In the protagonist’s absence, the boxes were exchanged resulting in a FB of the agent about the sticker’s location. The real location was still unknown to the child. Upon his return, the protagonist gave an advice in accordance to his FB and the child was allowed to search.

True Belief. After the sticker was hidden and the enclosure removed the protagonist left the scene. In the absence of the protagonist nothing happened. Upon the protagonist’s return protagonist the experimenter exchanged the boxes, while the protagonist watched and said “hmm, aha” and then gave his advice in accordance to his TB.

8.3.2. Results

Percentages of children solving both, one or none of the given controls is depicted in Table 10.

Table 10
Percentages of children solving both, one or none of the given controls in Experiment 1

	Control 1	Invisible Displacement	Ignore Communicator	Control 2
Both	71%	59%	91%	78%
One	25%	30%	5%	19%
None	4%	11%	3%	3%

8.3.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same belief were high (Φs > .27, *p*s < .05). Children who showed the same performance in both trials of a belief were 69% in FB and 73% in TB. Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses (see Table 11).

Table 11

Consistencies in children's performances over trials 1 and 2 in Experiment 1

		Trial 2			
		False Belief		True Belief	
		0	1	0	1
Trial 1	0	22	16	7	14
	1	7	30	6	48

1.2.2. Main Analyses

The mean sums of trials answered correctly as a function of conditions are depicted in Figure 11. As a group, children did not perform differently from chance in FB ($t(74) = 1.11$, $p = .27$) but in TB ($t(74) = 6.95$, $p < .001$, $d = . 80$).

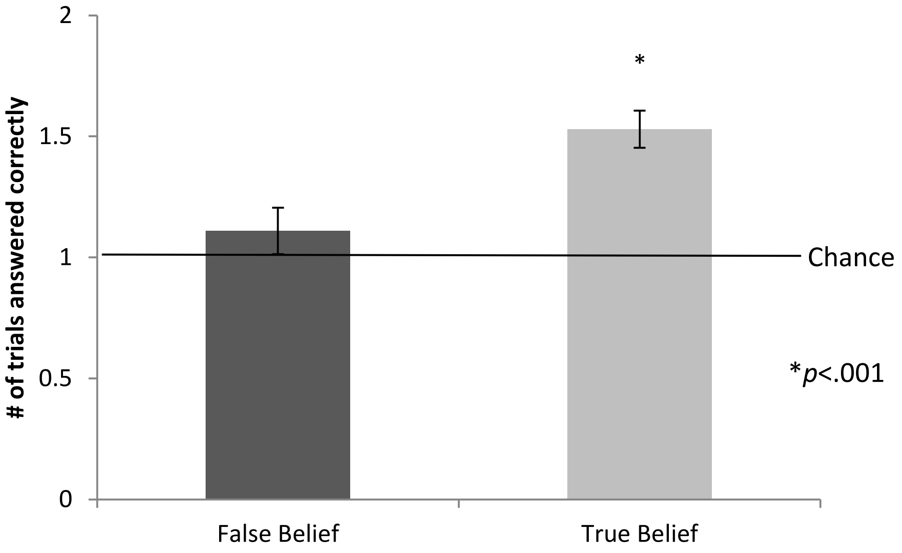


Figure 11. Children's performance on the different tasks in Experiment 1.

A 2 (belief: FB/TB) x 4 (age groups: 4-/5-/6- and 7-year olds) ANOVA revealed a main effect of belief ($F(1,71) = 10.01$, $p < .01$, $\eta p^2 = .12$), a main effect of age group ($F(3,71) =$ 5.96, $p < .01$, $\eta p^2 = .20$) and a interaction ($F(13,71) = 4.69$, $p < .01$, $\eta p^2 = .16$). Therefore, I analysed children's performance separately for the different age groups. While all age groups performed significantly above chance on TB (4-year-olds, $t(19) = 4.95$, $p < .001$, $d = 1.11$; 5-year olds, $t(19) = 2.18$, $p < .05$, $d = .49$; 6-year-olds, $t(19) = 3.33$, $p < .01$, $d = .74$; 7-year olds, $t(14) = 4.18$, $p < .01$, $d = 1.09$), on FB only 7-year olds performed significantly above chance ($t(14) = 6.21$, $p < .001$, $d = 1.60$), while 4-year olds performed significantly below chance (t(19) = -3.25, p < .01, d = -.74) and 5- and 6-year-olds performed at chance (5-year olds, t(19) = 1.83, p = .08; 6-year olds, t(19) = .25, p = .80) (Figure 12).

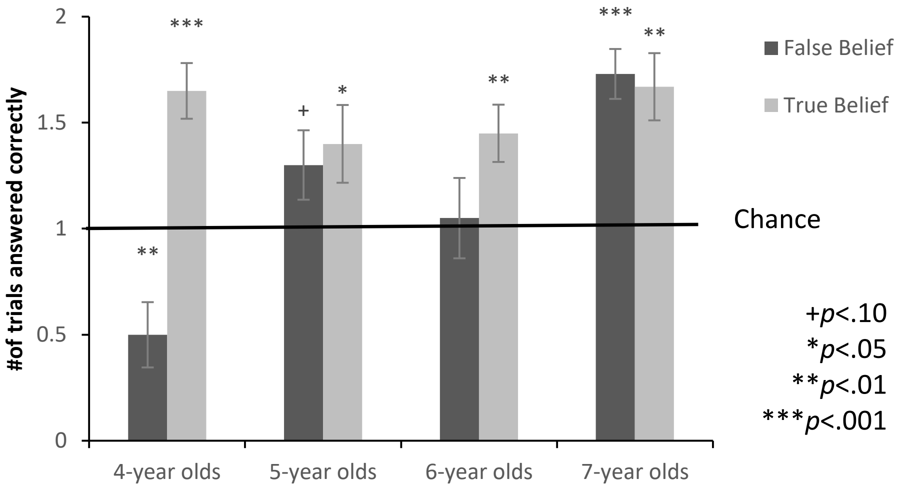


Figure 12. Different age groups' performances on the different tasks in Experiment 1.

Correlation of False and True belief for the different age groups is depicted in Table 12.

Table 12
Raw and partial correlations (controlling for age and verbal ability) between False and True Belief in Experiment 1

	All age groups	4-year-olds	5-year-olds	6-year-olds	7-year-olds
FB-TB	-.18 (-.31*)	-.20 (-.20)	-.21 (-.18)	-.54* (-.55*)	.67** (.67**; only corrected for age in months)

*p<.05, **p<.01

8.3.3. Discussion

In this experiment, all age groups performed above chance on TB, while only 7-year olds succeeded in FB. This findings show that (a) in this case FB is more difficult to solve than in classical direct measures and (b) no u-shaped curve is revealed in children's TB ascription development. The delay in FB may have been caused by the amount of trials children received before the critical test trials began. This may have allowed them to conclude behavioural rules (e.g. some children said “I always have to pick what the puppet says in order to win!”, note that this was not even correct for the warm-up trials) in order to solve the task but a control study in which I reduced the number of control trials did not reveal an improvement in children’s FB performance (Appendix E). Another reason may be that children already won or lost the most interesting stickers when the critical trials began and therefore lost interest in the game, but again a control study with special stickers for the critical test trials did not show a different pattern (Appendix F). It seems plausible that children’s FB performance classically profits from directs questions. Possibly because directly asking for a belief highlights the possibility, that this belief can be wrong. At the same time, the very same mechanism (asking directly for a belief and highlighting the possibility that it can be false) seems to hinder children from showing their real TB competence.

From my point of view, taken together these findings suggest that children’s initial difficulties in TB tasks are caused by the direct question about an agent’s belief. Moreover, this seems to be the same parameter facilitating the attribution of FB.

8.4 Experiment 3

To investigate whether children’s poor performance was caused by the usage of a test question in former experiments I tested 4- and 6-year-olds in FB and TB location change and aspectuality tasks in which the control and test questions were asked by a seemingly ignorant experimenter. This was done to transfer these test questions into genuine questions. Instead of an experimenter who knows the answers to the questions, this time questions are asked by a person who is genuinely seeking information.

8.4.1. Method

8.4.1.1. Participants. Seventy-three 4- and 6-year-olds (4-year-olds, N = 35, 48-59 months, *M* = 53 and 6-year-olds, N = 38, 71-84 months, *M* = 77) from mixes socioeconomic background were included in the final sample. Five additional children were tested but excluded from data analyses because they broke up or due to experimental errors. Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter in the laboratory, an additional female experimenter joint the experiment at a given time point.

8.4.1.2. Design and Procedure. The basic design was a 2 (condition: standard-aspectuality) x 2 (belief: FB-TB) design, with condition as a between- and belief as a within-subjects factor. Each child received two trials of each belief in the given condition resulting in four trials in total. The order of TB and FB blocks was counterbalanced across subjects.

8.4.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.4.1.2.2. Standard Location Change Task. The Location Change Task used in this experiment was very similar to the Belief Task used in Experiment 2. Again, either in the presence (TB) or the absence (FB) of a protagonist the initial location of an object was changed. The critical manipulation was the following. Upon the protagonist's return just before the experimenter could start asking control and test questions, experimenter 2 knocked the door, entered the room and told that there was an urgent phone call for the experimenter. The experimenter then suggested, that E2 could continue to play with the child, the protagonist and the object and left the scene. Now E2 continued with the following introduction: "Hmm… You played with [the protagonist] and the [object], I see." and asked the following control and test questions.

- "Where did you put [the object] in the beginning?" (Control Question 1)
- "Where is [the object] now?" (Control Question 2)
- "Did [the protagonist] see that you moved [the object] from there over there?" (Control Question 3)
- "If [the protagonist] wants [the object], where will she look for it?" (Test Question)

E2 then continued with the next trial and was interrupted by the experimenters return just before she could ask the control and test questions.

8.4.1.2.3. Aspectuality Task. The basic logic of these tasks (closely modelled after Study 3 of Rakoczy et al., 2015) is the following. In the presence of a protagonist an object was put into a box (box 1) under aspect A [e.g. pen]. In the protagonist's presence (TB) or absence (FB) it was revealed that the object had another identity B [e.g. rattle] and it was stored in the same box again. In the presence of the protagonist the object was now transferred to box 2 under its identity B (for example, the experimenter covered the object with her hands while taking it out of its initial box, rattled with it and then moved it to the other box such that the A-identity (pen) remained invisible throughout and only the B-identity (rattle) could be heard). In both belief conditions (TB and FB) the protagonist witnessed the object's transfer. Again the critical manipulation was realized by an experimenter change. Only one control question differed from the Standard Task. Instead of asking whether the protagonist saw the object' transfer (Control Question 3) the experimenter asked the child about the protagonist's knowledge about the object's dual identity ("Does [the protagonist] know that [A] is also a [B]?").

8.4.2. Results

Children answered control questions correctly in the following percentages of all trials: Standard Task FB 96%, Standard Task TB 89%, Aspectuality Task FB 88%, Aspectuality TB 78%.

8.4.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same kind of task were high for all tasks and conditions. The percentages of children who showed the same performance in both trials of a given type of tasks were 90% in Standard FB, 85% in Standard TB, 79% in Aspectual FB, 85% in Aspectual TB; all Φs > .60, despite of Standard FB Φ = .28) (see Table 13). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 13
Consistencies in children's performances over trials 1 and 2 of the different tasks in Experiment 3

Trial 2

		Location Change Task				Aspectuality Task			
		False		True		False		True	
		0	1	0	1	0	1	0	1
Trial	0	1	2	10	3	7	3	23	2
1	1	2	34	3	23	2	22	3	6

8.4.2.2. Main analyses. The mean sums of trials answered correctly as a function of conditions are depicted in Figure 13. When both age groups and tasks taken together, children performed significantly above chance on FB (t(72) = 8.16, p < .001, d = .95) but not on TB (t(72) = -.26, p = .80) (see Figure 13).

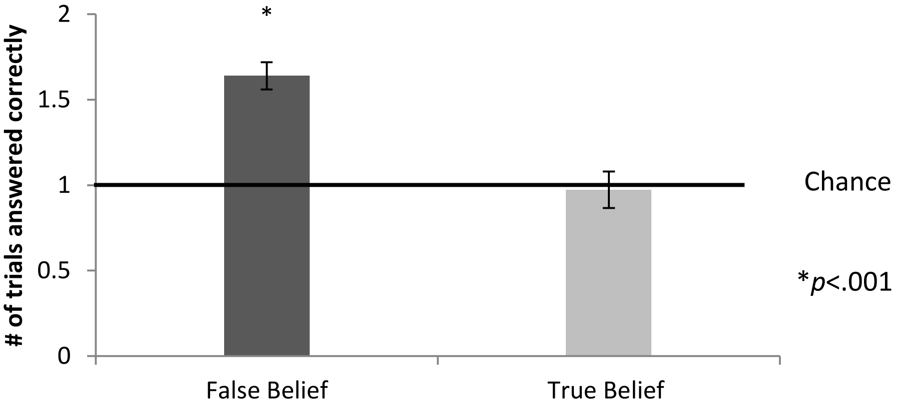


Figure 13. Children’s performance on the different tasks in Experiment 3.

Since preliminary analyses (2 (Belief: false/true; within) x 2 (Task: standard location/aspectuality; between) x 2 (age group: 4-/6-year-olds, between) x belief order (false-true/ true-false) mixed factors ANOVA on the mean sum of trials correct) failed to find any main or interaction effects for the order (false-true vs. true-false) of test blocks (all ps > .23), this factor was skipped from further analyses. A 2 (Belief: false/true; within) x 2 (Task: standard location/aspectuality; between) x 2 (age group: 4-/6-year-olds, between) mixed factors ANOVA revealed a main effect of belief type (F(1,69) = 20.47, p < . 001,ηp^2 = .23), a main effect of task type (F(1,69) = 45.95, p < .001, ηp^2 = .40), a main effect of age group (F(1,69) = 4.32, p < .05, ηp^2 = .06) and an interaction effect between age group and task type (F(1,69) = 7.71, p < .01, ηp^2 = .10) as well as between belief, age group and task type (F(1,69) = 6.55, p < .01, ηp^2 = .09). Therefore, separated tests against chance were conducted for the different age groups, task and belief types (see Figure 14). 4-year-olds performed significantly above chance only in Standard Location FB (t(19) = 6.84, p < .001, d = 1.53; TB: t(19) = -.25, p = .80; Aspectuality FB: t(14) = 1.17, p = .26 and Aspectuality TB: t(14) = -.90, p = .38). 6-year olds, however, performed significantly above chance on Standard Location Change FB (t(14) = 9.80, p < .001, d = 2.24) and TB (t(14) = 4.92, p < .001, d = 1.13) as well as on Aspectuality FB (t(18) = 3.28, p < .01, d = .75) and significantly worse than chance on Aspectuality TB (*t*(18) = -4.02, *p* < .01, *d* = -0.93) (see Figure 14).

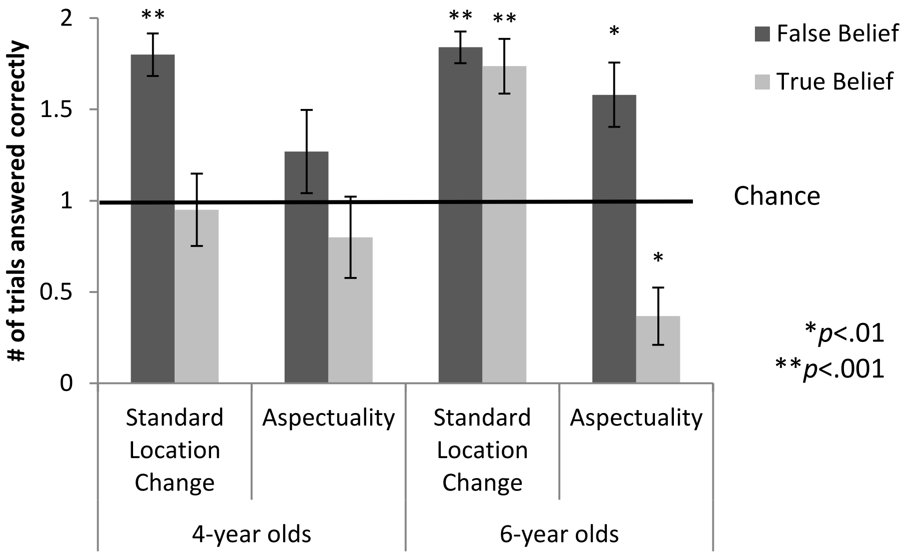


Figure 14. Different age groups' performances on the different tasks in Experiment 3.

Raw and partial correlations (correcting for age and verbal ability) of the sum scores (overall and for the different age group) between the different task and belief types were computed (Table 14). Overall FB and TB were significantly negatively correlated (*r* = -.29, *p* < .05). This was mainly driven by the negative correlations for 4-year-olds in both tasks and for 6-year-olds in the Aspectuality task (see Table 14).

Table 14
Raw and partial correlations (controlling for age and verbal ability) between False and True Belief in the different tasks used in Experiment 3

		Both tasks	Standard Location Change	Aspectuality Task
Correlation FB-TB	Overall	-.29* (-.34**)	-.11 (-.25)	-.77** (-.75**)
	4-year-olds	-.49** (-.51*)	-.36 (-.59*)	-.77** (-.75**)
	6-year-olds	-.11 (-.27)	.28 (.17)	-.75** (-.76**)

*p<.05; **p<.01

8.4.3. Discussion

Experiment 3 had the following results. For the location change task, the use of an ignorant experimenter improved 6-year-olds performance on TB, while 4-year-olds showed the classical pattern of mastering FB and failing TB. For the aspectuality task, however, the effect of our manipulation was different in the different age groups. 4-year-olds did not perform significantly different from chance at all. However, descriptively they seem to show the same pattern of mastering FB and failing TB. 6-year-olds in contrast do show this pattern significantly.

Therefore, this manipulation only succeeded in the standard task and was only helpful for 6-year-olds. There are different candidate causes for this. (i) While adding a second experimenter was helpful in the easy task, for the task with a higher cognitive demand, this manipulation may have not been clear enough. In some cases, children changed the cause of events when asked by the second experimenter, making a FB out of a TB in the Aspectuality task. Instead of saying that the protagonist knew about the object's dual identity, children often answered that the protagonist did not know. However, due to the use of an ignorant experimenter, the child could not be corrected (otherwise experimenter 2 would have had knowledge about the task making her no more ignorant). This may count for the difference between the tasks. Additionally (ii) this manipulation seemed to disrupt children's concentration on the task. 4-year-olds may have been more affected by the experimenter change in their concentration than 6-year-olds. This could explain the difference between the age groups. Therefore Experiment 4 aimed to realize the use of a genuine question in (a) a setting in which children do not have the chance to change a TB into a FB and (b) the task flow is not interrupted as it was by an experimenter change.

8.5. Experiment 4

To overcome the limitations of Experiment 3 in this experiment I introduced a silly puppet who asked questions with obvious answers. This aimed to establish that the puppet could ask genuine questions with obvious answers.

8.5.1. Method

8.5.1.1. Participants. Thirty-four 4- to 6-year-olds (49-80 months, *M* = 62, *SD* = 8) from mixes socioeconomic background were included in the final sample. Four additional children were tested but excluded from data analyses because they broke up or due to experimental errors. Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter either in the laboratory or in an appropriate room in their day care.

8.5.1.2. Design and Procedure. The basic design was again a 2 (condition: standard- aspectuality) x 2 (belief: FB-TB) design, with condition as a between- and belief as a within-subjects factor. Each child received two trials of each belief in the given condition resulting in four trials in total. The order of TB and FB blocks was counterbalanced across subjects.

8.5.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.5.1.2.2. Standard Location Change and Aspectuality Task. The Location Change Task used in this experiment was again designed closely after the classic paradigm by Wimmer and Perner (1983) and the task used in the experiments before. With the following modifications. Before the actual protagonist was introduced, the Experimenter introduced a hand puppet as her friend who often needs some time to understand what is going on and therefore tends to ask stupid questions. But for she really liked the hand puppet, if the hand puppet has questions they should help him. Then the experimenter introduced the actual protagonist, the boxes and the object. In this standard procedure, the hand puppet repeatedly asked questions to which the answer was obvious[8]. Finally, after the change of location (witnessed (TB) or not witnessed (FB) by the puppet) the hand puppet asked test and control questions. In the standard location change task the control and test questions were the following.

- "Hmm… Wait a minute. Where did we put [the object] in the beginning?" (Control Question 1)
- "Aha, okay. And where is it now again?" (Control Question 2)
- "Okay, I see. And what does [the puppet] think where the object is?" (Test Question)

[8] E.g. when the experimenter placed the boxes on the table the absentminded hand puppet asked where the boxes are. The experimenter then prompted the child to help the hand puppet by showing the boxes.

For the Aspectuality task an additional control question (Control Question 0) was asked by the hand puppet when the dual identity of the object was revealed (“Oh, wait a minute! Does the puppet know that [the X] is also a [Y]?”).

8.5.2. Results

Children answered control questions correctly in the following percentages of all trials. Standard Task FB 75%, Standard Task TB 75%, Aspectuality Task FB 75%, and Aspectuality TB 64%.

8.5.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same kind of task were high for all tasks and conditions. The percentages of children who showed the same performance in both trials of a given type of tasks were 88% in Standard FB, 76% in Standard TB, 88% in Aspectual FB, 94% in Aspectual TB; all Φs > .75, despite of Standard TB Φ = .38) (see Table 15). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 15
Consistencies in children’s performances over trials 1 and 2 of the different tasks in Experiment 4

		Trial 2							
		Location Change Task				Aspectuality Task			
		False		True		False		True	
		0	1	0	1	0	1	0	1
Trial	0	6	0	10	1	5	3	11	2
1	1	1	9	2	3	0	10	0	5

8.5.2.2. Main analyses. The mean sums of trials answered correctly as a function of conditions are depicted in Figure 15. Children did not perform significantly above chance on FB (t(33) = 1.45, p = .15) but on TB they performed significantly worse than chance (t(33) = -2.61, p < .05, d = -.45) (see Figure 15).

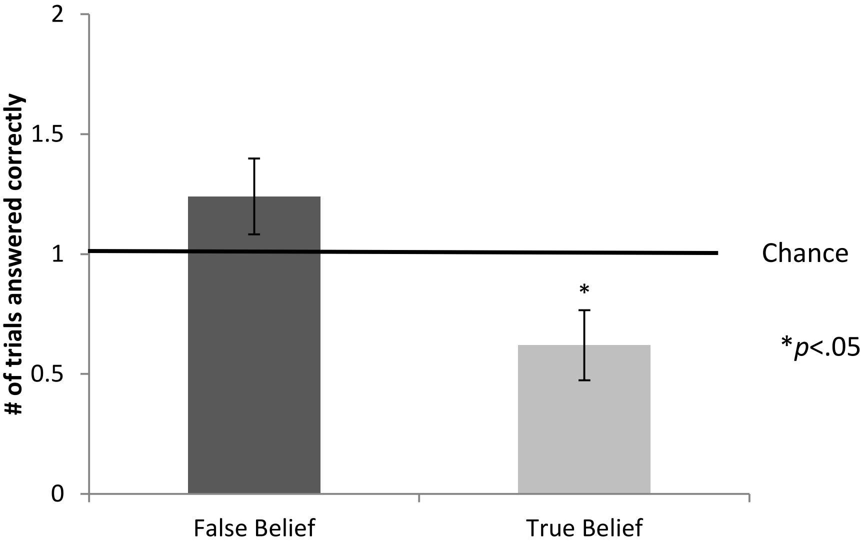


***Figure 15**.* Children's performance on the different tasks in Experiment 4 in Study 3.

Since preliminary analyses (2 (Belief: false/true; within) x 2 (Task: standard location/aspectuality; between) x belief order (false-true/ true-false) mixed factors ANOVA on the mean sum of trials correct) failed to find any main or interaction effects for the order (false-true vs. true-false) of test blocks (all *p*s > .27), this factor was skipped from further analyses. A 2 (Belief: false/true; within) x 2 (Task: standard location/aspectuality; between) mixed factors ANOVA revealed a main effect of belief type (*F*(1,32) = 4.93, *p* < . 05,ηp^2 = .13; FB was easier than TB), no main effect of task type (*F*(1,32) = .51, *p* = .48) and no interaction effect between belief and task type (*F*(1,32) = .00, *p*=.98). As a next step separated tests against chance were conducted for the different task and belief types. Children's performance did not differ significantly from chance in both FBs (Standard Location Change FB, *t*(15) = .76, *p* = .46; Aspectuality FB, *t*(17) = 1.32, *p* = .21)). Standard Location Change was significantly worse than chance (*t*(15) = -2.15, *p* <. 05, *d* = -.54), Aspectuality TB remained on chance level (*t*(17) = -1.56, *p* = .14).

Raw and partial correlations (correcting for age and verbal ability) of the sum scores between the different task and belief types were computed (Table 16). Overall FB and TB were significantly negatively correlated (*r* = -29, *p* < .05). This was mainly driven by the negative correlations for 4-year-olds in both tasks and for 6-year-olds in the Aspectuality task (see Table 16).

Table 16
Raw and partial correlations (controlling for age and verbal ability) between False and True Belief in the different tasks used in Experiment 4

	Both tasks	Standard Location Change	Aspectuality Task
Correlation	-.61**	-.48	-.75**
FB-TB	(-.57**)	(-.40)	(-.66*)

**p*<.01; ***p*<.001

8.5.3. Discussion

Experiment 4 has one main finding. In this experiment, children's TB performance was not improved by the manipulation and even the negative correlations could be replicated. The main difference between the procedures used in Experiment 3 and 4 seems to be the following. While in Experiment 3 the genuine question was asked by a person who could not have access to the relevant information (because she was absent when the critical events took place) in Experiment 4 the genuine questions were not that plausible. Despite the fact that the protagonist in Experiment 4 seemed to be confused and therefore asked real questions, children may not have interpreted the test question as a genuine question. To overcome the limitations of Experiment 3 and 4 in Experiment 5 I combined the used procedures. I introduced a puppet who was absent when the critical events occurred.

8.6. Experiment 5

Experiment 5 aims to overcome the limitations of Experiments 3 and 4 by combining their procedures.

8.6.1. Methods

8.6.1.1. Participants. Nineteen 4- and 5-year-olds (49-71 months, *M* = 60) from mixes socioeconomic background were included in the final sample. One additional child was tested but excluded from data analyses because she broke up. Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a male experimenter in an adequate room in their day care.

8.6.1.2. Design and Procedure. The basic design was a 1 (condition: standard) x 2 (belief: FB-TB) within-subjects design. Each child received two trials of each belief resulting in four trials in total. The order of TB and FB blocks was counterbalanced across subjects.

8.6.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.6.1.2.2. Standard Location Change Tasks. The Location Change Task used in this experiment was again very similar to the Belief tasks used before with the following modifications. Before the experimenter introduced the protagonist, he announced that he brought a friend of his to join the game they were going to play, but that this friend was sleeping. Therefore he asked the child to pay attention in order to be able to explain what was going on to his friend when he joins. In both belief conditions (False and True) after the change of location took place and the protagonist returned, the experimenter said that he heard his friend wakening and introduced him to the scene and the friend puppet asked the following control and test question.

- "Where did you put [the object] first?" (Control Question 1)
- "Where is it now?" (Control Question 2)
- "Did [the protagonist] see that?" (Control Question 3)
- "What does [the protagonist] think were the object is?" (Test Question)

8.6.2. Results

Children answered control questions correctly in the following percentages of all trials: Standard Task FB 68% and Standard Task TB 50%. The poor performance on control questions was driven by a low performance on Control Question 1 (70% of all trials, while Control Question 2 and 3 in > 89% of the trials) in the FB task and by a low performance on Control Question 3 (57% of the trials, while Control Question 1 and 2 in > 89% of the trials) in the TB task.

8.6.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same kind of task were high for all conditions. The percentages of children who showed the same performance in both trials of a given type of tasks were 79% in FB and 63% in Standard TB (Φs > .27) (see Table 17). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 17
Consistencies in children's performances over trials 1 and 2 of the different tasks in Experiment 5

		Trial 2			
		False Belief		True Belief	
		0	1	0	1
Trial 1	0	2	2	6	4
	1	2	13	3	6

8.6.2.2. Main analyses. The mean sums of trials answered correctly as a function of belief condition are depicted in Figure 16. Children performed significantly above chance in FB ($t(18) = 3.64$, $p < .01$, $d = .84$) but at chance in TB ($t\ (18) = 0$, $p = 1$) (see Figure 16).

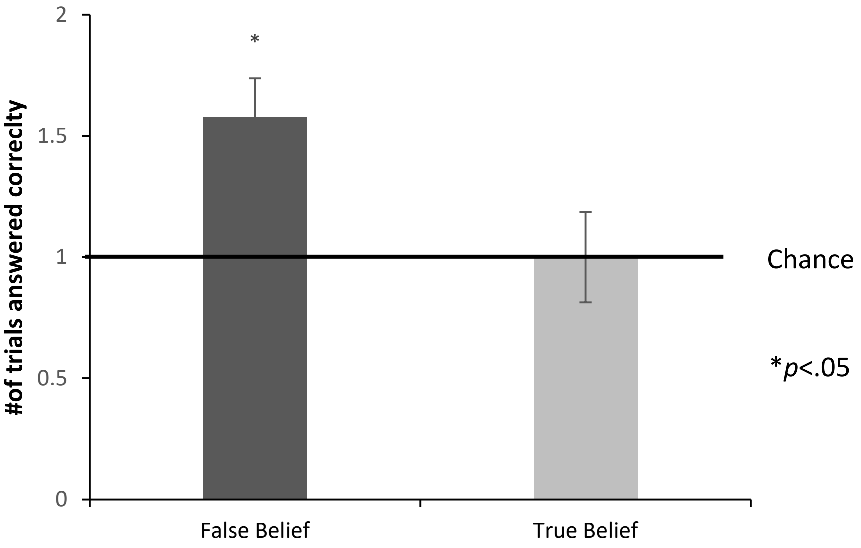


Figure 16. Children's performance on the different tasks in Experiment 5.

A repeated measures ANOVA revealed that children performed significantly better on FB than on TB ($F(1,18) = 5.07$, $p < .05$, $\eta p^2 = .22$). Children's FB and TB performances were not significantly correlated ($r = -.10$, $p = .69$; controlled for age and verbal ability, $r = .02$, $p = .94$).

8.6.3. Discussion

In this experiment, children again performed above chance on FB but not on TB. Therefore, this experiment could not extend the limited finding of a positive effect of genuine question usage from Experiment 3 (where at least 6-year-olds in the standard task profited from the manipulation).

To sum the three experiments on test-/genuine-question up, I investigated whether the usage of test questions can count as an explanation for children's difficulties in TB tasks. One of my manipulations (Experiment 3) in order to make the test question asking for the TB of an agent a genuine question seemed to have an effect- at least for the location change and only for the 6-year-olds. The other attempts (Experiment 4 and 5) did not have any positive effect. However, my data is not sufficient in order to judge about the role of test questions in children's TB performance.

8.7. Experiment 6

To investigate whether children's poor performance was due to differences in salience between FB and TB I designed an Experiment in which TB is also a belief update. In the classical procedures, TB does not contain any point in time where the belief of the protagonist becomes salient and therefore notable to the child. In FB, however, there is a point in time where the protagonist arrives at her FB and therefore, the child has a reason in order to ascribe a belief to the protagonist. If this is the reason for children's TB performance problem, manipulation this factor should cause an increase in children's TB performance.

8.7.1. Methods

8.7.1.1. Participants. Thirty-two 4- to 6-year-olds (48-71 months, *M* = 58, *SD* = 8) from mixes socioeconomic background were tested. Five additional children were tested but excluded from data analyses because they intervened into the procedure (N = 3) or broke up (N = 2). Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter either in the laboratory or an appropriate room in their day care.

8.7.1.2. Design and Procedure. The basic design was a 2 (belief order: FB-TB; TB-FB) x 2 (belief: FB-TB) design, with belief order as a between- and belief as a within-subjects factor. Each child received two trials containing both a FB and a TB, resulting in four measuring points per child.

8.7.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.7.1.2.2. Standard Location Change. The location change task used in this study was closely modelled after the belief tasks used before with the following modification. After one belief was tested, e.g. a FB trial where the object was first hid in box 1 and protagonist's absence moved to box 2, the trial continued with the other belief. After the child was asked control and test-questions about the initial and actual position of the object and the protagonist's belief, now the trial went on by the protagonist again leaving the situation. This time in the protagonist's absence nothing happened. After the protagonist's return the object was now transferred from box 2 back to box 1. In this scenario, the protagonist's TB is a belief update because the protagonist did not know the actual object's location (because he was holding a FB). Therefore both, the information that it was now in box 2 and its transfer to box 1 are reasons for the child to update the representation of the protagonist's belief. However, in the opposite order (TB first and FB second) TB does not have this characteristic.

8.7.2. Results

Children answered control questions correctly in 91% of the trials.

8.7.2.1. Consistency across trials. The consistencies in performance of children over trials 1 and 2 of the same belief was high (Φs = .80). Children gave consistent answers across trials in 97% of all cases (see Table 18). Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 18
Consistencies in children's performances over trials 1 and 2 of the different tasks in Experiment 6

		Trial 2			
		False Belief		True Belief	
		0	1	0	1
Trial 1	0	2	1	29	1
	1	0	29	0	2

8.7.2.2. Main analyses. The mean sums of trials answered correctly as a function of belief are depicted in Figure 17. Overall children performed above chance on FB (*t* (31) = 9.23, *p* < .001, *d* = 1.64) and significantly below change on TB (*t*(31) = -9.23, *p* < .001, *d* = -1.64) (see Figure 17).

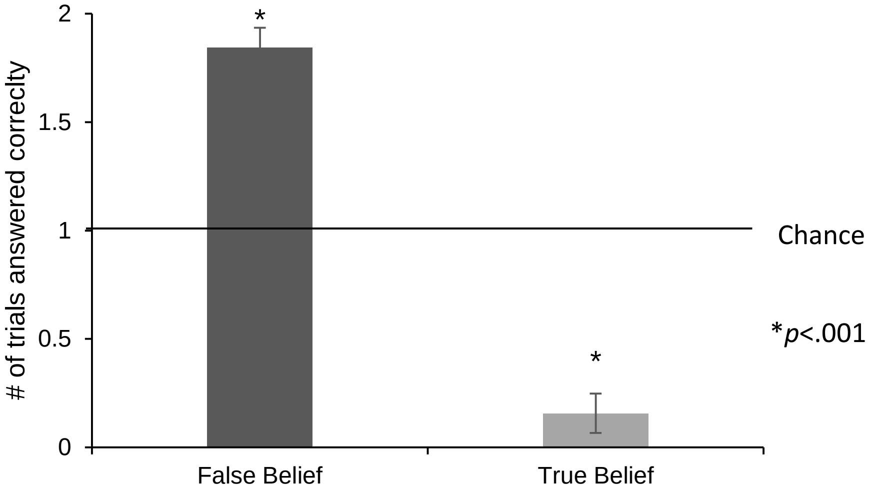


Figure 17. Children's performance on the different tasks in Experiment 6

A 2 (belief order: FB;TB- TB;FB) x 2 (belief: FB-TB) mixed factors ANOVA revealed a main effect of belief (*F*(1,30) = 101.52, *p* < .001, ηp^2 =.77), no main effect of belief order (*F*(1,30) = 0, *p* = 1) and no interaction effect between these factors (*F*(1,30) = .24, *p* = .63). Nonetheless, separated tests against chance were conducted for the different belief order groups (see Figure 18). Children who started with FB still succeeded in FB (*t*(16) = 7.5, *p* < .001, *d* = .1.82) but were significantly below chance on TB (*t*(16) = -7.5, *p* < .001, *d* = -1.82). Likewise children who started with TB performed significantly above chance on FB (*t*(14) = 5.53, *p* < .001, *d* = 1.07) but below chance on TB (*t*(14) = -5.23, *p* < .001, *d* = -1.07) (see Figure 18).

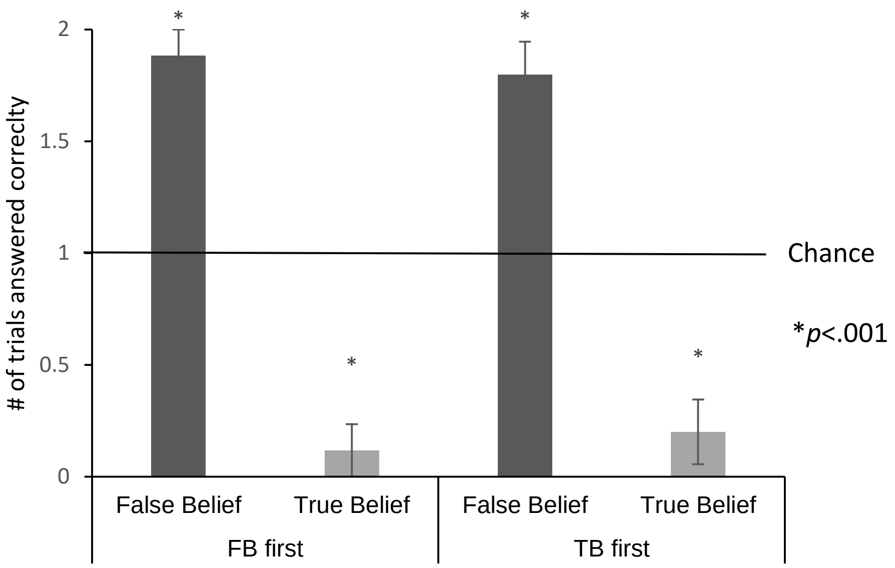


***Figure 18*.** Different belief order group's performances on the different tasks in Experiment 6

Raw and partial correlations (correcting for age and verbal ability) of the sum scores (overall and for the different belief orders) between the different belief types were computed (Table 19). Overall FB and TB were significantly negatively correlated (*r* = -64, *p* < .001; partial correlation, *r* = -.61, *p* < .001).

Table 19
Raw and partial correlations (controlling for age and verbal ability) between False and True Belief in the different conditions used in Experiment 6

	Overall	FB first	TB first
Correlation FB-TB	-.64*	-1*	-.32
	(-.61*)	(-1*)	(-.31)

**p*<.001

8.7.3. Discussion

Experiment 6 had the following results. Overall children performed better on FB than on TB. Additionally, our manipulation of conducting both kinds of belief trials within a trial did not have any effect. In both belief orders children performed above chance on FB and below chance on TB. For the belief order TB-FB, this is not surprising. However, if the root of children's TB performance problem was salience/belief update, children’s performance should have increased in the FB-TB condition. For this did not happen, this seems not to be the origin of children’s TB failure.

Another candidate explanation is that children may be confused by the simple question about the TB of an agent, tricking them into thinking that the experimenter implies something relevant and therefore the agent must have a FB.

8.8. Experiment 7

To investigate whether children’s poor performance was due to pragmatic confusion caused by the simplicity of the test question, I tested children’s TB performance in a testing situation where children had an alternative explanation for why the TB questions were so easy.

8.8.1. Method

8.8.1.1. Participants. Eight-teen 4- to 6-year-olds (49-81 months, *M* = 63, *SD* = 9) from mixes socioeconomic background were tested. Children were recruited from a databank of children whose parents had previously given consent to experimental participation. Children were tested by a female experimenter in the laboratory.

8.8.1.2. Design and Procedure. The basic design was a 2 (belief order: FB-TB; TB-FB) x 2 (belief: FB-TB) design, with belief order as a between- and belief as a within-subjects factor. Each child received two trials of each belief in the given order resulting in four trials in total.

8.8.1.2.1. Verbal Ability. At the beginning of the session, children were given a vocabulary test (the vocabulary subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999).

8.8.1.2.2. Intent Clarification. At the beginning, the experimenter told the following: “Today we are going to play three games. Some of them are really easy, like for 2-year-olds, some are still easy, like for 3-year-olds and some are just for your age, meaning they are no problem for you. Let’s start with something for babies!” The experimenter introduced a protagonist who was going to join the first game and showed the child and the protagonist an object (e.g. a bell) and asked the child what the object was. Then she asked what the protagonist thought what the object was. This was introduced to show that asking a question about the belief of an agent did not mean that the protagonist’s belief had to be different from the child’s knowledge/ the truth.

After two trials the next game was introduced by saying: “Now I have another game for you! This may be easier, a little bit harder or as easy as the one we just played.”

8.8.1.2.3. Standard location change. Four trials of standard change-of-location tasks with different stimuli were administered per child, 2 in TB and 2 in FB versions (Wimmer & Perner, 1983). The protagonist and the child were introduced to an object X. The object was then placed in one of two boxes (box1) before the protagonist left. Either in her absence (FB condition) or after her return (TB condition), the object was moved to the other box (box2) and the following control and test questions were asked.

- Control Question 1: Where did we put the X in the beginning? [correct answer: box1]
- Control Question 2: Where is the X now? [correct answer: box2])
- Test question: What does the puppet think where X is? [correct answer: box 2 (TB)/box 1(FB)]

8.8.2. Results

Children answered all control questions correctly in all trials.

8.8.2.1. Consistency across trial. The consistencies in performance of children over trials 1 and 2 of the tasks was high (Φs = 1, *p* < .001) (see Table 20). Children gave consistent answers across trials in 100% of all cases. Therefore, sum scores of trials answered correctly per condition [0-2] were used for further analyses.

Table 20
Consistencies in children’s performances over trials 1 and 2 of the different tasks in Experiment 7

		Trial 2			
		False Belief		True Belief	
		0	1	0	1
Trial 1	0	2	0	8	0
	1	0	16	0	10

8.8.2.2. Intent Clarification. Four-teen out of eight-teen children gave wrong answers in the first intent clarification trial, either denying to give an answer at all or claiming that the protagonist thinks that the object is something else (e.g. the given object was a bell and children said that the protagonist thought it was a nut). After it was clarified that the task is really easy and the protagonist was asked what she thinks what the object is and she gave the same answer as the child, in trial 2 seven-teen out of eight-teen children were able to answer that the protagonist thought the same as they themselves.

8.8.2.3. Main analyses. The mean sums of trials answered correctly as a function of belief are depicted in Figure 19. Overall children performed above chance on FB (*t(*17) = 5.1, *p* < .001, *d* = .97) but not on TB (*t*(17) = 1.11, *p* = -65) (see Figure 19).

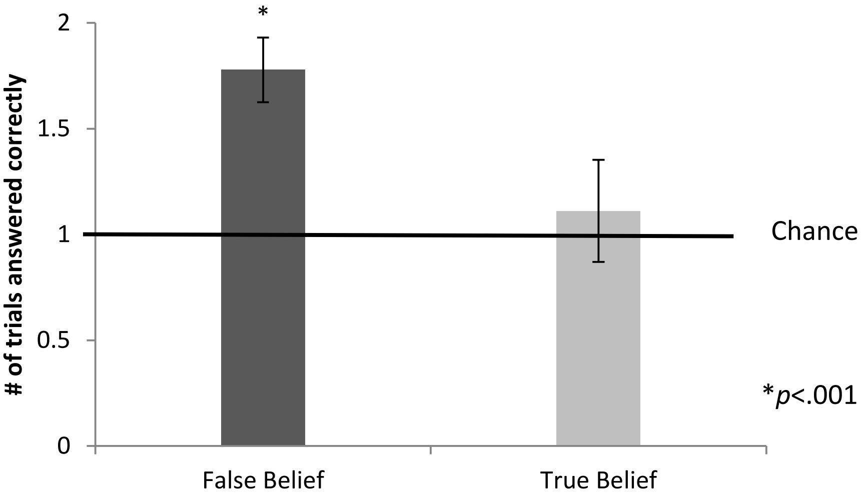


Figure 19. Children’s performances on the different tasks in Experiment 7

A 2 (belief order: FB;TB- TB;FB) x 2 (belief: FB-TB) mixed factors ANOVA revealed a main effect of belief (*F*(1,16) = 5.33, *p* < . 05, ηp^2 = .25), a main effect of belief order (*F*(1,16) = 4.0, *p* < .01, ηp^2 = .45) and an interaction effect between these factors (*F*(1,16) = 5.33, *p* < .01, ηp^2 = .10). Therefore separated tests against chance were conducted for the different belief order groups (see Figure 20). While children who started with FB still succeeded in FB (*t*(8) = 3.5, *p* < .01, *d*=.95) but not in TB (*t*(8) = .44, *p* = .095), children who started with TB performed significantly above chance on both belief types (FB, *t*(8) = 3.5, *p* < .01, *d* =.95 and TB, *t*(8) = 3.5, *p* < .01, *d* = .95) (see Figure 20).

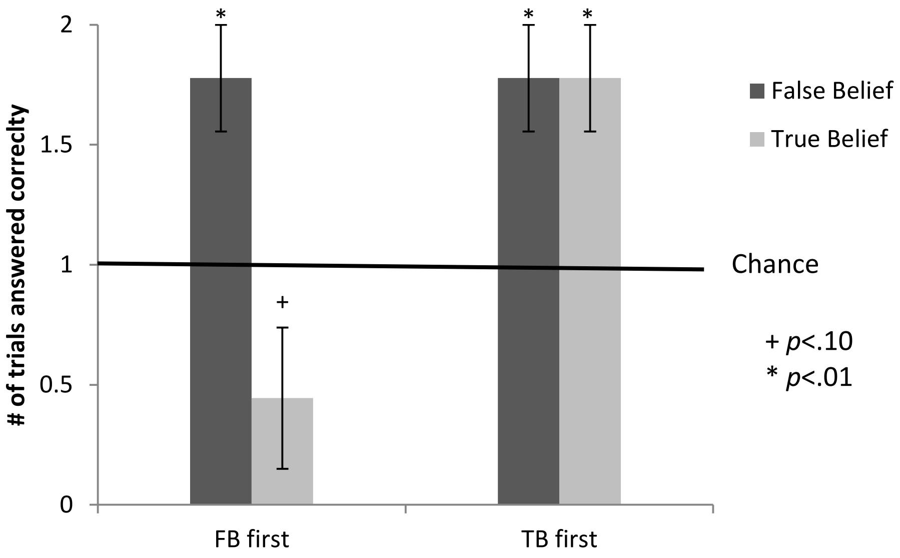


Figure 20**.** Different belief order group's performances on the different tasks in Experiment 7

Raw and partial correlations (correcting for age and verbal ability) of the sum scores (overall and for the different belief orders) between the different belief types were computed (Table 21). Overall FB and TB were not significantly correlated (*r* = -32, *p* = .20; partial correlation, *r* = -.37, *p* = .18) (see Table 21).

Table 21
Raw and partial correlations (controlling for age and verbal ability) between False and True Belief in the different conditions used in Experiment 7

	Overall	FB first	TB first
Correlation FB-TB	-.32 (-.37)	-.66^{+} (-.73^{+})	-.13 (not computable)

$^{+}$*p*<.10; ***p*<.01

8.8.3. Discussion

Experiment 7 had the following results. Overall children performed better on FB than on TB. Additionally, with this manipulation (intent clarification), belief order had a significantly positive effect on children's TB performance. Those children starting with a TB after the intent clarification performed better than children starting with a FB did. There was also a significant interaction such that children starting with a TB performed especially better on TB. This suggests the following. When it is established that the question about the belief of another person does not necessarily suggest that the other person has a belief different from the child's own belief/reality, children are able to solve TB tasks. However, this seems to be very fragile. When children encounter a situation in which the protagonist's belief is different from reality, they again relapse into the interpretation that the question implies a significant (and therefore in this case false) belief.

9. General Discussion

This dissertation aimed at integrating findings of children's failure in explicit ToM tasks into the standard picture of ToM development. To do so I reviewed these findings critically. I tested competence-based explanations of these findings against performance-based ones.

In Study 1, I tested 4- to 6-year-olds' understanding of the aspectuality of beliefs by using newly developed aspectual and non-aspectual FB tasks. Children in this study were able to solve both kinds of FB tasks. Moreover, performance on aspectual and non-aspectual versions were correlated. Experiment 2 in Study 1 asked more directly for children's understanding of the procedures' critical elements and confirmed the findings of Experiment 1. Children showed a clear understanding for the procedure's aspectual elements even when they were directly asked. I will discuss potential limitations of these findings, put them in relation to other studies on children's aspectuality understanding and focus on their impact on the standard picture of ToM development. Additionally, I will discuss the benefit of potentially investigating aspectuality understanding in implicit and explicit ToM in a parallel way.

In Studies 2 and 3 I investigated children's TB competence. Study 2 showed that once TB tasks are slightly modified, e.g. by setting TB questions into a meaningful context, children do handle TBs from four years of age on. None of the presented competence-based accounts can explain the pattern of increased TB performance. Therefore, this clearly speaks in favour of a performance-based explanation. However, Study 2 does not differentiate between different performance-based accounts.

Study 3 strengthens the performance view on TB failure by differentiating between several performance accounts. My findings support the idea that pragmatic mechanisms explain TB performance problems, especially when trivial questions are asked about beliefs. Connections between several findings will be established and discussed.

9.1. Investigating children's aspectuality understanding in explicit ToM

Several studies have shown that children, being older than 4 years of age and able to pass standard ToM tasks, have difficulties to understand the aspectuality of beliefs.

That is, that representations represent entities always only under some description but not under others. This has been interpreted as indicating a true lack of competence. A recent study by Rakoczy et al. (2015), in contrast showed that children are able to take into account the aspectuality of beliefs in simplified tasks. This speaks in favour of a performance-based explanation of children's failure in former studies. Study 1 aimed to investigate children's understanding of aspectuality using even more parallel aspectual and non-aspectual tasks.

In this new task, children were presented a scenario in which a protagonist ended up holding a FB about the number of identically looking objects in a box. The aspectual and non-aspectual versions differed in the way the protagonist had reached this FB. In the non-aspectual version, the protagonist did not witness a critical change of location. Therefore she falsely believed the object to be still in the initial box, leading to a FB about the number of objects in that box. In the aspectual version, however, the protagonist did witness the transfer but not as the transfer of this very object. She again arrived at a FB about the number of objects in that box because she thought she saw two different objects (instead of one object being transferred twice).

Experiment 1 compared these tasks to a classical location change task. Neither the non-aspectual nor the aspectual task were more difficult than the standard task. Furthermore, children's performance on the different tasks were correlated. This clearly speaks for performance-based explanations for former findings of children's failure. However, experiment 1 still leaves room for an alternative explanation of its findings. When children's success in Experiment 1 can be explained without aspectuality understanding, to interpret the results as showing aspectuality understanding constitutes a false positive. Even though this possibility seems unlikely, it was addressed in Experiment 2. When asked directly for the procedure's aspectual characteristics, children were still able to solve the task. This indicates that they had made use of aspectuality understanding in the first place. The results of this experiment even strengthen the assumption that aspectuality is not a limitation to the ToM competence children acquire around age four. Therefore, the ToM acquired around age four is full-blown.

Moreover, Study 1 has the potential to shed light on further ToM related research questions. The aspectuality task used in Study 1 allows to investigate aspectuality understanding in explicit and implicit ToM in direct comparison. The two-systems account of mindreading by Apperly and Butterfill (2009) especially weights toddlers' ability to understand aspectuality. It is assumed to be a signature limit to the early implicit ToM (System 1). A recent experiment by Schulz et al. (2016) used a helping paradigm version of the aspectuality task from Study 1 to test whether toddlers differentiated their helping behaviour according to an agent's non-aspectual or aspectual TB or FB. In this study, children showed a sensitivity in their helping behaviour for the agent's belief in the non-aspectual task. However, they were not able to take into account the agent's belief in the aspectual versions. This finding shows a limitation in children's early implicit ToM ability, and thus supports the two-systems view on early ToM development. This is especially interesting in the light of Study 1's finding that explicit ToM is unified and covers aspectuality understanding. Taken together, my findings from Study 1 and the study by Schulz and colleagues (Schulz et al., 2016) define the role aspectuality understanding plays in ToM development. On the one hand, understanding aspectuality seems to be a signature limit to the early implicit competencies found in children. On the other hand, understanding the aspectuality of beliefs does not challenge the explicit ToM competence acquired around age 4.

The findings of Study 1 fit into conceptual change accounts on ToM development, which suggest that children acquire a full-blown (meta-) representational conception of beliefs and other propositional attitudes around age four. This ability enables them to solve standard explicit ToM tasks and tasks requiring an understanding of the aspectuality of beliefs at the same time. Additionally, this task can be used to differentiate between conceptual change accounts in general and two-systems accounts on mindreading. While the former suggests, that early mechanisms (e.g. situation theories) are replaced by later developing more sophisticated abilities (real ToM competencies), two-systems accounts suggest that both, the early and the later system coexist. Implicit versions of the aspectuality task from Study 1 can be used to test whether there is still an implicit ToM competence in adults that additionally shows the same signature limits as in toddlers.

Taken together, the findings of Study 1 are in line with the standard picture of ToM development. They do not support the overestimation claim. In the case of aspectuality understanding the empirical foundation of this critique steams from extraneous task demands. Extraneous task demands mask children's real competencies and lead to an underestimation of their competence. Once these extraneous task demands are modified, children reveal their real competence.

9.2. Investigating children's TB understanding

Similar to previous findings on 4-to 6-year-olds failure to handle the aspectuality of beliefs, several studies have shown that children at that age window have difficulties attributing TBs. This has also been interpreted as indicating that the ToM competence children acquire around age four is not full-blown. Two different kinds of TB tasks, location change and aspectuality, were used and revealed comparable patterns of failure. Critically, there is no competence-based account that can explain struggle in both kinds of tasks. Different competence-limitation accounts can only explain failure in subclasses of TB tasks. However, one performance-based account can count for failure in both kinds of tasks. Studies 2 and 3 of this dissertation focussed on this foundation of the overestimation-claim.

The aim of Experiment 1 in Study 2 was to investigate children's performance in standard and aspectual FB and TB tasks. This experiment had three main findings. First, performance in different FB tasks develops in a strongly consistent and correlated manner and so does TB performance. Second, performance on both FB tasks increases with age, while TB performance in both tasks forms a u-shaped curve. At ages three and ten, children perform competently, while children of the ages in between perform poorly. Third, performance on FB and TB versions of the tasks are highly negatively correlated (between ages three and six). This findings are surprising despite the empirical background showing similarly poor performance.

The MFCT can explain children's performance pattern in the aspectual tasks in Experiment 1. It predicts the u-shaped age-related curve in TB and the negative correlation between FB and TB. But it cannot count for the whole picture. The account does not predict such a pattern for the standard location change task, which does not include aspectuality. The PAR account, however, does predict such a pattern for standard location change tasks but has one clear requirement. The protagonist must lack perceptual access to any kind of relevant information. Former studies by Fabricius et al. (2010), met this condition by adding unnecessary elements in the protagonist's absence in TB. In the location change TB I used, however, literally nothing happened in the protagonist's absence. The child and the experimenter just waited for the protagonist to come back. This clearly does not meet the criterion of a "comparable lack of perceptual access" (Hedger & Fabricius, 2011, p.432). For there is no event at all the protagonist could miss, using PAR should even lead to a correct answer. Given that the suggested heuristic is seeing → knowing; knowing → getting it right, the following should happen in this TB. There is nothing the protagonist did not see → she knows; she knows → she acts correctly. One could possibly modify the account in order to make it predict the same pattern in the case of this TB version as well. The only plausible possibility would be to add a high degree of automation to the account. A highly automatized PAR account could go like the following. Every time a protagonist leaves the situation, no matter what happens, a lack of perceptual access is attributed. Despite the fact that adding such a post-hoc extra premise would ask for a revision of the whole account, it would not even seem plausible. But still, even assuming that there is some kind of highly automated version, the account cannot sufficiently explain the data. In the TB in Experiment 1 (where the u-shaped curve and the negative correlation between FB and TB is given) the protagonist left the situation *before* the critical event took place. After she came back, the critical event was presented and directly followed up by the control and test-questions. There was no lack of perceptual access between the event and the questions at all. However, in Experiment 2 (where the u-shaped curve and the negative correlation between FB and TB is not given), the TB protagonist left the situation *after* the change of location was presented. One might assume that when she comes back, she might have had an interrupted perceptual access to the situation. Therefore she should get things wrong. Critically, in this TB children do predict successful behaviour. To sum it up, even with this modification the PAR account is not able to explain the results of Study 2 and even predicts the opposite of the pattern found in the data.

Additionally, the problem seems to be even much more serious. Only 10-year-olds were able to reliably master both belief types. What does this show? Following the overestimation claim, the interpretation of this result seems clear. It shows, that the four-year-revolution overestimates the ToM competence children acquire at age four. The age limit for a full-blown ToM is then ten years of age. Maybe this conclusion overpraises this finding. However, this pattern clearly reinforces the need for critically investigating whether this performance is caused by real competence limitations or mere performance problems.

Experiment 2 modified the testing procedures of the TB tasks by using two protagonists, one holding a FB and one holding a TB, within one task. From a competence-limitation point of view, there is absolutely no reason, why this should increase children's performance on TB. From a PAR perspective, there is no reason why using a heuristic twice (once for the FB protagonist and once for the TB protagonist) should improve children's performance. Likewise, from a MFCT point of view, there is no possible explanation, why handling vicarious files for two protagonists should suddenly enable children to coordinate horizontal and vertical linking in order to attribute an aspectual TB. Critically, from a performance problem view, this modification should make a difference. In contrast to former procedures, when two protagonists are used, TB becomes less trivial and the child has an opportunity to resolve its pragmatic confusion caused by the trivial test question about the belief of an agent (e.g. by thinking "The experimenter is asking me about the TB of this protagonist because it contrasts to the other protagonist's FB and is therefore worth mentioning."). The results of Experiment 2 fit perfectly to the predictions of a performance-based account. With this modification children older than four years of age performed competently in both tasks (standard and aspectuality) and belief conditions (FB and TB). This clearly indicates that the initial problem was a performance-, not a competence-limitation problem.

The aim of this study was to test competence-limitation accounts against performance-limitation accounts of children's TB failure in general. From a competence point of view children's failure in initial studies and experiments reveals a true competence limitation. It indicates that children do not have a full-blown ToM, even if they can solve FB tasks. This conclusion really conflicts with the standard picture of ToM development. However, Experiment 2 in Study 2 shows that children's failure does not originate in real competence limitation but is caused by extraneous task demands hindering them from performing in accordance to their real competence. From this point of view children's former failure is a false negative because children older than four years of age do have a full-blown ToM but fail to show this competence in both kinds of TB tasks due to extraneous (most likely pragmatic) task demands.

Taken together the results of Study 2 clearly speak for some kind of performance account and against the two competence-limitation accounts described before. However, this study leaves open three broader questions. First, is there another competence account that can explain the present performance patterns? This account then should offer a plausible explanation, why children’s performance on the TB task in Experiment 2 is just a false positive. To my knowledge there is no such account that can explain both, that the pattern is the same for non-aspectual and aspectual TBs and that it breaks down in both cases when a second protagonist is added. Second, can additional evidence for performance and against competence accounts be found using different methodological approaches? One possibility to test this can be to remove the test question and design a task in which belief attribution is not directly tested. And finally, if the problem is really driven by performance factors, which factors are masking children’s real competences?

One basic difference between FBs and TBs is that while FBs are interesting and salient, TBs are rather boring and trivial. When FBs are realized within the setup, there is some point in time at which it becomes clear (at least to a person holding a ToM) that the protagonist’s belief is interesting because it is different from reality. This may highlight the protagonist’s belief and be an occasion to attribute a belief and keep track of this belief. TBs in contrast totally lack such a moment at which the protagonist’s belief seems relevant in any way. In TBs, therefore, children are confronted with the idea of attributing a mental state to the agent only when they are asked for the protagonist’s belief. This might confuse them. Out of this confusion they then might use the general assumption that the basic idea of belief discourse is that beliefs can be false. This again may cause them to assume that the targeted belief is false. This idea is perfectly compatible with the findings of Study 2. Being confronted with both, a protagonist holding a FB and one holding a TB automatically highlights the TB due to the given contrast (like “A falsely believes the object to be in its initial location while B knows it is at its real location” or “B believes the object to be where it really is, but A has a false belief”). In this context the TB is highlighted too and the child has a reason to represent this TB. However, this idea cannot explain the negative correlations between FB and TB in Experiment 1. Why would children with an advanced ToM have more problems with TBs? The belief attribution caused by salience or relevance of the agent’s TB cannot explain this pattern.

Another, possibly complementary, candidate performance factor is the usage of a test-question. Children can differentiate test-questions from genuine questions from early on (see Siegal & Beattie, 1991). Children’s pragmatic assumptions about test questions have recently been addressed especially in relation to ToM tasks (in order to explain, from a nativist point of view, why children younger than four years of age fail to attribute FBs, see Helming et al., 2014). In the TB debate one possibility how these assumptions may be obstructive is the following. When children are asked stunningly trivial test questions about the belief of a cognizant agent, this violates the basic point of belief discourse, namely that it refers to or at least highlights the possibility of mistakes. It is important to note that this is the very same mechanism that may boost children's FB performance in general. When they are asked about the belief of an agent, the default is that the belief may be or even is wrong. In FB this is indeed the case and, therefore, the test question facilitates the attribution of a FB. But in TB, this may mislead children by causing confusion ("why is she asking me about the agent's belief? The only possible way for the belief to be interesting is that it is wrong."). This account can count for both, the poor performance on TB as well as the negative correlation between FB and TB. When children between four and six years of age are asked for the belief of an agent holding a TB (as it was done in Experiment 1 in Study 2) the triviality confuses them. In a setting with a possibility to resolve this confusion, however, children are able to solve TB tasks (like in Experiment 2, where one could explain the test question's triviality by thinking "she is asking me for the TB of this agent because it is significantly different from the FB of the other agent."). This account can also explain the negative correlation between FB and TB. The more competent children get in their ToM competence, the more pragmatically sensitive they may get to the experimenter's conversational aims ("when she is asking me about this mental state, she maybe wants to give me a hint, that it can be false."). That is why children with advanced ToM are more affected by the TB problem.

Study 3 aimed to test for different performance-based explanations more directly. Besides the possibilities discussed above, Study 3 also investigated further plausible mechanisms causing the effect. Experiment 1 focussed on the question whether children's TB problem is specific to belief discourse or caused by a general problem with easy questions. If a general problem causes their struggle in TB tasks, children should show similar problems with simple questions in any kind of task. This possibility does not seem very likely because children did not have comparable problems with (even easier) control questions. Nevertheless, this is still an alternative explanation that needed to be ruled out. Experiment 1 in Study 3 addressed this idea by using simple test questions in a task that does not involve belief ascription. In addition to location change TB and FB, children were tested in true and false conditions of a photo task. The basic idea is that the photo task is perfectly parallel to the location change task but it does not involve belief ascription. If the problem is a general one, a comparable pattern should be found in the photo task. If the problem is, however, specific to belief discourse, children should not show any difficulties. This experiment had two results. First, this experiment replicated the basic finding on TB and FB. Second, no such pattern was found for the photo task. Children performed equally competently on both photo tasks. This clearly supports the hypothesis that children's difficulties with TB tasks result from a mechanism that is specific to belief discourse.

Experiment 2 investigates the effect of the usage of direct test questions. If the reason for children's difficulties with TB really is connected to some characteristics (e.g. easiness) of the test questions in belief discourse, removing the test question completely should resolve the problem. Instead of asking children directly for a protagonist's belief, Experiment 2 used a non-verbal task that stripped the task's pragmatics altogether. In this setup keeping track of the agent's belief was relevant for a real decision the child had to make. If directly asking for the belief plays a critical role in children's TB failure, in this indirect measure they should succeed in TB. This experiment had three main findings. First, children between four and seven years of age had no difficulties at all with TB tasks. Second, there was a delay in children's FB performance. Only 7-year-olds were able to solve FB tasks significantly above chance. Finally, for the age group performing competently on FB performance on FB and TB were positively correlated. What do these findings show? The basic finding seems to be that asking directly for the belief of a protagonist boosts children's performance on FB tasks. This is probably due to the fact that asking for the protagonist's belief highlights the possibility that this belief is wrong. Children profit from this highlighting and are able to reveal their ToM competence. When they are not directly asked for the belief, this highlighting does not apply anymore and therefore children's competence is masked. At the same time, the absence of the very same mechanism seems to enhance children's TB performance. Without the pragmatic hints connected to direct belief questions, children do not have difficulties assuming that a relevant belief may be true. Taken together, this experiment shows that explicitly asking for an agent's belief plays a role in children's TB performance problem.

One characteristic of the direct measures used in TB tasks is that they were classical test questions. Combined with the triviality of TB, the usage of test-questions may also play a role in masking children's TB competence. Experiments 3 to 5 focussed on this possibility. If children's pragmatic confusion is caused by the usage of such easy questions as test questions, the problem should disappear once test questions are replaced by genuine questions. The potential problem with test questions is that they may mislead children. They might think "Why is she asking me such a stupid test question? The reason cannot be that she doesn't know the answer because she does know it. There must be something else, maybe I missed something." In Experiments 3 - 5 I tried to implement genuine questions into the test procedure. Genuine questions should not cause children to think critically about the test question's triviality. If someone lacks information and asks to seek information, the triviality of her questions should be plausible. In order to implement genuine questions, I used several methods in the different experiments. Experiment 3 involved two experimenters. One experimenter played the scenario with the child until the other one entered the testing room just before control and test questions were asked and went on playing with the child. The second experimenter was not present when the scenario was played and pretended to being ignorant. Therefore, her questions could be genuine information seeking questions. However, this manipulation only had an effect on 6-year-olds performance on location change TB. 4-year-olds in both tasks as well as 6-year-olds in the aspectuality task did not profit. This seems surprising, because the different tasks showed the same pattern in Experiment 1 in Study 2 and were similarly sensitive to the manipulation in Experiment 2 in Study 2. So in sum, the manipulation was not really successful. The procedure used in this experiment has some blind spots which may explain this failure. For instance, the second experimenter may not be perceived as really ignorant. When she entered the situation she said, "Oh, I see, you are playing with [the protagonist] and the boxes!" in order to show interest. Children may have interpreted this as indicating superior knowledge about the situation. Additionally, some children had difficulties to correctly answer the control question, whether the protagonist was present when the object was transferred (location change)/ the dual identity was revealed (aspectuality). The main mistake they made was claiming absence/ignorance in TBs. This, however, makes TBs FBs. The experimenter could not directly correct the child because she was supposed to seem ignorant. For those children claiming that the protagonist had a FB, their wrong answers in TB were consistent with their version of the events. Critically, none of these limitations can explain the very isolated effect of this manipulation (on 6-year-olds in location change TB). At this point, besides the possibility that it was an incidental finding, I do not have any sophisticated explanation for that.

Given the limitations of Experiment 3 (possibility of manipulation failure) Experiment 4 struck a new path of converting the test question into a genuine question. In this procedure, the control and test questions were asked by a silly puppet who proved his absent-mindedness by asking obviously silly questions. In this case it might seem plausible that this puppet asks such a trivial question about the protagonist (TB test question). In contrast to Experiment 3, in this scenario the experimenter could correct the child without breaking with the procedure’s basic idea. However, children did not profit from this manipulation at all. Children still failed in TB tasks, even if they were asked by a silly puppet. Again, there are several potential problems regarding the procedure. One may be that children did not really belief the puppet to be silly. Instead they may have interpreted his questions in a different way. An alternative explanation for the puppet’s behaviour can be something like “Oh, the puppet is asking me about things like e.g. where the boxes are. He might be directing my attention. These boxes may be relevant.” In this case, even the former questions asked by the puppet (to show his absentmindedness) can be interpreted in a meaningful way. Additionally, children might have perceived the question as it was asked directly by the experimenter (e.g. “The experimenter is directing my attention by using the puppet and its questions.”). The questions are not perceived as genuine questions but rather as means of interaction between the child and the experimenter. In this case, the question would be a test question and the manipulation would fail. Again it is unclear what these results show because of these limitations. It is not possible to exclude test questions as a source of the problem because I cannot take the success of the manipulation as granted. Therefore, in a third version of the attempt to change the test question character, I combined the former procedures. In Experiment 5 a puppet was used, who was absent while the events took place and entered the situation just in time in order to ask control and test questions (as genuine questions). This is kind of a combination of the former two studies. Instead of a real second experimenter, a puppet was used and instead of this puppet being stupid, he was legitimately ignorant. In this experiment, again, this manipulation did not have any effect. Children showed the classical pattern of high performance on FB and poor performance on TB. However, the same problems as in Experiment 4 may be responsible for a manipulation failure. The manipulation failed, if children again perceived the puppet's questions as attempts of the protagonist or the experimenter to direct their attention. For I cannot exclude this possibility, it is difficult to interpret the findings of this experiment.

What do these experiments show? I cannot make a final judgment about the role test questions play in the TB problem. I cannot exclude the possibility that the modified test questions of these experiments were still perceived as test questions. This is caused by the weak points of each procedure I described above. These experiments do not provide a proper basis in order to exclude test question as a potential factor in children's problem with TB tasks. Future research is needed to shed light on the real role of this factor. Instead of replacing test questions by genuine questions (as I tried to do), one might change the assumptions children have about test questions. This can be done by setting test questions into a context where children do not expect them to imply something meaningful. The first setup that comes to one's mind is school. In this environment children are used to be asked trivial questions by a cognizant person, the teacher. A trivial test question does not imply anything in this setup. Therefore, children should not be irritated when they are asked a stunningly trivial question in school contexts. They should even be used to questions that are asked to give them an opportunity to disclose their abilities and knowledge. The concrete hypothesis is the following. When children are tested in a school setting, this activates the school associated expectation about test questions. This expectation is that they are used to test the child's knowledge and therefore, can sometimes be really easy. Children should not get confused by the test question's triviality and therefore, be able to solve TBs. Of course, the possibility of implementing this manipulation is limited to a very critical sub class of participants, 6- to 8-year-olds, who fail TBs in my experiment and already attend school. If children's performance differs between the laboratory and the school setting, this can be an indicator that the querist's supposed (and pragmatically concluded) intent plays a role in children's poor TB performance. This special subset of participants is additionally critical because they fail to attribute TBs only in my data. However, they are still relevant and may give hints to the role test questions play.

Experiment 6 focussed on another possible origin of the effect. As mentioned before in contrast to FBs, in TBs the protagonist's belief is less salient. The salience of the FB protagonist's belief makes the child ascribe a belief in the first place. In TB, however, this is missing because the belief is not salient. When asked for the protagonist's belief, children in FB are well prepared whereas in TB children may be surprised by the question. This may lead them to choose the most likely possibility, namely that the protagonist is holding a FB. If this is the case, children should succeed in a TB scenario with a point in time where it is necessary to ascribe or update the protagonist's belief. Experiment 6 tested this possibility by realizing a TB that required a belief update. This was done by making the TB protagonist first hold a FB and then arrive at the TB. However, this manipulation did not have any positive effect on children's TB performance. Even if TBs were tested in a context where they require a belief update, children still struggle. Therefore, it seems unlikely, that salience plays a role in children's hassle with TBs.

Finally, the last idea I investigated was that alternative explanations for the test question's triviality should help children to solve TBs. This idea is already supported by Experiment 2 of Study 2. The mechanism how introducing a second protagonist helped children can be the following. When there is one protagonist holding a FB, this is an explanation for why the experimenter is asking such a trivial test question (TB). To further investigate this possibility, Experiment 7 used a setup, where the experimenter shows that she does not intent to ask challenging questions. The basic idea is that when children see that the experimenter does not have any subliminal pragmatic messages hidden in her question, their performance on TB should increase. Basically, the idea is to hinder children from even asking why the experimenter is asking them such trivial questions. In Experiment 7, I realized this using three tasks differing in their complexity. The first task, the so-called baby-task, was meant to clarify the experimenter's intent. Although, this task was meant to be the manipulation, children's answers to the questions in this game reveal some information about the investigated mechanism. Descriptively, children had some difficulties with this task. Children were shown familiar objects, e.g. a bell, asked what this was and asked what the protagonist believed the object to be. The majority of the children either refused to give an answer to the belief question or gave wrong answers (e.g. saying that the protagonist believes the bell to be a nut, probably because he was a suricate; or using a synonym, probably to avoid using the same expression for both questions, e.g. they said it was a “Glocke” and the puppet thought it was a “Klingel” which are two different German words for bell). When children gave wrong answers, they were corrected, in contrast to TB and FB tasks. Therefore, children did not have problems solving the second trial of this task by simply saying that the protagonist believed the object to be what it really was. The mechanism that I suggest fits to their performance in this task. When children are asked for the belief of a protagonist, no matter how easy the question was, they assume that the experimenter is giving them a pragmatic hint that the protagonist’s belief is significant (“Why is she asking me this? There must be something significant about this belief.”). In this experiment, however, they were able to solve TBs after they learned that the experimenter did not imply anything significant by asking about the protagonist’s belief (e.g. by thinking “Obviously, she does not intent anything by asking such trivial questions.”). Critically, this was only the case when TB followed directly on this baby-game. When children were presented with the baby-game, a FB and then a TB, they again failed to solve TBs. What does this show? From my point of view, this shows how strong the pragmatic assumption is. When children are confronted with FBs, they realize that their initial assumption about the experimenter’s intention was correct and they fall back to the usage of this pragmatic heuristic. When there was a situation, in which the agent had a FB, they again interpret the experimenter’s question as indicating something similarly interesting. Taken together, children seem to be able to overcome their pragmatic assumptions when they are presented with an alternative explanation for the triviality of the TB test question (increasing task complexity when they are presented with the baby-task, TB and FB in this order). However, they fall back to their common pattern, when they are remembered on the fact that beliefs can be false (in the baby-task, FB, TB order).

Taken together Study 3 has six main findings. First, children’s TB problem is specific to belief discourse. Second, the usage of test questions is decisive. When the question is removed completely, children perform competently on TB. Third, the very same mechanism (using direct test questions) boosts children’s FB performance. They performed less competently on FB when they were not directly asked for the agent’s belief. Fourth, the possibility that the usage of test questions (in contrast to genuine questions) has an effect on children’s performance cannot be definitively answered yet. Fifth, the difference of salience between FB and TB (that FB involves a belief update while TB does not) cannot count for the initial performance pattern. Children’s

TB performance was not increased when TB was modified into being a belief update and therefore salient. Sixth, when children were given the possibility to resolve the pragmatic assumption that the belief must be false, they are able to successfully ascribe TBs.

The picture of the TB performance problem therefore seems to be the following. When children are asked directly trivial test questions about beliefs, they assume that this belief is most probably wrong. The performance problem disappears, when belief discourse is not the topic, the question is removed, or children are given the possibility to resolve the pragmatic assumption based on the test question's triviality. The role of the usage of test question, however, remains unclear and needs further research in order to clear its real role.

What does this mean for the overestimation claim concerning the standard picture of ToM development? Study 3 clearly supports the conclusion based on Study 2 that the given problem is performance-based rather than competence-based. However, Study 3 goes beyond simply supporting this general position. It offers an empirically underpinned framework for the problem's appearance / disappearance. This makes the performance account comparably in-depth as its competence-based competitor. Taken together, this is clearly a great stride in defence of the four-year-revolution against the overestimation claim.

9.3. Implications for the overestimation claim

Taken together all three studies I have conducted aimed to defend the standard picture of ToM development against differently motivated aspects of the overestimation claim. I have shown for both areas (aspectuality and TB understanding) that children from four years on are able to succeed on both once tasks are suitably modified.

In the case of aspectuality understanding I have presented work, supporting the initial findings of Rakoczy et al. (2015). Namely, that children had rather performance- than competence-problems in former studies. Study 1 investigated aspectuality in a new way and confirmed the findings of Rakoczy et al. (2015). Additionally, the paradigm used in Study 1 offers the possibility to investigate aspectuality understanding in implicit ways in order to explore the early mindreading ability's scope and limits.

In a more detailed way, I investigated children's TB competence. Experiment 1 in Study 2 disclosed an even more surprising age-related pattern than former studies. Only 10-year-olds were able to succeed on FB and TB tasks. However, the other experiments on TB understanding clearly support a performance-based explanation for this pattern. Children's pragmatic assumptions about belief discourse seem to overlay their real competence. This seems to be a problem, especially when they are asked direct trivial test questions about beliefs without the possibility to find an alternative explanation for the test question's triviality. It remains open which role test questions usage plays. Future studies are needed in order to explore this.

What remains from the critical findings of limited explicit ToM competence in children older than 4 years of age, are performance problems caused by extraneous task demands. Therefore, the empirical basis of the overestimation claim collapses. In the light of these results, the standard picture of ToM development is warranted to persist. Additionally, my work has some implications for the underestimation claim as well.

9.4. Implications for the underestimation claim

Besides the overestimation claim, the 4-year-revolution has been blamed by the opposite, namely underestimating children's real competence. How is my work relatable to this claim? In the following I will compare both claims and suggest implications of my findings for the underestimation claim.

The overestimation claim was motivated by children's performance shown in different areas (aspectuality and TB understanding). Similarly, the underestimation claim is motivated by findings of children's implicit competences found in studies using a broad range of methods (e.g. VoE, AL). In both cases this variety can be (and for the implicit findings it has been) interpreted as evidence for the effect's robustness (see Carruthers, 2013).

Especially in the case of TB understanding, my data does not only show that the problem can be overcome when extraneous task demands are adopted. Moreover, it shows how robust the initial findings of incompetence are. In several studies, the performance problem persisted even though I varied superficial aspects that were justified candidate causes. This is comparable to sets of studies on implicit ToM using the same measure in superficially different settings and finding comparable results. For

AL for instance, several studies with different implementations of the procedure show similar findings (Kovacs et al., 2010; Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Surian et al., 2007; for an overview see Baillargeon et al., 2010). This can be interpreted as indicating a robustness comparable to findings showing that explicit ToM is robust and insensitive to superficial modifications of the task structure (see Wellman et al., 2001).

In spite of findings of robustness, performance-based explanations are sufficient in order to explain limitations of children's explicit competence. It is reasonable, that something similar is possible for the underestimation claim. Heyes (2014), already critically reviewed implicit findings and offered alternative low level explanations for children's successful implicit behaviour. These alternative explanations do not involve any kind of belief ascription. The main argument against her approach is that there are only local low-level alternative explanations that cannot explain whole classes of findings. However, alternative explanations could experience a revival due to a growing body of non-replications (mainly unpublished) (for an overview see Kulke, L. & Rakoczy, H., 2017). If the replicability of implicit findings is limited, this questions their robustness and validity. Therefore, low-level explanations for the initial positive findings should be reconsidered.

A low level of replicability may be less surprising, given the complex structure of studies measuring children's eye-movement. For this is kind of a fragile dependent measure, it may seem plausible, that a high degree of the specific now how is needed to be able to capture these implicit indicators of children's early competence. However, even in the case of studies using less implicit methods (helping paradigms and referential communication), replicability seems to be a problem. For referential communication there is already one published non-replication (Wiesmann, Friederici, Singer & Steinbeis, 2017). Children's failure in this task does not seem surprising, given the complexity of the conclusions children are asked to make. Children are expected to take into account the FB of a communication partner in an act of referential communication, which is not exactly non-complex per se. It is rather impressive that children are able to do this in the original study. Similarly, in case of helping behaviour there are several studies supporting the existence of children's early competence ((Buttelmann et al., 2009; Fizke et al., 2014). However, for these studies there is also a promising low-level explanation which has not been ruled out yet. An alternative low-level explanation seems legit because of the complex cognitive process that it imposed on children in order to solve the task. Recall that in this task the FB procedure goes like the following. The protagonist places an object in box 1 and leaves the scene. In his absence the object is transferred to box 2. Upon his return the protagonist tries to open box 1 and needs the child's help. When the child helps the protagonist by opening box 2, this is considered as indicating that the child thought something like the following: "He tries to open box 1 because he thinks the object is in there. He wants the object, therefore I help him by leading him to the object in box 2.". Even though this is already complex, the analyses for TB seem even more challenging. In this case the protagonist saw the object's transfer but still tries to open box 1 upon his return. In this case, the child is expected to have the following chain of thoughts: "he is trying to open box 1 but he knows the object is in box 2, therefore he probably wants to do something else with box 1, and therefore I help him by opening box 1." Unpublished data by Rakoczy and colleagues (Rakoczy, Funken, & Oktay-Guer, 2016) shows that when presented with the scenario even adults do not show a comparable chain of thoughts for the TB version. A less mentalist explanation for children's successful behaviour, inspired by the PAR account by Fabricius et al. (2010), could be the following. Instead of implicitly reasoning about beliefs, children may reason about the agent's perceptual access. In the case of FB this means that because the agent lacked perceptual access to the change of location, he is getting things wrong. A reasonable way of helping is doing the opposite of what he is doing (helping to open box 2 when he is trying to open box 1). However, in TB when he had perceptual access to the change of location, he is getting things right. A reasonable way of helping is supporting him in his action (helping to open the targeted box). The present studies on implicit ToM using helping paradigm cannot differentiate between belief and perceptual access reasoning because they have the same predictions. One possibility to distinguish between these options is to use three instead of two possible locations. When presented with three boxes instead of two, only implicit ToM predicts that children would help to search in the correct box in FB. From a PAR point of view, however, children should help to search somewhere else than the box targeted by the protagonist because the protagonist is getting it wrong. Critically, PAR does not make any predictions about the correct way (concrete box) of helping. If children fail to show correct helping behaviour in a three options FB task, this clearly speaks against a (implicit) mechanism that operates with beliefs.

Taken together, what I present here is a much more fragile picture of implicit ToM competences. Given the parallelism between the empirical basis of the underestimation and overestimation claims presented above, my work can be an impulse in order to re-evaluate the underestimation claim. This re-evaluation can probably lead to the same conclusion as in case of the overestimation claim, namely that the standard picture of ToM development can again be successfully defeated.

10. References

Anscombe, G. E. M. (1957). *Intention*. UK: Basil: Oxford.

Apperly, I., & Robinson, E. J. (1998). Children's mental representation of referential relations. *Cognition*, *67*(3), 287–309. https://doi.org/10.1016/S0010-0277(98)00030-4

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

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of cognitive neuroscience*, *16*(10), 1773–1784. https://doi.org/10.1162/0898929042947928

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

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

Butterfill, S. A., & Apperly, I. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: the performance of children and great apes. *Child development*, *70*(2), 381–395.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S.,. . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning: evidence from five cultures. *Psychological science*, *16*(5), 378–384. https://doi.org/10.1111/j.0956-7976.2005.01544.x

Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014

Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y

Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, *194*(3), 673–688. https://doi.org/10.1007/s11229-015-0792-3

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Developmental psychology*, *33*(5), 781–789.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Dennett, D. C. (1978). Beliefs about beliefs [P&W, SR&B]. *Behavioral and Brain Sciences*, *1*(04), 568. https://doi.org/10.1017/S0140525X00076664

Fabricius, W. V., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: do 5-year-olds understand belief? *Developmental psychology*, *46*(6), 1402–1416. https://doi.org/10.1037/a0017648

Fizke, E., Butterfill, S. A., van der Loo, L., & Rakoczy, H. (2014). Signature limits in early theory of mind: Toddlers spontaneously take into account false beliefs about an object's location but not about its identity. *Unpublished Manuscript*.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*. (44), 283–296.

Goldman, A.I. (1992). In defense of the simulation theory. *Mind Language*. (7), 104–119

Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child development*, *59*(1), 26. https://doi.org/10.2307/1130386

Gordon, R.M. (1986). Folk psychology as simulation. *Mind Language*. (1), 158–171

Grosse, G., & Tomasello, M. (2012). Two-year-old children differentiate test questions from genuine questions. *Journal of child language*, *39*(1), 192–204. https://doi.org/10.1017/S0305000910000760

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Harris, P. (1992). From simulation to folk psychology: the case for development. *Mind Language*. (7), 120–144

Hedger, J. A., & Fabricius, W. V. (2011). True Belief Belies False Belief: Recent Findings of Competence in Infants and Limitations in 5-Year-Olds, and Implications for Theory of Mind Development. *Review of Philosophy and Psychology*, *2*(3), 429–447. https://doi.org/10.1007/s13164-011-0069-9

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

Heyes, C. (2014). False belief in infancy: a fresh look. *Developmental science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

Kaufman, A., & Kaufman, N. (1999). *Kaufman Assessment Battery for Children (4th ed.).* Franfurt am Main, Germany: Swets Test Services.

Kaufman, A., & Kaufman, N. *Kaufman AssessmentBattery for Children* (4th ed.). Frankfurt am Main, Germany.

Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental science*, *15*(1), 113–122. https://doi.org/10.1111/j.1467-7687.2011.01098.x

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kulke, L. & Rakoczy, H. (2017, January). *How robust and replicable are implicit Theory of Mind tasks? Criticism and alternative explanations.* Budapest CEU Conference on Cognitive Development, Budapest.

Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*(1-3), 211–238. https://doi.org/10.1016/0010-0277(94)90029-9

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in cognitive sciences*, *9*(10), 459–462. https://doi.org/10.1016/j.tics.2005.08.002

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive psychology*, *50*(1), 45–85. https://doi.org/10.1016/j.cogpsych.2004.06.002

Lewis, S., Hacquard, V., & Lidz, J. (2012). The samantics and pragmatics of belief reports in preschoolers. *Proceedings of SALT*. (22), 247–267.

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: limits on preschoolers' efficient ability to track perspective. *Child development*, *85*(4), 1519–1534. https://doi.org/10.1111/cdev.12224

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007

McKay, T., & Nelson, M. (2014). "Propositional Attitude Reports", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/>.

Newton, A. M., & Villiers, J. G. de. (2007). Thinking while talking: adults fail nonverbal false-belief reasoning. *Psychological science*, *18*(7), 574–579. https://doi.org/10.1111/j.1467-9280.2007.01942.x

Oktay-Gür, N. & Rakoczy, H. (2017). Children’s difficulty with true belief tasks: Competence deficit or performance problem? *Cognition, 166,* 28-41. https://doi.org/10.1016/j.cognition.2017.05.002

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

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: the acquisition of belief verbs. *Cognition*, *105*(1), 125–165. https://doi.org/10.1016/j.cognition.2006.09.008

Perner, J. (1991). *Understanding the representational mind* (1st MIT Press pbk. ed.). *Learning, development, and conceptual change*. Cambridge, Mass.: MIT Press.

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, *145*, 77–88. https://doi.org/10.1016/j.cognition.2015.08.006

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: key to children's understanding of belief. *Science (New York, N.Y.)*, *333*(6041), 474–477. https://doi.org/10.1126/science.1201216

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in cognitive sciences*, *16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Perner, J., & Ruffman, T. (2005). Psychology. Infants' insight into the mind: how deep? *Science (New York, N.Y.)*, *308*(5719), 214–216. https://doi.org/10.1126/science.1111656

Pham, K., Bonawitz, E., & Gopnik, A. (2012). *Seeing who sees: Contrastive access helps children rason about other minds.* Proceedings of the Thirty-fourth Cognitive Science Society.,

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515. https://doi.org/10.1017/S0140525X00076512

Rakoczy, H., Funken, I., & Oktay-Guer, N. (2016). *Investigating implicit Theory of Mind: Validation of the helping paradigm on adults.*

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind is even more unified than previously assumed: belief ascription and understanding aspectuality emerge together in development. *Child development*, *86*(2), 486–502. https://doi.org/10.1111/cdev.12311

Recanati, F. (2012). *Mental Files.*: Oxford University Press.

Russell, J. (1987). “Can we say …?: ” Children's understanding of intensionality. *Cognition*, *25*(3), 289–308. https://doi.org/10.1016/S0010-0277(87)80007-0

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of experimental psychology. Human perception and performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Schulz, A., Oktay-Guer, N., & Rakoczy, H. (2016). 2.5-year olds performance in aspectual and non-aspectual helping tasks. *Unpublished Data*.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child development*, *80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x

Scott, R. M., Baillargeon, R., Song, H.-j., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive psychology*, *61*(4), 366–395. https://doi.org/10.1016/j.cogpsych.2010.09.001

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental science*, *15*(2), 181–193. https://doi.org/10.1111/j.1467-7687.2011.01103.x

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: UK: Cambridge University Press.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science (New York, N.Y.)*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*(1), 1–12.

Song, H.-j., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology*, *44*(6), 1789–1795. https://doi.org/10.1037/a0013774

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

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. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Sprung, M., Perner, J., & Mitchell, P. (2007). Opacity and Discourse Referents: Object Identity and Object Properties. *Mind & Language*, *22*(3), 215–245. https://doi.org/10.1111/j.1468-0017.2007.00307.x

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

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*(3), 239–277. https://doi.org/10.1016/0010-0277(88)90021-2

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

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

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Westra, E. (2016). Talking about Minds: Social Experience, Pragmatic Development, and the False Belief Task. *Unpublished Manuscript*.

Westra, E., & Carruthers, P. (2016). The theory-of-mind scale: A pragmatic approach. *Unpublished Manuscript*.

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

Winner, E., & Gardner, H. (2012). Metaphor and irony: Two levels of understanding. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 425–444). Cambridge [u.a.]: Cambridge Univ. Press. https://doi.org/10.1017/CBO9781139173865.021

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and “false” photographs. *Cognition*, *35*(1), 41–68. https://doi.org/10.1016/0010-0277(90)90036-J

Appendix A: Detailed event sequences of the Numerical Location Change and Aspectuality tasks in Study 1

1 Warm-up

If the testing takes place in a day care, the experimenter introduces herself to the child in the child's group. After a short ice-breaking talk or game, the experimenter askes the child to play another game in the testing room. If the testing takes place in the laboratory, the experimenter picks up the child and its family in the entrance hall of the department and accompanies them to the rooms of the department. After a short ice breaking game the experimenter askes the child to play another game in the testing room.

2 Verbal ability

Vocabulary test (subscale of the Kaufman Assessment Battery for Children; Kaufman & Kaufman, 1999). The experimenter announces that she brought a picture book, suggests to look and that book and explains that she will show the child the pictures and it has to name the objects shown on the pictures.

3 Introduction of the protagonist

The experimenter says, "I brought someone who really wants to play with us, do you want to see him?" and takes out the first protagonist, e.g. the rabbit. The rabbit says: "Hello 'name of the child', I am the rabbit and I really want to play with you!" The child is allowed to touch the puppet and the experimenter shows the child the home of the puppet.

3 Introduction of the boxes/box

The experimenter shows the child and the protagonist two boxes: "Look I have two boxes here. Do you want to check if there is something in this one?" She handles the child the empty box first. After the child announces that the box is empty, the experimenter rattles the second box, showing that this box contains something and handles it to the child saying: "And what about this one?" When the child opens the box the experimenter says, "Look, the box contains ‘objects, e.g. green blocks’. We need two for the game we are going to play.”

4 **Introduction of the game/ Warm-up trial**

After the child takes out two green blocks, the experimenter takes them and starts explaining the game: “Look, ‘child’ and rabbit, the game we are playing now goes like this. I take one green block, put it in the middle first and then into this [the empty] box. How many blocks are in this box now? [Correct answer 1]” If the child answers correctly, the rabbit repeats the answer. If the does not answer correctly, the child is allowed to open the box and check the content. After the child’s final correct answer the game continues. The experimenter says: “Okay, now we take the other green block and put it in the middle first and now into the [target] box, too. How many blocks are in the box now?” After the child gives the correct answer, the two blocks are taken out of the box and a new round/trial begins.

5 **Test trials**

The test trials starts like the warm-up trial. The experimenter takes the first object, places it in the middle first and puts it in the target box. She again asks, “how many blocks are in the box now?” and continues with the second block. But right after the experimenter puts the second block in the middle, the protagonist says, „OH no! I forgot something in my house; I have to go home for short. I will be back soon.” The experimenter responds: “Okay, rabbit. I already put the object in the middle and we will wait for you to continue with the game.” After the rabbit leaves the scene, the experimenter explains that the puppet cannot hear them and suggests playing a trick on him.

5 a/b	**Aspectuality task:** The experimenter takes the block from the middle, puts it to the initial box containing all the blocks, and replaces it in the middle by the first object from the target box. And asks the first control question: “Does the rabbit know that we put the block from the middle back to this box and took the other one out of the other box and replaced it?”	**Numerical Location Change task:** The experimenter takes the block from the target box and moves it to the initial box. The object in the middle remains untouched. And the experimenter asks the first control question: “Does the rabbit know that we took the block out of this box and put it back to the initial box?”
5 c	**Upon the protagonist’s return** The experimenter announces that they waited for the rabbit and the game continues by the experimenter putting the object from the middle into the target box.	
6	**Test questions** After putting the object into the box, the experimenter holds the ears of the protagonist and askes the following control and test questions: Control Question 2; repetition of Control Question 1 [correct answer: no] Control Question 3:“How many objects are in box 1? [the target box] [correct answer: 1]” Test Question: “How many objects does the rabbit think are in the box?” [Correct answer: 2].	

Appendix B: Analyses only for children mastering all control questions (N = 37) and only including tasks with all control questions correct in Study 1

The consistencies in performance of children over trials 1 and 2 of each task were high (Φ = .84 in the Location Change task; Φ = .90 in the Aspectuality Task and Φ = .44 in the Numerical Location Change Task). Therefore, sum scores of trials solved correctly per task [0-2] were computed for further analyses. The mean values of these sum scores in the different tasks are depicted in Figure B1. First, in order to test whether the tasks differed in difficulty, a univariate ANOVA with task as factor was conducted but did not reveal any effect, (F(2,72) = 1.18 , p = .31). Comparisons against chance performance showed that children gave the correct answer significantly more often than expected by chance in all tasks (Standard Location Change Task, t(36) = 4.99 , p <.001, d = .82; Aspectuality Task, t(36) = 6.10, p <.001, d = 1.00 and Numerical Location Change Task, t(36)= 8.93, p <.001, d = 1.46)(see Figure 6). Correlation of the sum scores of correct answers in each task is depicted in Table B1.

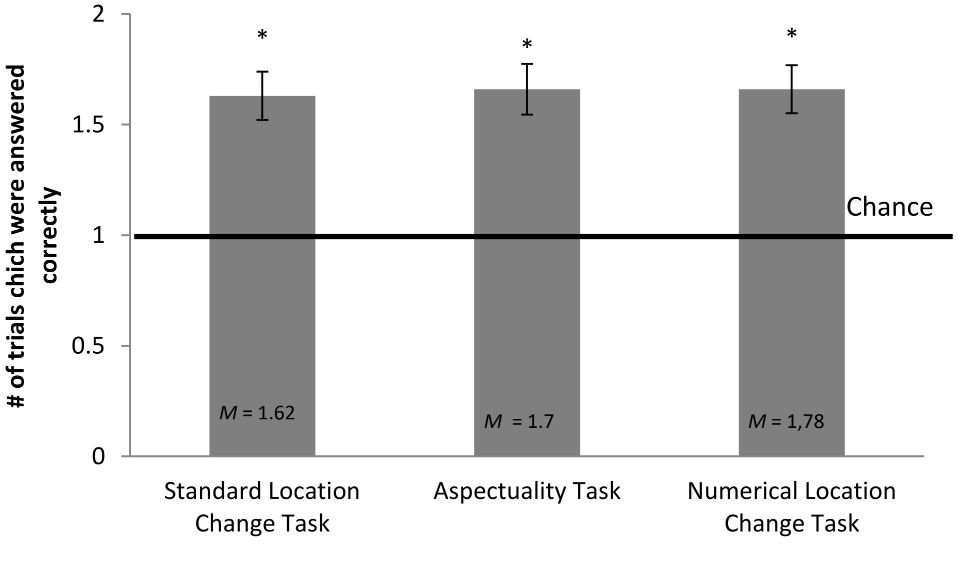


Figure B1. Mean number of trials answered correctly in the different tasks in Study 1 (only children mastering all control questions) (*p < .05).

Table B1

Correlations (and Partial Correlations Correcting for Age and Language Ability in brackets) between the different tasks in Study 1 (only children mastering all control questions).

	Aspectuality Task	Numerical Location Change Task
Standard Location Change Task	.51** (.45*)	.34** (.21)
Aspectuality Task		.86** (.85**)

* *p* < .01 ; ** *p* < .001

Analyses only for tasks with all control questions correct. The consistencies in performance of children over trials 1 and 2 of each task were high (Φ = .83 in the Location Change task; Φ = .84 in the Aspectuality Task and Φ = .66 in the Numerical Location Change Task). The mean number of trials (0-2) in which children answered the test question correctly as a function of task type is depicted in Figure B2 (Standard Location Change Task *M* = 1.63, *SD* = 0.74; Aspectuality *M* = 1.66, *SD* = 0.73 and Numerical Location Change task *M* = 1.66, *SD*= .70). Comparisons against chance performance showed that children gave the correct answer significantly more often than expected by chance in all tasks (Standard Location Change Task, *t*(45) = 5.77 , *p* <.001, *d* = .85; Aspectuality Task, *t*(40) = 5.79, *p* <.001, *d* = 0.91 and Numerical Location Change Task, *t*(41)= 6.08, *p* <.001, *d* = 0.95)(see Figure 7).

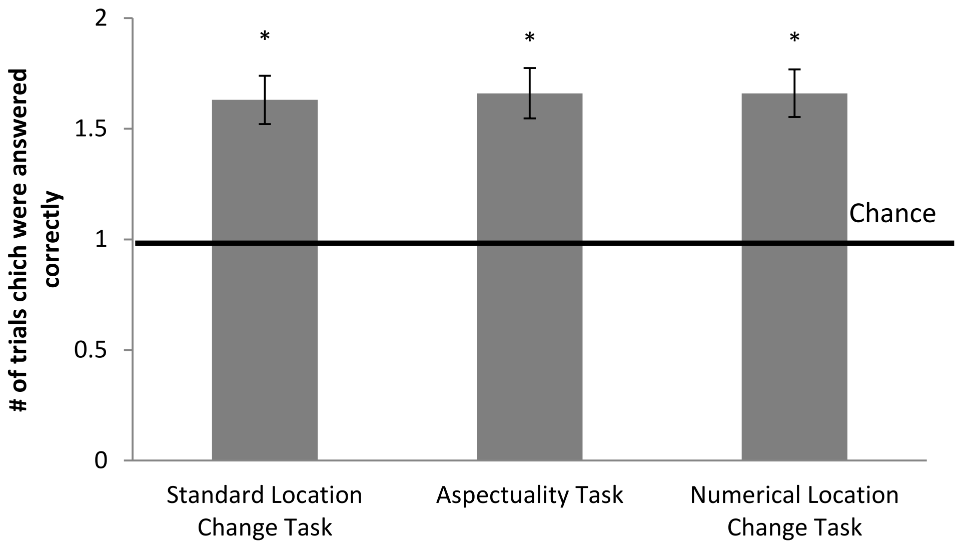


Figure B2. Mean number of trials answered correctly in the different tasks in Study 1 (only children mastering all control questions) (**p* < .05).

Appendix C: Analyses only for children mastering all control questions (N = 19) in Experiment 2 in Study 1

Analyses only for children mastering all control questions (N = 19). The consistencies in performance of children over trials 1 and 2 of each test question were high (Aspectuality task identity question Φ = .72 and number question Φ = 1.00; Numerical Location Change task location question Φ= .44 and number question Φ = 1.00). Therefore, trials 1 and 2 per test questions were combined to yield sum scores [0-2]. In addition, within each trial we computed an *aggregate score* that took into account whether children solved both the identity/location and the number question. A given trial received the *aggregate score* “correct” only if children answered both questions correctly (with a chance level of guessing correctly of 1/4). The mean sum scores for the different tests questions as well as the mean sum of aggregate scores across trials 1 and 2 of a given type of task are depicted in Figure C1as a function of conditions. First, in order to test whether there were differences between tasks or test questions, a 2 (Aspectuality vs. Location Change task) x 2 (question: identity/location vs. number) ANOVA was conducted on the mean sum of correct trials. This analysis yielded no main effect of task (Aspectuality vs. Location Change, $F(1,18) = 1.36$, $p =$ 26), and no main effect of test questions ($F(1,18) = 2.94$, $p = .10$), and no interaction effect ($F(1,18) = 0$, $p = 1$) between the factors. Correlations between the tasks are depicted in Table C1. The first test questions were not correlated (Identity and Location, $r = .31$, $p > .05$) but the second test questions were (Number Questions, $r = 1.00$, $p < .001$). We also aggregated new scores indicating that children solved both test questions within a trial of a task (called aggregate scores).

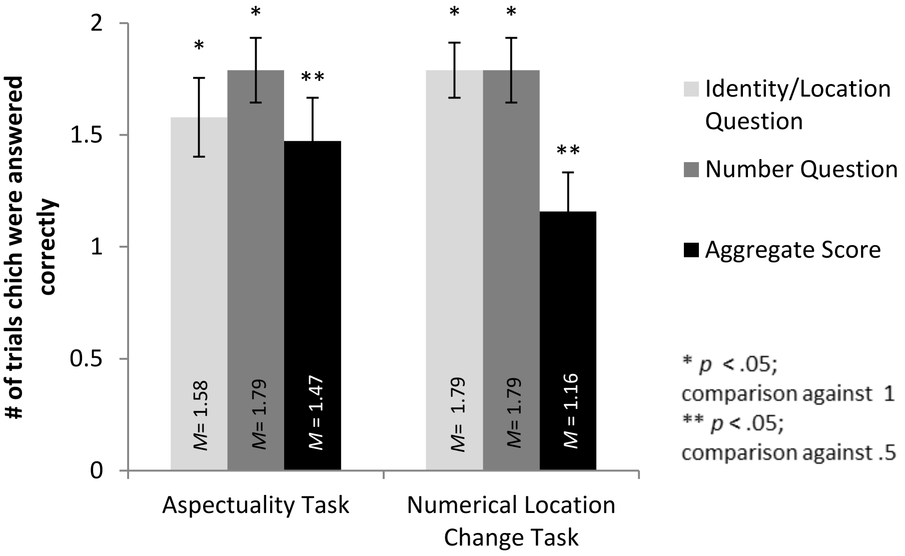


Figure C1. Mean number of trials answered correctly in the different tasks in Study 2 (only children mastering all control questions).

Table C1. *Correlations (and partial correlations correcting for age and language ability in brackets) between the different tasks in Study 2 (only for children mastering all control questions).*

Correlations Aspectuality and Numerical Location Change Task		
Identity/ Location Questions	Number Questions	Aggregate Scores
.31 (.25)	1.00* (1.00)*	.66* (.64)*

* *p* < .01,

Appendix D. Control Analyses for Study 3

In this control analysis, only the data of those children (N = 63, *M* = 68 months, 32 female) were included who answered all control questions correctly. The consistency in performance of children over trials 1 and 2 of the same type of task and belief was high for all conditions (Φs > .42). Therefore, again sum scores of trials answered correctly per condition [0-2] were used for further analyses.

The mean sum scores of trials in which children answered TB questions correctly and the sum scores of trials in which they answered FB questions correctly as a function of task type are depicted in Figure A1.

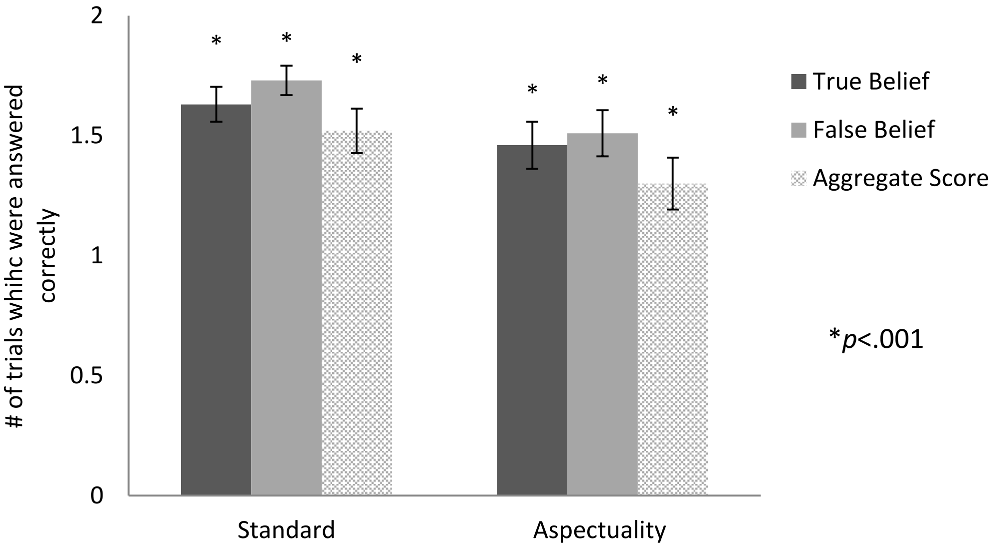


Figure D1. Mean number of trials in which true and FB questions were answered correctly and aggregate scores as a function of task type. [note that the chance level of guessing correctly differed between TB/FB (chance level = 50%, i.e. 1) and the aggregate score combining both measure (chance level = 25%, i.e. 0.5)]

A 2 (belief type: TB/FB) x 2 (task type: standard change-of-location/aspectuality) repeated measures ANOVA on these mean sum scores of correct trials yielded a main effect of task type (*F*(1,62) = 4.84, *p* < .05 , ηp^2 = .08), no main-effect of belief type (*F*(1,62)= 1.43, *p*=.24) and no interaction (*F*(1,62) = .356, *p* = .55). To test for children's competence as a function of task type separate planned comparisons against chance were conducted. These analyses revealed that these children performed significantly above chance on all TBs (standard TB, M=1.63, *t*(62) = 8.74, *p* < .001, *d* = 1.09 and aspectuality TB, *M* = 1.46, *t*(62) = 4.69, *p* < .001, *d*=.59) and FBs (standard FB, *M* =

1.73, *t*(62) = 12.02, *p* < .001, *d* = 1.52 and aspectuality FB, *M* = 1.51, *t*(62) = 5.31, *p* < .001, *d* = .67). FB and TB performance overall was highly correlated for both tasks (standard, *r* = .74, controlled for age and language ability, *r* = .72 and aspectuality, *r* = .58, controlled for age and language ability, *r* = .55

We computed aggregate scores that took into account whether children solved both TB and FB within a given trial. The sum aggregate scores (of trials in which children answered both TB and FB questions within a given trial) as a function of condition are depicted in Figure A1 as well. An ANOVA for task type (standard/aspectuality) on these mean aggregate scores revealed that there was a main effect of task type (*F*(1,62) = 6.13, *p* < .05, ηp^2 = .09), children performed better on the standard task than on the aspectuality task.

Post-hoc tests against chance showed that children performed above chance on both of the tasks (standard, *M* = 1.60, *t*(62) = 14.35, *p* < .001, *d* = 1.80 and aspectuality, *M* = 1.33, *t*(62) = 7.85, *p* < .001, *d* = .99) (see Figure A1). Furthermore, aggregate scores for the standard and the aspectuality tasks were correlated (*r* = .32, *p* < .05; controlled for age and verbal ability *r* = .24, *p* = .06).

Appendix E. Control Experiment using fewer trials for Experiment 2 in Study 3.

In order to control for the effect of the amount of trials children received, by removing the warm-up-trials. I tested seven-teen additional 4- and 5-year-olds (*M* = 59, 51-70 month olds). However, this manipulation had a negative effect on children's FB performance. Children's mean score on TB was 1.41, which was significantly above chance (*t*(16) = 2.38, *p* < .05, *d* = .58), while the mean score on FB was .47, which was significantly below chance (*t*(16) = -3.04, *p* < .01, *d* = -.74). For these preliminary results did not indicate an effect of reducing the amount of trials, this was not further investigated.

Appendix F. Control Experiment using new rewards for Experiment 2 in Study 3

In order to control for the effect of that children may have lost interest because they won all interesting stickers before the test trials, I tested nine additional 6-year-olds (*M* = 79, 73-84 month olds). In this experiment, children received a new set of stickers before the test trials began and therefore were more interested in winning them. However, this manipulation did not have a positive effect on children's FB performance. Children's mean score on TB was 1.67, which was significantly above chance (*t*(8) = 4.00, *p* < .01, *d* = 1.34), while the mean score on FB was 1.22, which was again at chance (*t*(8) = 1.00, *p* = .35). For these preliminary results did not indicate an effect of introducing new rewards before the test trials, this was not further investigated.

Curriculum Vitae

Personal Data

Name	**Nese Oktay- Gür, née Oktay**
Date of Birth	September 30th, 1989
Place of Birth	Türkeli, Sinop, Turkey
Citizenship	Turkish
Marital Status	Married

Academic Career

Since 04/2014	PhD-student in the BeCog-program of the University of Goettingen (GAUSS, Grundprogramm Biologie) Thesis supervisor: Hannes Rakoczy
04/2014 – 04/2017	Research scientist at the Department for Cognitive Developmental Psychology, University of Göttingen
04/2014	M. Sc. Degree in psychology, University of Göttingen
11/2013 – 12/2013	Research assistant in the Courant Research Centre for text structures, University of Göttingen
10/2013 – 03/2014	Student tutor for “Evaluationsmethoden”- classes, University of Göttingen
11/2012	B.Sc. degree in psychology, University of Göttingen

12/2010 – 09/2013	Student assistant at the department “Kulturen, Migration und psychische Krankheiten” at the Asklepios Hospital Göttingen
10/2010 – 03/2011	Student tutor for “Sozialpsychologie II”- classes, University of Göttingen

Teaching experiences

Spring 2016	M.Ed. seminar on diagnostics, assessment and support
Spring 2015	M.Ed. seminar on diagnostics, assessment and support
Winter 2014/2015	M.Ed. seminar on diagnostics, assessment and support
Spring 2014	B.Sc. seminar on developmental psychology

Scientific Papers

Oktay-Gür, N., & Rakoczy, H. (2017). *Investigating the roots of children’s true belief performance-problem.* Manuscript in preparation.

Oktay-Gür, N. & Rakoczy, H. (2017). Children’s difficulty with true belief tasks: Competence deficit or performance problem? Cognition, 166, 28-41. https://doi.org/10.1016/j.cognition.2017.05.002

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2016). *Children exhibit different performance patterns in explicit and implicit theory of mind tasks*. Manuscript submitted for publication.

Posters/ Talks

Oktay-Gür, N., Wenzel, L., & Rakoczy, H. (2017). The understanding of false and true belief in two-year-old children – Issue of competence or pragmatics? CEU conference on cognitive development, Budapest, January 5th -January 7th. [Poster]

Oktay-Gür, N., & Rakoczy, H. (2017). Children succeed in True Belief if Failure is Costly. CEU conference on cognitive development, Budapest, January 5th -January 7th. [Poster]

Oktay-Gür, N., & Rakoczy, H. (2017). The Role of Pragmatic Factors in Children's True Belief Competence. CEU conference on cognitive development, Budapest, January 5th -January 7th. [Poster]

Oktay-Gür, N., & Rakoczy, H.(2016). Is true belief a problem to 4- to 6-year olds? CEU conference on cognitive development, Budapest, January 7th -January 9th. [Poster]

Oktay-Gür, N., & Rakoczy, H. (2015). Competence depends on the right question. Conference of the German Educational Psychology Society, Frankfurt, September 14th – September 16th. [Poster]

Oktay-Gür, N., & Rakoczy, H. (2015). Theory of Mind und der Umgang mit Aspekthaftigkeit. Conference of the German Developmental Psychology Society, Frankfurt, August 31st - September 2nd [Talk]

Oktay-Gür, N., & Rakoczy, H. (2015). When they pass standard false belief tasks, children also master identity related numerical tasks. CEU conference on cognitive development, Budapest, January 8th -January 10th. [Poster]