



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com



Original Articles

# Children exhibit different performance patterns in explicit and implicit theory of mind tasks

Nese Oktay-Gür, Alexandra Schulz, Hannes Rakoczy\*

Institute of Psychology, University of Göttingen, Germany

## ARTICLE INFO

## Article history:

Received 16 January 2017

Received in revised form 19 December 2017

Accepted 3 January 2018

Available online xxx

## Keywords:

Theory of mind

Two-systems theory

Signature limits

Aspectuality

## ABSTRACT

Three studies tested scope and limits of children's implicit and explicit theory of mind. In Studies 1 and 2, three- to six-year-olds ( $N=84$ ) were presented with closely matched explicit false belief tasks that differed in whether or not they required an understanding of aspectuality. Results revealed that children performed equally well in the different tasks, and performance was strongly correlated. Study 3 tested two-year-olds ( $N=81$ ) in implicit interactive versions of these tasks and found evidence for dis-unity: children performed competently only in those tasks that did not require an understanding of aspectuality. Taken together, the present findings suggest that early implicit and later explicit theory of mind tasks may tap different forms of cognitive capacities.

© 2017.

## 1. Introduction

One of the biggest puzzles in recent theory of mind (ToM) research is this: how can we reconcile decades of findings that children fail explicit verbal false belief (FB) and related ToM tasks before age 4 with a growing body of evidence that even infants can perform successfully in implicit versions of such tasks?

## 1.1. The puzzle

In standard verbal FB tasks children are required to make explicit predictions of a protagonist's action on the basis of her mistaken belief. In change-of-location scenarios, for example, the child witnesses a protagonist put an object into box 1. In the protagonist's absence, the object is then transferred to box 2, and the child has to predict where the protagonist will search for it. Children younger than 4 years of age tend to fail in this task by claiming that the protagonist will look in box 2 while older children pass by predicting that she will mistakenly search in box 1 (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

Less explicit versions of such tasks, using looking and non-verbal interactive behavior as dependent measures have revealed competence much earlier than age 4. In violation-of-expectation looking time tasks infants look longer at an event if a protagonist performs an action which does not fit with her (false) belief (e.g., Onishi &

Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; see Baillargeon, Scott, & Bian, 2016; Baillargeon, Scott, & He, 2010 for review). Furthermore anticipatory looking studies show that two-year-olds form an expectation about the behavior of an agent based on her (false) belief (Clements & Perner, 1994; Southgate, Senju, & Csibra, 2007). Studies using helping behavior have revealed that infants and toddlers spontaneously help their interaction partners in ways that suggest that they are sensitive to the partners' beliefs (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Over, Carpenter, & Tomasello, 2014; Knudsen & Liszkowski, 2012a, 2012b; Southgate, Chevallier, & Csibra, 2010).

## 1.2. Theoretical responses to the puzzle: early competence versus conceptual change

How, then, might these two sets of findings be theoretically reconciled? From the point of view of early competence accounts (often nativist in spirit), the new findings with implicit measures suggest that the core competence for belief ascription operates from early on, is perhaps even innate, and does not itself undergo fundamental qualitative changes (Baillargeon et al., 2010; Carruthers, 2013; Leslie, 2005). The fact that children fail standard verbal FB tasks until several years later does not reflect, according to such views, any lack of conceptual competence or any significant conceptual development (the conceptual apparatus for belief ascription is present early and thus does not need to undergo substantial development). Rather, standard verbal tasks pose a number of extraneous task demands (in terms of inhibition, linguistic proficiency etc.) and thus mask children's early competence. Such tasks are then only mastered once children have acquired the requisite yet extraneous capacities (execu-

\* Corresponding author at: University of Göttingen, Institute of Psychology, Waldweg 26, 37073 Göttingen, Germany.

Email address: hannes.rakoczy@psych.uni-goettingen.de (H. Rakoczy)

tive function etc.) required to meet these task demands (Carruthers, 2013; Leslie, 2005).

Conceptual change accounts, in contrast, assume that there may be different forms and levels of conceptual capacities, with some crucial qualitative conceptual change from infancy onwards. The two kinds of tasks (early implicit and later explicit) may actually not tap the very same kinds of conceptual abilities. Rather, the implicit tasks may tap a more basic form of ToM that develops earlier and may constitute a foundation for the fully-fledged ToM capacities developing subsequently (e.g., Gergely & Csibra, 2003; Perner, 1991; Perner & Roessler, 2012; Wellman, 2011). If such a general picture was accurate, there should be clear differences between the scope and limits of the earlier and more basic, compared to the later-developing and more complex capacities. In particular, there should be signature limits of the early-developing capacities: Agents operating only on the basis of these more basic capacities should be able to master only a sub-set of simpler ToM tasks while failing more complex ones. A recent two-systems-account, in particular, makes clear, theoretically motivated and testable predictions of specific signature limits (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). According to this account, there are at least two systems for tracking mental states – in analogy, for example, to the widely shared assumption that in the domain of numerical cognition there are at least two systems for tracking numbers (e.g. Carey, 2009; Feigenson, Dehaene, & Spelke, 2004). The capacities tapped in implicit tasks reflect the workings of a simpler, evolutionarily and developmentally more ancient system (System 1, S1) that operates fast and independently of central cognitive resources (such as language or executive function). In contrast, the capacities tapped in explicit tasks reflect the workings of System 2 (S2) that develops later, is dependent on language and executive function, and operates with the fully-fledged conceptual grasp of subjective mental (mis-)representation. More specifically, the two systems differ crucially in their representational capacities in the following ways: S1 enables a subject to track so-called *relational attitudes*. These are relations that agents hold to situations like registering (in the sense of: being in perceptual contact with) an event. S1 allows an agent to represent, for example, that from his vantage point Peter can see (is in perceptual contact with) the cake on the table whereas Paul and Mary, from their perspective, cannot. On the basis of S1, an agent can thus engage in level-I perspective-taking. Importantly, though, keeping tracking of what another agent has or has not registered does not yet involve understanding a crucial form of the subjectivity of mental representation, namely their so-called *aspectuality*. Mental (and linguistic) representations are aspectual in the sense that agents represent objects (e.g., Clark Kent, who in fact is Superman) and situations always only under some aspects (e.g., “Clark Kent”) and not under others (e.g., “Superman”). Imagine, for illustration, that Peter (ignorant of the Clark Kent=Superman identity) witnesses the following sequence of events: First, he sees Clark Kent enter the house; then he sees Superman exit the house and fly to the beach. In order to understand what Peter believes about Clark Kent’s whereabouts, we need to take into account how he has represented the events in this sequence. De facto, he has seen Clark Kent first enter and then leave the house. But crucially, he only saw the person entering the house as Clark Kent. The person leaving was not represented as Clark Kent, but only as Superman. In consequence, Peter believes that Clark Kent must still be in the house.

S1, with its restriction to the representation of relational attitudes such as registering an event, does not enable agents to make such fine-grained distinctions regarding the question under which aspects an agent has encountered an object. If an agent has registered Clark Kent leaving the house, she has ipso facto registered Superman leaving

the house. Registration is not aspectual. S2, in contrast, recruits fully-fledged propositional attitude concepts like “belief” which are inherently aspectual: Ascription of a belief about a given object to an agent is sensitive to the aspects under which the agent subjectively represents the object in question. To ascribe the belief “Clark Kent is at home” to an agent is fundamentally different from ascribing to her the belief “Superman is at home”.

Empirically, these differences in the representational repertoire of the two systems should thus manifest themselves in distinctive and differential patterns of performance. S1 should have characteristic signature limits such that on the basis of this system, agents can master (only) those FB tasks that can be solved by tracking agents’ purely relational attitudes. This will apply to tasks for which it is not strictly required to grasp the aspectuality of mental representation. Level-I-perspective-taking tasks fall into this class, for example. In such tasks one merely has to track *whether* someone has seen an object or not, but not how she has seen that object. Similarly, many simpler change-of-location FB tasks fall into this class, too. Here, subjects only have to keep track of *which* events a protagonist has or has not registered (and not *how* she has represented these events). In a classical change-of-location FB task (Wimmer & Perner, 1983), the protagonist at time 1 puts an object into box 1, which is then at time 2 transferred in her absence to box 2, and the crucial question is where the agent, upon return at time 3, will search for her object. In order to solve this task, an infant may only need to represent that the agent at time 1 registers (stands in perceptual contact with) O in box 1. Since the protagonist subsequently does not register any other or competing information, this registration is not updated, and thus the infant can predict at time 3 that the agent will act on the basis of this registration. But since registration is a relational attitude, it does not allow the infant to distinguish how the protagonist may have represented the object and thus would not allow mastery of tasks that would require such a more fine-grained understanding.

S2, in contrast, should not be subject to such signature limitations. Rather, it should enable the mastery of a great variety of tasks the common denominator of which is that they require an understanding of mental representation and its aspectuality. That is, subjects operating with S2 should be able to solve standard change-of-location FB tasks just as much as more complex tasks that require explicit representation of aspectuality (such as answering the question “Where does Peter believe Clark Kent is now?”). While there should thus be divergence and dis-unity in different types of implicit FB tasks (such that young children consistently master those FB tasks that do not strictly require an understanding of aspectuality but fail those that do), for explicit FB tasks there should be convergence and unity (such that all kinds of such tasks begin to be mastered at the same time and in correlated fashion).

### 1.3. The empirical situation so far

Turning first to young children’s performance in implicit FB tasks of various types and topics, is there any evidence for disunity and disassociation? The empirical situation so far is complex. On the one hand, some studies suggest that infants and toddlers are able to solve some implicit FB tasks that require an understanding of aspectuality around the same time that they master implicit non-aspectual change-of-location FB tasks (Buttelmann, Suhrke, & Buttelmann, 2015; Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015). For example, in Buttelmann et al. (2015), a protagonist reached toward an object with misleading appearances (e.g. an A that looked like a B). In some cases she was aware of the real nature of the object and thus knew that it was an A that only looked like a B (TB condition)

whereas in other cases she was not aware of the true nature of the object and thus took it by its appearance as a B (FB condition). Infants then, in some cases, and for some sub-sample of the stimuli, helped the protagonist differentially in TB and FB conditions (they tended to give her another B-object in the TB condition more often than in the FB condition, and tended to give her another A-object in the FB condition more often than in the TB condition).

These studies taken by themselves, however, are very difficult to interpret. One reason is that all of them have used a single isolated vignette each of which leaves room for alternative, more parsimonious explanations, either in low-level terms (Heyes, 2014a, 2014b) or in terms of children's tracking belief-like states rather than fully-fledged aspectual beliefs (e.g., Butterfill & Apperly, 2013). In the absence of additional data, the interpretation of such results remains ambiguous and the validity of these tasks difficult to establish. One way to disambiguate these findings would be tests for convergent validation: We know from numerous studies that performance across many various explicit ToM tasks reveals clear convergence and correlation and thus evidence for the convergent validation of the individual tasks (Gopnik & Astington, 1988; Perner & Roessler, 2012). So far, there are hardly any correlational data on infants' performance across implicit tasks, and the only published study suggests that performance in different types of implicit ToM tasks does not show the systematic unity and correlation typical of explicit tests (Yott & Poulin-Dubois, 2016). Another way to disambiguate the infant tasks purportedly showing an understanding of aspectuality would be to validate them in analogous yet explicit form with older children and adults. This form of validation is given for standard change-of-location scenarios implemented as violation-of-expectation (Onishi & Baillargeon, 2005) or anticipatory looking tasks (e.g. Southgate et al., 2007) that are closely modeled on existing explicit tasks (Wimmer & Perner, 1983). For the infant studies interpreted as showing an appreciation of aspectuality, in contrast, since they have used novel and very peculiar isolated vignettes, there are so far no explicit versions of the vignettes with older children and/or adults that would corroborate their interpretation by showing that adults indeed understand the scenarios along the lines of the task analysis.<sup>1</sup>

Another line of recent research, in fact, has begun to systematically implement this very rationale to use the same or structurally analogous aspectuality tasks in implicit form with younger children and contrast this with older children's/adults' performances in explicit tasks. In a comprehensive series of studies, Low and colleagues tested children at age 3 and 4 years and adults in implicit (anticipatory looking) and explicit FB tasks that do or do not involve aspectuality (Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013; Wang, Hadi, & Low, 2015). These studies have shown that competence in explicit change-of-location scenarios and in modified scenarios that involve aspectuality strongly converges: 3-year-olds consistently failed both non-aspectual and aspectual versions of explicit FB tasks, whereas 4-year-olds and adults consistently mastered both types of tasks. In anticipatory looking versions of the tasks, in contrast, all age groups showed signs of tracking the agent's belief in the non-aspectual change-of-location FB task, but not in the aspectual task version – suggesting clear signature limits in the early developing System 1.

<sup>1</sup> In fact, data from a recent study aiming just at this question suggest that adults may not see the events in the Scott and Baillargeon (2009) according to the task analysis at all (Low & Edwards, 2017). Similarly, it was recently found that older children, in contrast to younger ones, do not pass the Buttelmann et al. (2009) tasks (Buttelmann & Buttelmann, 2015) – putting into doubt the original task analysis and the very validity of this task.

However, the interpretation of the results by Low and colleagues in terms of signature limits has recently been subject to some debate (Carruthers, 2013, 2015, 2016; Csibra, 2012; Jacob, 2012). Critics have been concerned whether the differences in performance in the aspectual FB vs. standard FB tasks may have been due to performance factors (such as working memory demands, or the fact that only the aspectual FB task involves certain forms of mental rotation and simulation).

In a similar approach as Low and Watts (2013), two preliminary studies have recently contrasted older children's performance in explicit verbal task with younger children's performance in implicit helping tasks in false belief scenarios that do or do not involve aspectuality (Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Rakoczy, Bergfeld, Schwarz, & Fizke, 2015). The results, similar to those of Low and colleagues, suggested convergence between various kinds of explicit tasks, but signature limits in early implicit ToM such that children performed successfully in non-aspectual while failing aspectual FB helping tasks. However, these results are also difficult to interpret given the way the contrast between aspectual and non-aspectual FB tasks was implemented. The non-aspectual FB tasks were standard change-of-location FB tasks in which an object was transferred from one box to another in the presence (TB) or absence (FB) of the protagonist. The aspectual task, in contrast, differed from the non-aspectual one in many respects. An object with two identities (reversible soft toys, e.g. a bunny on one side, and a carrot once turned inside out) was put into one box under its one identity. It was then revealed in the presence (TB) or absence (FB) of the protagonist that the object had another identity as well. And finally, the object, under its second identity, was transferred to the other box, and the crucial test question was asked where the protagonist would think the object under its first identity was. One problem with the aspectual task is that it is unclear whether these stimuli are really good instances of objects with dual identities (in fact, in some sense the objects are neither a carrot nor a bunny). Secondly, the two purported identities of the objects always go along with perceptual differences, and thus, theoretically, one could keep track of them in merely perceptual ways. More generally, however, the most fundamental problem is that the contrast between aspectual and non-aspectual conditions was confounded in many ways with differences in complexity, and was thus far from a desirable minimal contrast.

#### 1.4. Rationale of the present study

In sum, existing research presents complex patterns of partly inconclusive evidence concerning unity and disunity of implicit and explicit forms of ToM. Against this background, the rationale of the present study was to test the predictions of the two-systems-theory with a novel, comprehensive and stringent design, the first one based on minimal contrast pairs between non-aspectual tasks (which should be mastered both implicitly and explicitly) and closely matched aspectual ones (which should only be mastered in explicit form). To this end, structurally analogous aspectual and non-aspectual versions of FB tasks were designed and equated in terms of performance factors as far as possible: Children watched scenarios that involved several qualitatively identical, perceptually indistinguishable objects, in which the protagonist at some point formed a false belief about the number of objects present in a box. In both conditions, the basic task for the participant was to keep track of the protagonist's belief as to how many objects are in the box. The crucial difference between aspectual and non-aspectual conditions was how the protagonist arrived at this (false) belief: in the non-aspectual condition she failed to witness a transfer of an object and thus formed a false belief (assuming

there were two objects in a box when in fact there was only one). In the aspectual condition, the protagonist arrived at the same mistaken belief, but in crucially different ways: she failed to witness the transfer of an object *as* transfer of this very object (she, in fact, saw the same object enter a box twice, but did not see it *as* the same object) and thus formed a corresponding false belief. That is, children could pass the non-aspectual condition by keeping track of what the protagonist has or has not registered. The aspectual condition required participants to make the very same judgment in the end (concerning the protagonist's belief as to how many objects are in the box). However, in contrast to the non-aspectual condition, participants could arrive at this judgment only by reasoning in aspectual ways about *how* the protagonist has seen the objects. Importantly, this need to keep track of *how* the protagonist has represented the objects is now disentangled from any perceptual differences in the objects' appearances. Rather, and more simply, what one has to keep track of is *how* the protagonist represents the objects *indexically* ("this object here now", in relation to "that object that was here before" etc.).

Different measures were used for the explicit and implicit tests: in the explicit version, in Studies 1 and 2, children were verbally asked about the belief of the protagonist, whereas in the implicit version in Study 3, their spontaneous helping behavior (following Buttelmann et al., 2009; Southgate et al., 2007) served as dependent measures. The two-systems-theory would predict convergence and unity in the explicit versions of the FB tasks (such that children pass or fail both non-aspectual and aspectual versions), but dis-unity and dissociation in the implicit versions (children pass the non-aspectual but fail the aspectual versions).

## 2. Study 1

### 2.1. Method

#### 2.1.1. Participants

Fifty-three- to six-year-olds (twelve three-year-olds, seventeen four-year-olds, eighteen five-year-olds and three six-year-olds; range: 38–72 months;  $M=57$ ,  $SD=9.7$ ; 22 girls) 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 participate in experiments. Children were tested by a female experimenter (E) either in an appropriate room in their daily childcare or in the lab.

#### 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 location of the boxes were counterbalanced (see Appendix A for details).

##### 2.1.2.1. Verbal ability

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

##### 2.1.2.2. Standard false belief task (SFB)

Each child was tested in two trials of a standard location change false belief task (after Wimmer & Perner, 1983). The child and the protagonist, were shown two boxes and an object. The object was hidden in one of two boxes (box 1) and the protagonist left the scene.

The experimenter (E) suggested playing a trick<sup>2</sup> on the protagonist who was absent and transferred the object to the other box (box 2). When the protagonist returned, control questions (CQ1 Location 1: "Where did we put [the object] in the beginning?" and CQ2 Location 2: "Where is it now?") and the test question (TQ: "When the protagonist wants [the object], where will he look for it?" [correct answer: "box 1"]) were asked. Each child received two trials with two different objects. The order of the objects and the location of box 1 were counterbalanced.

##### 2.1.2.3. Aspectual task (AT)

In addition, children were tested in a new aspectuality task where the identity of an object caused a false belief in the protagonist about the number of objects hidden in a box (see Fig. 1; for details see Appendix A). The child and the protagonist 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 (E) and the protagonist were going to play together.

###### 2.1.2.3.1. Introducing the rules of the game

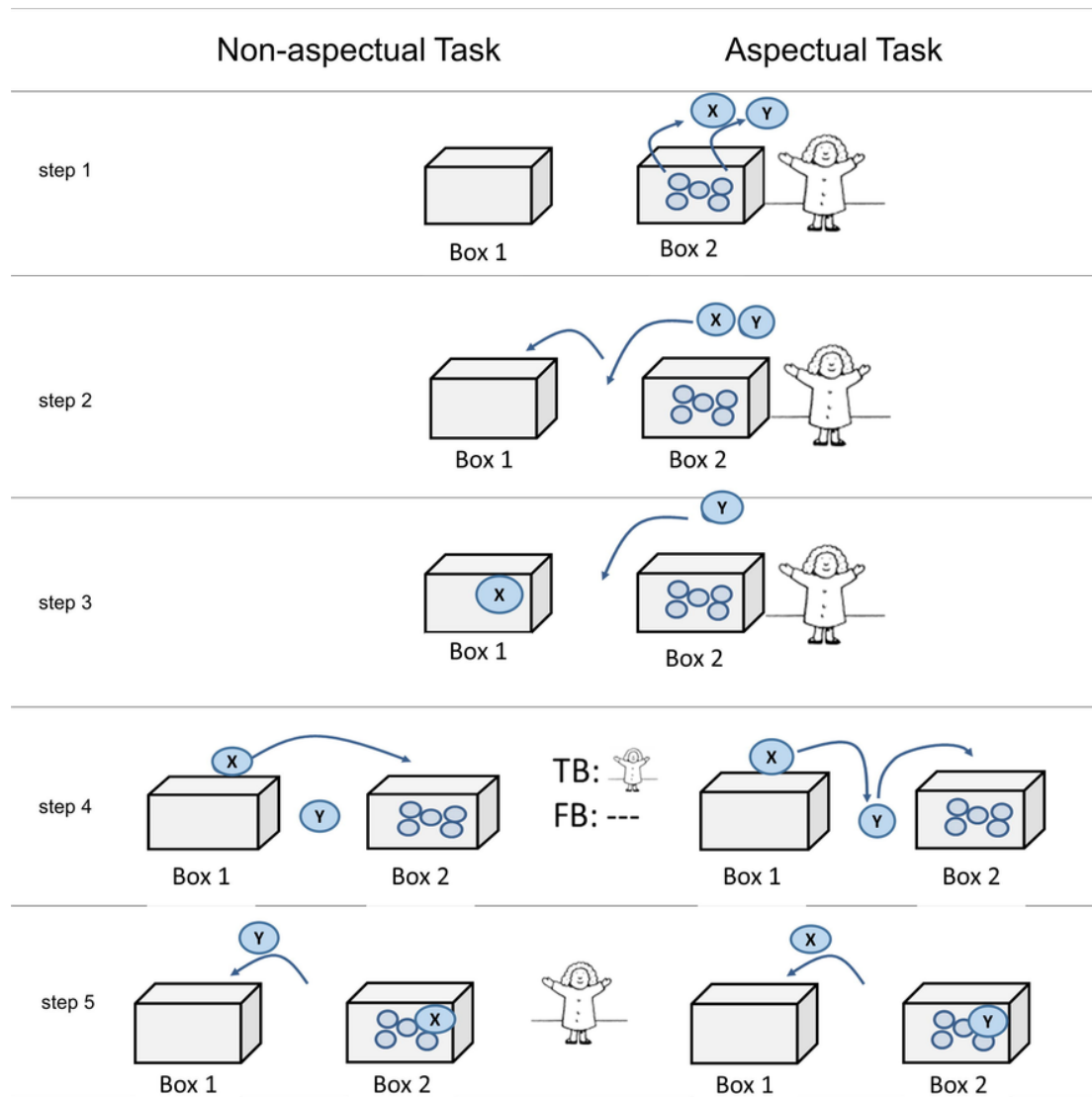
E explained that in this game an object was first put in the middle (between the two boxes) and then in box 1. After doing so, the child was asked how many objects were in the box. When the child gave the right answer ["one"], the game continued and the experimenter put the second object in the middle and then in box 1. E again asked how many objects were in box 1 [correct answer: "two"]. If the child gave any incorrect answer at any time, the experimenter opened box 1 and allowed the child to count the contents. 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.

###### 2.1.2.3.2. Critical test trial

The test trial started as described above by E putting the first object into box 1 (steps 1 and 2). But before E could ask the first number question, the protagonist claimed she forgot something and had to leave. In response, E put the second object in the middle (step 3) and said they were going to wait for the protagonist to return in order to continue with the game, upon which the protagonist left. In the absence of the protagonist, the experimenter suggested the following trick: the object from the middle was put back to box 2 (step 4). The object from box 1 was taken out and replaced the former object from the middle. The first control question (CQ1: "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?" correct answer: "no") was asked. Upon the protagonist's return, this manipulation resulted in a false belief on the part 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 (step 5). Now E covered the ears of the protagonist with her hands and asked the remaining control questions and the test question:

- CQ2: Does the protagonist know that we exchanged the objects when he was absent?" [correct answer: "no"]

<sup>2</sup> Acting out the transfer of the object in change-of-location false belief tasks in ostensibly deceptive ways has been shown to be helpful to younger children in some studies (Wellman et al., 2001).



**Fig. 1.** Procedure of the Aspectual and Non-aspectual Tasks used in Study 1. (Note that box 1 contains a collection of qualitatively identical objects that cannot be perceptually discriminated; “X” and “Y” are only marked here for the reader but were visually indistinguishable to children and the protagonist.)

- CQ3: How many objects are in box 1? [correct answer: “one”]
- Test question: How many objects does the protagonist think are in the box? [correct answer: “two”].

If they failed to answer a control question, they were asked again (at most twice) and corrected after the second repetition.

#### 2.1.2.4. Non-aspectual task (NAT)

This task was designed to test for children’s performance in a structurally analogous false belief task, that was closely matched in terms of complexity and task demands to the AT but differed in the one crucial respect (see Fig. 1 and Appendix A): In both tasks, the protagonist arrived at a false belief about the number of objects in a box, but in order to understand how this belief came about, NAT did not require an understanding of the aspectuality of the protagonist’s belief. In the crucial step 4 in the NAT, in the absence of the protagonist, E 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 false belief on the part of the protagonist upon her return concerning the identity of the object in the middle, but simply a false

belief about the location of the object formerly in box 1 and thus a false belief about the content of box 1. The following control and test questions were asked in this task:

- CQ1: Does the protagonist know that we put back that object? (Asked in the absence of the protagonist after putting the object from box 2 to box 1) [correct answer: “no”]
- CQ2: Does the protagonist know, that we put back that object in her absence? (Asked upon the protagonist’s return) [correct answer: “no”]
- CQ3: How many objects are really in that box? [correct answer: “one”]
- Test question. How many objects does the protagonist think are in the box? [correct answer: “two”]

## 2.2. Results

### 2.2.1. Control questions

Children answered 94% of all control questions correctly and 74% of the children (33% of the three-year-olds ( $n=4$ ), 77% of the

four-year-olds ( $n=13$ ), 94% of the five-year-olds ( $n=17$ ) and all six-year-olds ( $n=3$ ) answered all control questions correctly on the first request. On average, children answering all control questions correctly were 4;11 years old ( $SD=0;8$ ), while the mean age of children who failed at least one control question was 3;11 ( $SD=0;6$ ). Table 1 depicts the percentages of children solving the different types of control questions.

### 2.2.2. Main analysis

For the main analyses, data from all children were included (for supplementary control that take into account performance in control questions and that show converging results, see Appendix B). The consistency in performance of children over trials 1 and 2 of each task were moderate to high ( $\Phi=0.87$  in the Standard FB task;  $\Phi=0.85$  in the Aspectual Task and  $\Phi=0.60$  in the Non-aspectual 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 Fig. 2. 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 ( $F(2, 98)=0.34, p=.71$ ).

Second, children's performance was compared to chance performance (if children just guessed in both trials, chance performance would be 1). These analyses showed that children gave the correct answer significantly more often than expected by chance in all tasks (Standard FB Task,  $t(49)=4.21, p<.001, d=0.60$ ; Aspectual Task,  $t(49)=3.78, p<.001, d=0.53$  and Non-aspectual Task,  $t(49)=5.07, p<.001, d=0.71$ ). 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).

### 2.3. Discussion

Study 1 had three main results: First, children performed above chance in both the Aspectual as well as the Non-aspectual Task. Second, all of the tasks (AT, NAT, SFB) were equally difficult. And third, children's performances on the different tasks were strongly correlated.

These findings taken together seem to speak for the unity and convergence in explicit ToM competence predicted by the 2-systems view. However, one possibility regarding these findings is that they perhaps do not show that children can solve aspectual and non-aspectual FB tasks in analogous ways. Children may in fact be unable to understand the aspectuality of beliefs and have 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 ball" at time 1, and "another ball" at time 2) and thus arrived at a false belief ("there are two different balls in the box"). But

perhaps children here arrived at the correct solution in a much simpler way that has nothing to do with understanding the aspectuality of beliefs. Children may not have paid attention to the identity of the objects at all. Rather they may have used some kind of belief-book-keeping. 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 (along the following lines: he witnessed "+1", he did not witness the "-1", but then did witness the second "+1" again, therefore his belief is "+2").

Study 2, therefore, was designed to address this concern: If children were unable to understand the aspectuality of beliefs in principle, and were thus restricted to representing the Aspectuality Task in such simpler, non-aspectual ways, then they should be unable to explicitly ascribe to the protagonist beliefs about the identity of the object in the middle. If however, they are capable of solving 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.

## 3. Study 2

In this Study, the same closely matched aspectual (AT) and non-aspectual (NAT) belief ascription tasks as in Study 1 were used, but with one crucial modification: In addition to the test question concerning the protagonist's belief at the end, another test question was added that could not be solved by using such simpler alternative strategies (see below).

### 3.1. Method

#### 3.1.1. Participants

Thirty-four 4- to 6-year-olds (range: 51–80 months;  $M=62$ ; 17 girls) from mixed socioeconomic backgrounds were tested. Two additional children were tested but excluded from data analysis because 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 participate in experiments. No child from Study 1 participated in Study 2. Children were tested by a female experimenter in an appropriate room in their daily childcare or in the lab.

#### 3.1.2. Design and procedure

In a within-subjects design, children were tested in the Aspectuality and the Non-aspectual Tasks and received two trials of each task. The order of the tasks as well as the sides of the relevant boxes were counterbalanced across subjects.

##### 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.

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

# trials correct	Standard FB Task		Aspectual Task		CQ3	Non-aspectual Task		
	CQ1	CQ2	CQ1	CQ2		CQ1	CQ2	CQ3
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=2)	–	8% (N=4)	6% (N=3)	–	8% (N=4)	4% (N=2)	–

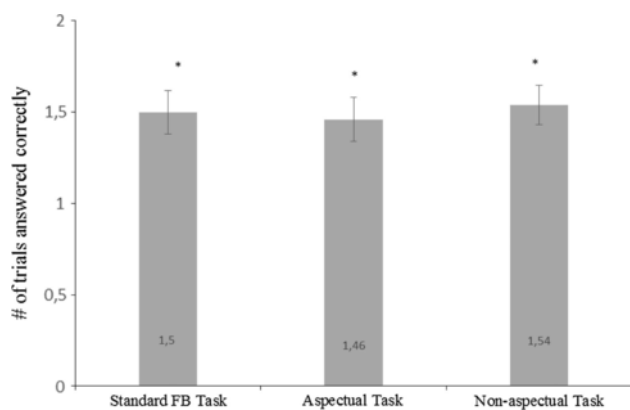


Fig. 2. Mean number of trials answered correctly as a function of task in Study 1 (\* comparison against chance (=1),  $p < .05$ ).

Table 2

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

	Aspectual Task	Non-aspectual task
Standard FB Task	0.58** (0.40 <sup>+</sup> )	0.50* (0.25 <sup>+</sup> )
Aspectual Task		0.88** (0.85**)

<sup>+</sup>  $p < .10$ .

\*  $p < .01$ .

\*\*  $p < .001$ .

### 3.1.2.2. Aspectuality task (AT)

This AT was the same as in the first experiment (see Fig. 1) with the following modifications:

- (1) one of the two objects used in the game was assigned to the child and the other one was assigned to 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 2 and it was replaced by the child's object from box 1
- (4) in the absence of the protagonist the child was asked the first control question (CQ1: "Whose object is the one in the middle?" [correct answer: "the child's"])
- (5) upon the protagonist's return the first test question (*identity test question*) was asked (T1: "Whose object does the protagonist think is in the middle?")
- (6) finally, children were asked two control questions (CQ2: "Does the protagonist know that we exchanged the objects in her absence?" and CQ3: "How many objects are in that box?") and the numerical test question ("How many objects does the protagonist think are in that box?")

### 3.1.2.3. Non-aspectual task (NAT)

This task was a modification of the task used in Study 1, with control and test questions equivalent to the ones used in the AT in Study 2. In the absence of the protagonist the object of the child from box 1 was moved to box 2 and the first control question was asked (CQ1: "Where is your object now?"). Upon the protagonist's return the first *location test question* (TQ1: "Where does the puppet think your object is?" [correct answer: "in box 1"]) was asked and the game went on analogously to the AT. Children were asked two additional control questions (CQ2: "Does the protagonist know that we moved your ob-

ject from box 1 to box 2 in her absence?") and CQ3: "How many objects are in that box?") and the numerical test question ("How many objects does the protagonist think are in that box?").

Children were directly corrected if they answered the second control question (CQ2), asking for the knowledge of the protagonist about the manipulation, incorrectly. Each child received two trials per task. Task order and side of Box 1 was counterbalanced between subjects.

## 3.2. Results

### 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 56% of the children ( $N=19$ ) consistently answering all control questions correctly. On average, children answering all control questions correctly were 5;6 years olds ( $SD=0;8$ ), while the mean age of children who failed at least one control question was 4;11 years ( $SD=0;7$ ). 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 (why children performed worse here on CQ2 than in Study 1 is currently unclear. One plausible possibility is that this may have been due to the generally more demanding set of control and test questions used in Study 2).

### 3.2.2. Main analyses

For the main analyses, data from all children were included (for supplementary control that take into account performance in control questions and that show converging results, see Appendix C). The consistency in performance of children over trials 1 and 2 of each test question was high (AT identity question  $\Phi=0.81$  and number question  $\Phi=0.82$ ; NAT location question  $\Phi=0.77$  and number question  $\Phi=0.90$ ). Therefore, trials 1 and 2 of each 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 (since chance level for each of the two questions was 1/2, the chance level for the aggregate score was 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 Fig. 3 as a function of conditions.

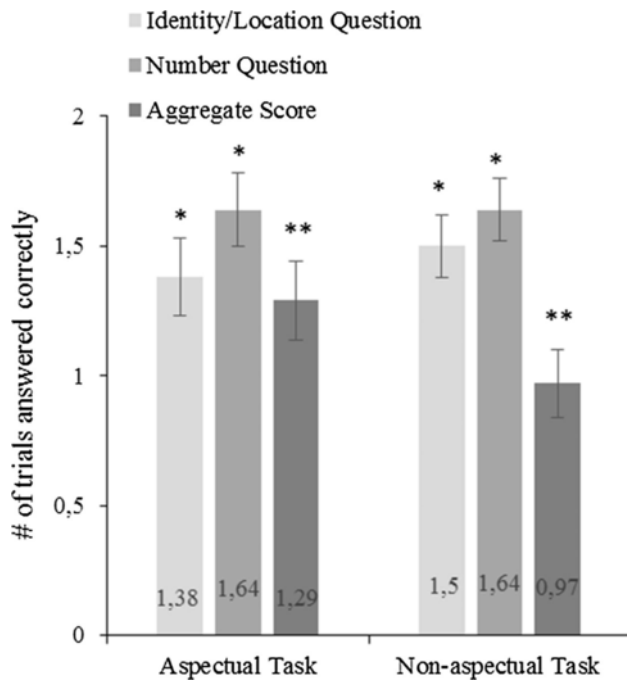
First, in order to test whether there were differences between tasks or test questions, a 2 (AT vs. NAT)  $\times$  2 (question: identity/location vs. number) ANOVA was conducted on the mean sum of correct trials. This analysis yielded no main effect of task (AT vs. NAT,  $F(1, 33)=0$ ,  $p=1$ ), a main effect of test questions (such that the number question was easier than the identity/location question,

Table 3

Control questions answered correctly (in%) in Study 2.

# of trials correct	Aspectuality Task			Non-aspectual Task		
	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)	–





**Fig. 3.** Mean number of trials answered correctly as a function of task and question type in Study 2. (\*  $p < .05$ , comparison against chance (=1); \*\*  $p < .05$ , comparison against chance (=0.5).

( $F(1,33)=5.38, p < .05$ ), 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,  $t(33)=2.51, p < .05, d=0.43$ ; and number question,  $t(33)=5.14, p < .001, d=0.88$ ; NAT location question,  $t(33)=3.53, p < .01, d=0.61$  and number question,  $t(33)=5.14, p < .001, d=0.88$ ). With regard to the aggregate score, children's performance was also significantly different from chance in the AT ( $t(33)=5.11, p < .001, d=0.88$ ) and NAT ( $t(33)=3.44, p < .01, d=0.60$ ).

Third, in order to analyze 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  $r=0.68, p < .001$ ; partial correlation, controlling for age and verbal ability,  $r=0.61, p < .001$ ) and for the NAT (Location and Number Question  $r=0.60, p < .001$ ; partial correlation, controlling for age and verbal ability,  $r=0.37, p < .05$ ). Performance on a given test question, and on both test questions per trial combined, also correlated substantially across the different tasks (see Table 4).

**Table 4**

Correlations (and partial correlations correcting for age and language ability) of performance in a given questions type and the aggregate scores between Non-aspectual and Aspectual tasks in Study 2.

Identity/Location Question	Number Question	Aggregate Scores
0.60** (0.51*)	0.89** (0.86**)	0.60** (0.53*)

\*  $p < .01$ .

\*\*  $p < .001$ .

### 3.3. Discussion

Study 2 replicated the main findings of Study 1: AT and the NAT did not differ in difficulty, children performed competently in both, and performance was strongly correlated across tasks. However, Study 2 also extended the results of Study 1 in crucial ways: The new test question revealed that children did pay attention to the object's identity, and the protagonist's corresponding representations. Children's performance showed convergence and unity across different FB tasks even with such explicitly aspectual questions. This clearly speaks against more parsimonious strategies of solving the AT without understanding aspectuality. Studies 1 and 2 together thus supply converging evidence for unity and consistency in performance across various explicit ToM tasks.

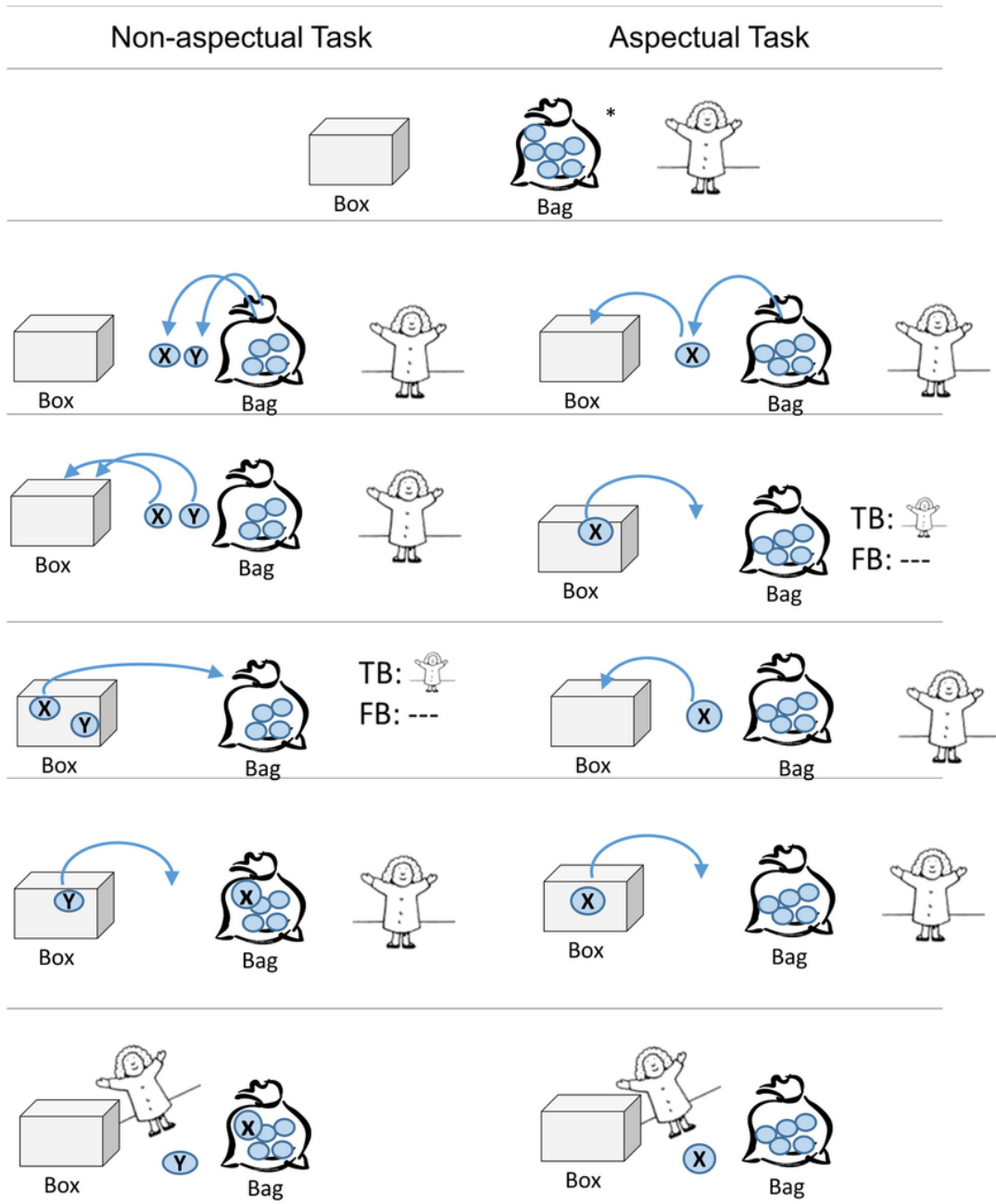
According to the 2-systems account, this unity and convergence found in explicit ToM in Studies 1 and 2 should contrast with patterns of dis-unity and dissociation in implicit ToM measures. Study 3 was designed to test this prediction. To this end, non-verbal analogues of the AT and NAT tasks were devised, building on the spontaneous interaction methodology used by Buttelmann et al. (2009). The central prediction of the 2-systems view would be that younger children in these tasks should solve the NAT but fail the AT.

### 4. Study 3

In this study, non-explicit false and true belief tasks were used that matched the explicit Aspectuality and Non-aspectual Tasks as closely as possible. These tasks build on an implicit FB study with spontaneous helping as dependent measure (Buttelmann et al., 2009). Children see a protagonist P put an object O into box 1 which is then transferred in her presence (TB) or absence (FB) to box 2. P then tries to open box 1, fails and looks in help-seeking manner toward the child. The underlying logic is the following: if children in the FB condition understand P's mistaken belief ("she thinks O is still in box 1"), they should help her by opening box 2 so that P can get O. In the TB condition, in contrast, children should think "She knows O is not in box 1 anymore, so she must be trying to open this box for some other reason" and thus help her by opening box 1. And indeed, the authors found that children responded differently in FB and TB conditions (they tended to open box 2 in FB and box 1 in TB).

For the present study, we build on this logic to create closely matched aspectual and non-aspectual versions of such helping tasks (see Fig. 4). The non-aspectual version was structurally analogous to the original Buttelmann et al. (2009) task, yet somewhat more complex in order to match it as closely as possible to the aspectual version: P put 2 objects from a bag into a box; then one object was removed and put back into the bag in P's presence (TB) or absence (FB); P then removed one object from the box in both TB and FB, resulting in her belief (just like in the original Buttelmann et al. (2009) study) that there was one of the original objects left in the box (FB) or that there was none of the original objects left in the box (TB). P then continued searching in the box. If children ascribe the corresponding beliefs in TB and FB, respectively, they should then respond differentially to the agent's attempt to open the box: in the TB condition they should reason "She knows none of the original objects is in the box anymore, so she must be looking for something else" and accordingly help her to open the box. In the FB condition, in contrast, they should reason "She is looking for one of the original objects that she thinks is still in the box" and accordingly help her by referring her to the bag and/or giving her one of the objects from there.





\* Note that the bag 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.

Fig. 4. Schematic event sequences of the implicit Non-aspectual and Aspectual Task used in Study 3.

The aspectual version also tests whether children arrive at these belief ascriptions and thus respond differentially. But the crucial difference is that they can arrive at this ascription only via ascribing aspectual beliefs to the agent: First, P put an object from the bag into the box. Then this object was taken from the box and put on the table in P's presence (TB) or absence (FB). P then put this object back into the box, but either expressing her belief that this was a different object in the FB ("Another one! I'll put it in here as well") or her true belief about the object's identity in TB ("I'm putting this one into the box again"). P then searched in the box and removed the object, thus resulting in her belief that there was one of the original objects left in the box (FB) or that there was none of the original objects left in the

box (TB). P then continued searching in the box. If children ascribe the corresponding beliefs in TB and FB, respectively, they should then respond differentially to the agent's attempt to open the box exactly like in the non-aspectual FB and TB conditions.

From a theoretical point of view, the crucial task analysis of the 2-systems-view is the following: The non-aspectual tasks can be solved by proper belief ascription, but they need not be so solved. Rather, simply tracking belief-like states such as registration is sufficient to differentiate TB and FB. The aspectual conditions, in contrast, strictly require the ascription of aspectual beliefs that take into account *how* the agent has seen a given object. The prediction that follows from this task analysis is that younger children should be able

to respond differentially in TB vs. FB in the non-aspectual conditions (like in the original Buttelmann et al. (2009) study), but fail to do so in the aspectual conditions.

#### 4.1. Method

##### 4.1.1. Participants

Eighty-one children (40 girls) were included in the final sample. ( $M=27.4$  months,  $SD=2.68$ ; range 23–33). One additional child was tested but excluded due to experimental error. Another twenty-two children (nine in AT, thirteen in NAT) were tested but excluded from the analysis because they gave ambiguous responses ( $N=7$ ), did not show any reaction ( $N=10$ ) or were uncooperative ( $N=5$ ).

##### 4.1.2. Design

Children were randomly assigned to one of four conditions resulting from a 2 (TB–FB)  $\times$  2 (AT–NAT) between subject design. Children received two trials (see below) (except 4 children who ended the session after one valid trial).

##### 4.1.3. Materials and procedure

Children were tested individually in an interactive play setting with two experimenters, closely modeled after Buttelmann et al. (2009). At least one parent was present during the session. Sessions began by playing two warm-up games (e.g., building a toy zoo). After the child interacted freely with the experimenters, the child, one parent and one experimenter moved to the table where the subsequent testing took place. The child sat on the parent's lap facing experimenter 1 (E1). Experimenter 2 (E2) hid behind a wall. Another two warm-up games (puzzle and picture book) were played. Parents were told not to interfere during the test trials.

For each task, a set that contained some qualitatively indistinguishable objects (toy ladybeetles or pigs) and a box (yellow or green) were used. The boxes had several openings and were covered with tissue on the inside – to make it appear plausible that an agent could be unsuccessfully looking for objects hidden and difficult to retrieve. Children were left ignorant about the initial content of the box (which was, in fact, empty).

##### 4.1.3.1. Spontaneous helping tasks

###### 4.1.3.1.1. Non-aspectual task

After the warmup, E2 brought out one of the bags containing several objects of the same kind (e.g. toy pigs), put it on one side of the table and placed two of the objects in the middle of the table. E1 took one of the objects and gave the other one to the child making sure that the child was not afraid of touching it. Then E1 announced that she had to leave. She took a box and placed it on the table on the opposite side of the bag while claiming that she wanted to put both of the objects in there. Before E1 left she put both objects one by one inside the box. The following sequences varied between the true and false belief conditions (see Fig. 4 and Appendix D):

- False Belief Condition (NAT\_FB): E1 left and in her absence E2 entered the situation and suggested to play a trick on E1. She took one of the objects out of the box and moved it to the bag containing the other objects and stressed that E1 could not see what she was doing.
- True Belief Condition (NAT\_TB): Before E1 left, E2 entered the scene and transferred one of the objects from the box to the bag by saying: “Look what I am doing now!”. E1 observed the change of location attentive, saying: “Hmm, aha, okay!”. And left the situation for about the same time as in the false belief condition.

After E1 returned she declared that she wanted to take the objects out of the box. Without naming them, she took out the first (and only one) and went on searching inside the box and said: “Huh? I don't understand. This is strange. Huh? Can you help me?”. If the child did not show any reaction, she repeated the help question at most two times. Children's reactions after E1 began searching was coded. The task was repeated with the other sort of object and the other box.

###### 4.1.3.1.2. Aspectuality task

Despite that only one object was used, the Aspectuality Task had the same procedure as the Non-aspectual task until E1 claimed that she had to leave and went on in the following ways:

- False Belief Condition (AT\_FB): E1 left and in her absence E2 entered the situation and suggested to play a trick on E1. She took the object out of the box and placed it in the middle of the table and stressed that E1 could not see what she was doing.
- True Belief Condition (AT\_TB): Before E1 left, E2 entered the scene and took the object out of the box and placed it in the middle. After attentively observing this, E1 left the scene.

After her return E1 either said “Oh, another pig! Now I've got two. I'll put this one into the box, too.” (AT\_FB) or said “Okay, I'll put this into the box again.” (AT\_TB). Then she announced that she would need to leave and that she wanted to take her objects out of the box and with her. She took the first object out and searched for the second one in the box like in the Non-aspectual Task.

###### 4.1.4. Coding

Sessions were coded from videotape. The child's reaction after E1 started searching was coded as belonging to one of the following categories.

- Help to search in box: if the child showed any reaction, like pointing, touching or trying to open or opening the box, behavior was coded as helping to search in box.
- Help to search in bag/give object: if the child showed any bag or object directed reaction, like pointing to the bag or the object, giving the object or searching in the bag, it was coded as helping to search in bag/giving the object.

Additionally a given trial could be coded as ambiguous or invalid if the following criteria were met.

- Ambiguous trial: showed a reaction to the searching behavior which did not fit either of the two categories described above because the child showed both behaviors simultaneously or did something totally irrelevant.
- Invalid trial: did not show any reaction, left the table during the trial, just said, “Yes.” as an answer to the helping question but did not show any additional behavior.

A second (blind to hypotheses) coder coded 35 tapes (33%) for reliability which was  $\kappa=0.86$  for the four category coding. In cases of mismatch the codes of the first coder were used.

#### 4.2. Results and discussion

For the final analysis, data of 81 children were used. Mean age was 27.61 months ( $SD=2.91$ ;  $n=21$ , 11 girls) in the NAT\_FB condition,  $M=28.05$  months ( $SD=2.92$ ;  $n=20$ , 10 girls) in the NAT\_TB condition,  $M=26.77$  months ( $SD=2.2$ ;  $n=20$ , 10 girls) in the AT\_FB condition and  $M=27.21$  months ( $SD=2.7$ ;  $n=20$ , 10 girls) in the AT\_TB condition. There was no significant age difference between conditions ( $F(3, 77)=0.84$ ,  $p=.48$ ).

#### 4.2.1. Main analyses

The main analyses focused on children's first valid and unambiguous trials. If the first trial was invalid or ambiguous, it was replaced by the second trial if this was a valid one. This analysis revealed that children's helping behavior, in line with previous findings by Buttelmann et al. (2009), differed significantly between FB and TB condition in the Non-aspectual Task ( $\chi^2(1, N=41)=14.752, p<.001$ ): Children were more likely to help the agents search in the box or give her the object in the FB than in the TB condition. In the Aspectuality task, in contrast, helping behavior did not differ significantly between FB and TB condition,  $\chi^2(1, N=40)=1.111, p=.29$  (see Fig. 5).

In a secondary analysis including ambiguous responses as a third category, the difference between conditions in the Non-aspectual Task remained significant,  $\chi^2(2, N=46)=14.967, p=.001$ , while, again FB and TB condition did not differ significantly in the Aspectuality task,  $\chi^2(2, N=42)=1.111, p=.574$ .

#### 4.2.2. Supplementary analysis on aggregate scores over two trials

In addition to the children who only received one trial ( $N=4$ ), a large number of children had at least one trial which has been coded invalid or ambiguous ( $N=11$  in the Non-aspectual Task;  $N=13$  in the Aspectuality task). Eight of those children (two in the Aspectuality task and six in the Non-aspectual Task) gave an ambiguous response in the first trial but then an unambiguous response in the second trial. There were thus 53 children who had two valid trials with unambiguous responses. The patterns of performance of these children over the two trials are depicted in Table 5. The distribution of these patterns differed between FB and TB conditions in the Non-aspectual Task ( $\chi^2(2, N=27)=6.3, p<.05$ ) but not in the Aspectuality task ( $\chi^2(2, N=26)=2.1, p>.05$ ).

Taken together, the present findings suggest that toddlers were able to take into account a partner's belief in the Non-aspectual Task in which they adapted their helping behavior between the belief conditions. In the Aspectuality Task, in contrast, they showed no indications of taking into account the belief of the protagonist. This disunity in children's behavior in the closely matched non-aspectual and as-

pectual tasks contrasts with the unity in analogous explicit tasks found in Studies 1 and 2.

## 5. General discussion

The aim of the present study was to investigate the scope and limits of implicit and explicit theory of mind. Two-system and conceptual change accounts predict that children's early implicit ToM capacities enable them to track others' belief-like epistemic states but not ascribe fully-fledged aspectual beliefs proper. And thus young children are expected to pass those implicit false belief tasks that do not require an understanding of aspectuality but fail those that do. Children's later-developing explicit ToM capacities, in contrast, operate with a full-fledged concept of belief and its aspectuality and should thus enable children to master all kinds of FB tasks in analogous and consistent ways.

These predictions were tested in a comprehensive design with a novel method: children witnessed scenarios in which a protagonist formed a false belief about the number of qualitatively identical objects in a box. In the non-aspectual condition this false belief came about since the protagonist merely failed to witness one event (transfer of an object) whereas in the aspectual condition the protagonist arrived at the same mistaken belief in crucially different ways: he did witness a given crucial transfer of an object, but failed to witness it *as* transfer of this very object (seeing "this" object as different from "that one" (previously seen) when in fact it was the very same object).

The main results were the following: In Study 1 we investigated the characteristics of our novel Aspectual task. It was shown that it was no more difficult than a closely matched Non-aspectual task or the Standard FB task. Furthermore, children's performances in the different tasks were substantially correlated. Study 2 ruled out a more parsimonious alternative explanation to the effect that children solved the aspectuality task in simpler, non-aspectual ways, and produced additional and converging evidence that children's competences on various FB tasks emerge together. Study 3 tested younger children in analogous implicit versions and found that toddlers were able to differentiate between non-aspectual false and true beliefs by adapting

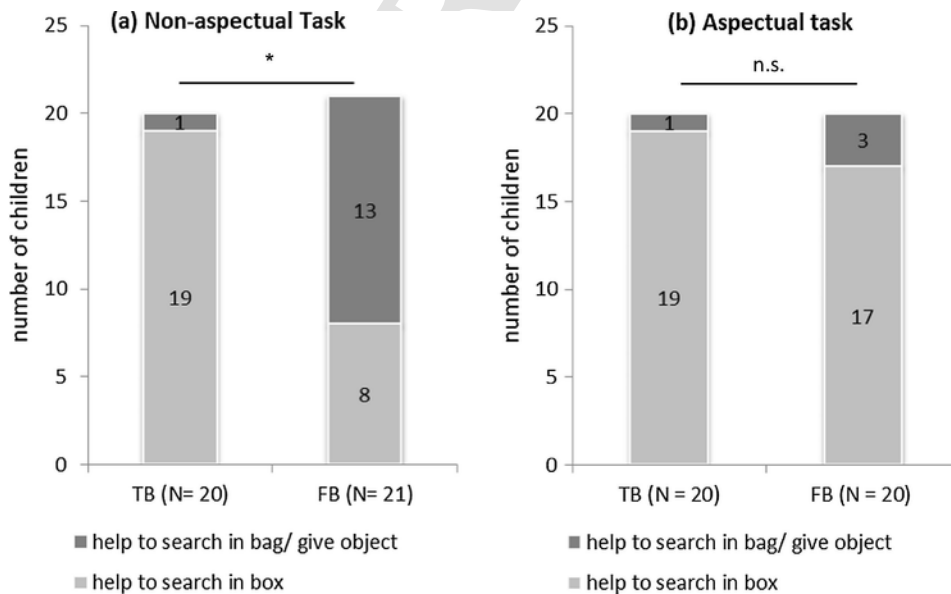


Fig. 5. Helping behavior in the first valid and unambiguous trial of the (a) Non-aspectual Task and (b) the Aspectual Task in Study 3 (\*  $p<.05$ ).

**Table 5**

Patterns of performance across both trials of those children (N=53) who completed two trials in Study 3.

	Non-aspectual Task		Aspectuality task	
	TB	FB	TB	FB
Consistently referred to bag/gave object	1	5	1	0
Consistently helped to search in box	13	5	14	10
Mixed pattern	1	2	0	1

their behavior to the belief of an interaction partner but did not do so in the aspectual version.

What is the upshot of these findings? First, the present results corroborate recent findings that, in contrast to previous assumptions, explicit aspectuality tasks, once suitably modified are not more difficult than standard false belief tasks (see Rakoczy et al., 2015). Second, the results provide converging evidence, in line with other recent findings, for signature limits in early implicit ToM (see Fiske et al., 2017; Low & Watts, 2013; Low et al., 2014; Wang et al., 2015).

But might there be alternative interpretations of the current findings that warrant different theoretical conclusions? One possibility is that the findings from the explicit aspectuality tasks present false positives such that children solve these tasks without understanding their aspectuality. While it may be theoretically possible, this alternative seems unlikely given that it was explicitly addressed and ruled out in Study 2 which probed children's ascription of aspectual beliefs directly. Conversely, the findings with the implicit tasks might be seen to reflect false negatives due to task demands or other performance factors (see Carruthers, 2013, 2016). This possibility cannot be ruled out conclusively here. But in light of the fact that we matched the implicit aspectual and non-aspectual tasks as far as possible, it is currently quite unclear, what candidate performance factors there may be left.

All in all, the present findings constitute *prima facie* evidence compatible with predictions of conceptual change and two-system accounts. However, many fundamental and important questions remain open: First of all, how robust and generalizable are the present findings? Do they extend more broadly to various kinds of test material, vignettes and measures? More specifically, do the present findings of signature limits in implicit ToM extend to other situations for which the two-systems view (and to some degree, conceptual change accounts) would predict analogous patterns? These include, in particular, situations in which adult subjects – capable of operating with fully-fledged belief ascription in principle – are expected to fall back to their simpler ToM capacities, such as situations of spontaneous (uninstructed) mindreading, and situations of dual-task performance. These questions need to be addressed in future research with the same kinds of scenarios used here with implicit tasks suitable for adults such as eye-tracking (along the lines of Low & Watts, 2013).

Second, how do the present findings (suggesting signature limits in young children's performance tasks) relate to positive findings suggesting that infants master all kinds of implicit ToM tasks, including aspectual ones? The current empirical picture is complex and puzzling. On the one hand, a number of studies have developed novel

vignettes and scenarios to test infants and toddlers on implicit looking time and interaction tasks and have produced evidence that infants pass ToM tasks that seem to require an understanding of aspectuality (Buttelmann et al., 2015; Scott & Baillargeon, 2009; Scott et al., 2015). But none of the tasks in these studies has been validated so far with older children and adults. On the other hand, a number of studies have administered the very same contrasts between aspectual and non-aspectual ToM tasks in implicit form with younger children and in explicit form with older children and adults (Low & Watts, 2013; Rakoczy et al., 2015 and Fiske et al., 2017; and the present study). These studies tend to find converging performance in explicit aspectual and non-aspectual ToM tasks, but dissociation and signature limits in implicit form (young children master non-aspectual but fail aspectual ToM tasks).

How can these seemingly inconsistent results be explained and reconciled? Currently, we do not know. Clearly, what is needed are more systematic and comprehensive designs in future research. Ideally, more comprehensive studies should test various types of scenarios, vignettes and tasks used in previous studies with positive and negative findings, and implement them in stringent, closely matched minimal contrast pairs between aspectual and non-aspectual versions in both implicit and explicit form. Such designs are necessary to gain more conclusive evidence whether young children's failure in aspectual ToM tasks, the crucial evidence for signature limits, rests merely on some performance factors pertaining to the specific tasks of the present and related studies, or rather reflect some more fundamental competence limitations of early implicit ToM.

If the latter turned out to be the case, the most exciting theoretical question for future research will be which of the different kinds of accounts in the general category of conceptual change and two-system views best explains these general patterns of findings. Conceptual change accounts suggest that early competences subsequently become replaced by later and more sophisticated ones. Two-systems-accounts, in contrast, assume that earlier and later developing systems of mindreading continue to operate in parallel. To be able to differentiate between these possibilities, experimental work with adults is needed to test whether there continues to be an implicit, automatic form of ToM with infant-like signature limits in operation across the lifespan – an issue currently very much debated (see Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Heyes, 2014a, 2014b; Phillips et al., 2015; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014).

## 6. Uncited references

Flavell (1988), Low et al. (2016), Perner and Howes (1992) and Sprung et al. (2007).

## Acknowledgements

We would like to thank Carina Neumann, Kira Sagolla and Lisa Wenzel for help with testing and coding. Thank you very much to Marlen Kaufmann and Konstanze Schirmer for the organization of the studies.

## Appendix A. Detailed event sequences of the Non-Aspectual and Aspectual Tasks in Study 1

- 1 **Warm-up**If the testing takes place in a daycare, the experimenter introduces herself to the child in the child's group. After a short ice-breaking talk or game the experimenter asks 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 asks the child to play another game in the testing room.
- 2 **Verbal ability**A vocabulary test (subscales of the Kaufman Assessment Battery for Children; Kaufman and Kaufman, 1999) was used. The experimenter announces that she brought a picture book, suggests to look at that book together and says, "Let's look at that book. I will show you some pictures and you will tell me what we see on these 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 protagonist and the experimenter shows the child the home of the protagonist.
- 4 **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 hands 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."
- 5 **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 child 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.

- 6 **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. Before she can ask for the number of objects in the box, the protagonist says: "Oh no! I forgot something in my house, I have to go home for a short while. I will be back soon." The experimenter responds: "Okay, rabbit. I put this 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 protagonist cannot hear them and suggests to play a trick on him.
- 6 a/ **Aspectual task:**The experimenter takes the block from the middle and 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?"
- b **Non-aspectual 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?"
- 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.
- 7 **Test questions**After putting the object into the box, the experimenter holds the ears of the protagonist and asks 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. Control analyses on the sub-sample of children who mastered all control questions (N=37) in Study 1

The consistencies in performance of children over trials 1 and 2 of each task were moderate to high ( $\Phi=0.84$  in the Location Change task;  $\Phi=0.90$  in the Aspectual Task and  $\Phi=0.44$  in the Non-aspectual 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 Fig. 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,  $t(36)=4.99, p<.001, d=0.82$ ; Aspectual Task,  $t(36)=6.10, p<.001, d=1.00$  and Non-aspectual Task,  $t(36)=8.93,$

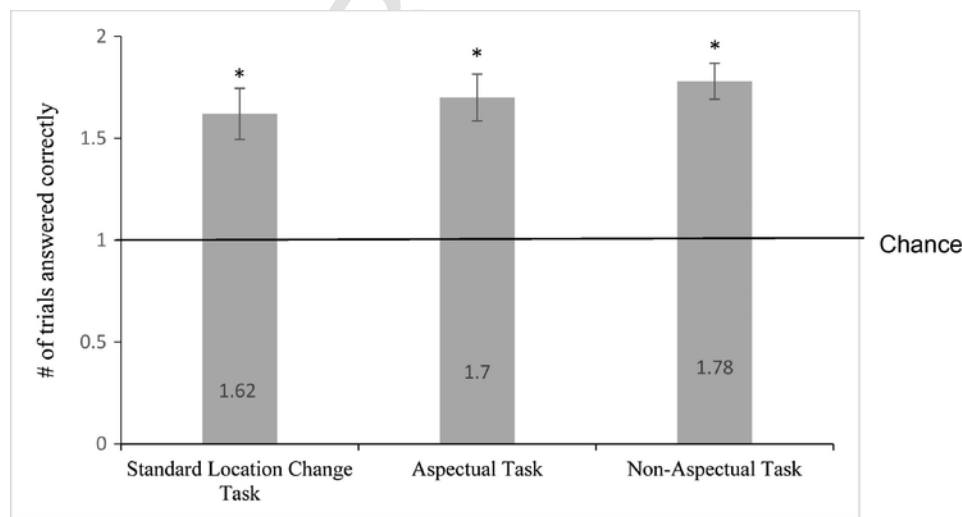


Fig. B1. Mean number of trials answered correctly in the different tasks in Study 1 (sub-sample of children mastering all control questions) (\* comparison against chance (=1),  $p<.05$ ).

$p < .001$ ,  $d = 1.46$ ) (see Fig. B1). Correlation of the sum scores of correct answers in each task is depicted in Table B1.

**Appendix C. Analyses only for children mastering all control questions (N=19) in Study 2**

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 (Aspectual Task identity question  $\Phi = 0.72$  and number question  $\Phi = 1.00$ ; Non-aspectual Task location question  $\Phi = 0.44$  and number question  $\Phi = 1.00$ ). 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 Fig. C1 as 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)  $\times$  2 (question: identity/locati-

**Table B1**  
Correlations (and Partial Correlations Correcting for Age and Language Ability in brackets) between the different tasks in Study 1 (sub-sample of children mastering all control questions).

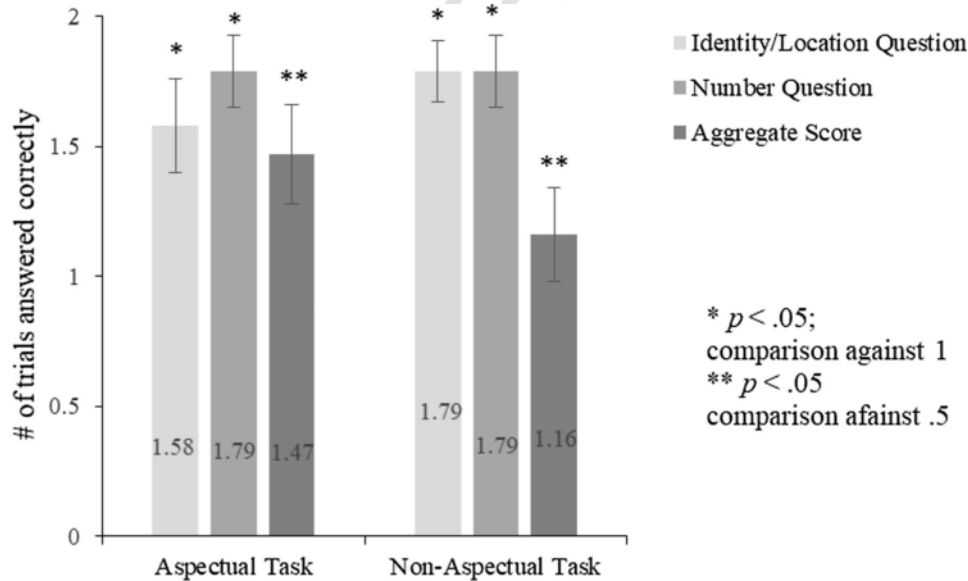
	Aspectual Task	Non-aspectual Task
Standard FB Task	0.51** (0.45*)	0.34** (0.21)
Aspectual Task		0.86** (0.85**)

\*  $p < .01$ .  
\*\*  $p < .001$ .

tion 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,  $r = 0.31$ ,  $p > .05$ ) but the second test questions were (Number Questions,  $r = 1.00$ ,  $p < .001$ ). We also aggregated mean scores indicating that children solved both test questions within a trial of a task (called aggregate scores).

**Appendix D. Detailed event sequences of the non-aspectual and aspectual spontaneous helping tasks in Study 3**

	Aspectual Task	Non-aspectual Task
1	<b>Warm-up</b> Experimenters E1 and E2 pick up the child and its parents in the entrance hall of the department and accompany them to the testing room where the first warm-up game is already prepared. The child and the two experimenters play together with a toy train: the experimenters sit at different sides of the train route and push the train forward until they cannot reach it anymore and the other experimenter or the child continues. While playing the game E1 pretends to have difficulties reaching the train and asks the child for help by saying “Oh, uh, I can’t reach the train! Can you help me?”. After a while one of the experimenters suggests to play with toy animals and build up a zoo. Again E1 asks the child for help several times. When the child interacts with both experimenters freely, E1, one parent and the child move to a table where the testing will take place. E1 and the child on the lap of the parent sit vis à vis at the table, E2 hides behind a curtain. E1 and the child play together with a puzzle and look at a picture book.	
2	<b>Introduction of the objects</b> E1 says: “And now, let’s see what we will play next!” and E2 enters the scene from behind the curtain with a bag containing several e.g., soft toy pigs and says: “Look, you two, what I got here!” and puts the bag on e.g., the right side of the table with the opening open and the objects visible.	



**Fig. C1.** Mean number of trials answered correctly in the different tasks in Study 2 (sub-sample of children mastering all control questions). (\*  $p < .05$ , comparison against chance (=1); \*\*  $p < .05$ , comparison against chance (=0.5).

**Table C1**

Correlations (and partial correlations correcting for age and language ability in brackets) between the different tasks in Study 2 (sub-sample of children mastering all control questions).

Correlations Aspectual and Non-aspectual Tasks		
Identity/Location Questions	Number Questions	Aggregate Scores
0.31 (0.25)	1.00* (1.00)*	0.66* (0.64)*

\*  $p < .01$ .

- 3 Identity Change** E2 takes one toy pig out of the bag, places it in the middle and leaves the scene. E1 takes the toy pig and says: "Oh look, what a nice toy! It is so cuddly!" and offers the child to touch it. When the child touches the object, E1 says: "Oh no! I forgot something outside! I have to leave for a short time! But before I leave I will put the toy pig into this box!" and takes a box from under the table and places the box on the opposite side to the bag and puts the toy pig into the box. And says: "Okay, I will be back, soon."
- True Belief:* Before E1 leaves, E2 comes to the table and says: "Look, what I am doing now!" and takes the object out of the box and places it in the middle of the table. E1 watches the action attentively and says "Oh, okay, aha. I have to leave now." And goes out. After her return E1 says: "Okay, I'll put this into the box again." and puts the object into the box.
- False Belief:* After E1 leaves, E2 comes to the table and says: "Look, what I am doing now! 'E1' [points to the door] is outside and can't see what we are doing!" and takes the object out of the box and places it in the middle, says "Shhh, she does not know!" and leaves behind the curtain. After her return E1 says: "Oh, another pig! Now I've got two. I'll put this one into the box, too" and puts the object into the box.
- 4 The signal** E1 announces that she would need to leave and that she wants to take her objects out of the box and with her: "But this time I will take it/them [in German: "die" (ambiguous between singular and plural)] out of the box!" and takes out the first one and places it in the middle of the table. Then she goes on searching, opens the different doors of the box and says: "Huuuh?! I don't understand! This is strange! Huuhuh?!" and sits down on her chair, looks at the child and says: "Can you help me?" and waits. If the child does not react, she repeats "Can you help me?" If the child does not show any reaction, E1 puts away the box and the bag and says: "Let's see what comes next!" and the second trial begins.
- Location Change** E2 takes two toy pig out of the bag, places them in the middle and leaves the scene. E1 takes the toy pigs and says: "Oh look, what nice toys! They are so cuddly!" and offers the child to touch one of them. When the child touches the object, E1 says: "Oh no! I forgot something outside! I have to leave for a short time! But before I leave I will put the toy pigs into this box!" and takes a box from under the table and places the box on the opposite side to the bag and puts both toy pigs into the box. And says: "Okay, I will be back, soon."
- True Belief:* Before E1 leaves, E2 comes to the table and says: "Look, what I am doing now!" and takes one of the objects out of the box and places it in the bag containing all the other objects. E1 watches the action attentively and says "Oh, okay, aha. I have to leave now." And goes out. After her return E1 says: "Hm, I am putting this into the box again" and puts the object into the box.
- False Belief:* After E1 leaves, E2 comes to the table and says: "Look, what I am doing now! 'E1' [points to the door] is outside and can't see what we are doing!" and takes one of the objects out of the box and places it in the bag containing all the other objects says "Shhh, she does not know!" and leaves behind the curtain. After her return E1 says: "Hm, I am putting this into the box again" and puts the object into the box.

## Appendix E. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cognition.2018.01.001>.

## References

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>.

- Baillargeon, R., Scott, R.M., Bian, L., 2016. Psychological reasoning in infancy. *Annual Review of Psychology* 67, 159–186.
- 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, F., Buttelmann, D., 2015. Can one false-belief paradigm simply be applied to different age groups? Children's performance in an interactive theory-of-mind task at three different ages. In: Poster presented at the IMPRS workshop: Perspectives on the ontogeny of mutual understanding, Nijmegen, Netherlands.
- 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>.
- Buttelmann, D., Over, H., Carpenter, M., Tomasello, M., 2014. Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology* 119, 120–126. <https://doi.org/10.1016/j.jecp.2013.10.002>.
- Buttelmann, F., Suhrke, J., Buttelmann, D., 2015. What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology* 131, 94–103. <https://doi.org/10.1016/j.jecp.2014.11.009>.
- Butterfill, S.A., Apperly, I.A., 2013. How to construct a minimal theory of mind. *Mind & Language* 28 (5), 606–637. <https://doi.org/10.1111/mila.12036>.
- Carey, S., 2009. *The origin of concepts*. Oxford University Press, New York.
- Carruthers, P., 2013. Mindreading in infancy. *Mind & Language* 28 (2), 141–172. <https://doi.org/10.1111/mila.12014>.
- Carruthers, P., 2015. Mindreading in adults: Evaluating two- systems views. *Synthese* 192.
- 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>.
- Clements, W., Perner, J., 1994. Implicit understanding of belief. *Cognitive Development* 9, 377–397.
- Conway, J.R., Lee, D., Ojaghi, M., Catmur, C., Bird, G., 2017. Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance* 43 (3), 454–465. <https://doi.org/10.1037/xhp0000319>.
- Csibra, G., 2012. Revising the belief revision paradigm. Retrieved from <<http://www.cognitionandculture.net/home/blog/44-pierre-jacobs-blog/2455-do-we-use-different-tools-to-mindread-a-defendant-and-a-goalkeeper>>.
- Feigenson, L., Dehaene, S., Spelke, E., 2004. Core systems of number. *Trends in Cognitive Sciences* 8 (7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>.
- Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., Rakoczy, H., 2017. Are there signature limits in early theory of mind?. *Journal of Experimental Child Psychology* 162, 209–224. <https://doi.org/10.1016/j.jecp.2017.05.005>.
- Flavell, J., 1988. The development of children's knowledge about the mind: From cognitive connections to mental representations. In: Astington, J., Harris, P., Olson, D. (Eds.), *Developing theories of mind*. Cambridge University Press, Cambridge, England.
- Gergely, G., Csibra, G., 2003. Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences* 7 (7), 287–292.
- Gopnik, A., Astington, J.W., 1988. Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development* 59 (1), 26–37.
- Heyes, C., 2014. False belief in infancy: A fresh look. *Developmental Science* 17 (5), 647–659. <https://doi.org/10.1111/desc.12148>.
- Heyes, C., 2014. Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science* 9 (2), 131–143. <https://doi.org/10.1177/1745691613518076>.
- Jacob, P., 2012. Another look at the two-systems model of mindreading. Retrieved from <<http://cognitionandculture.net/blog/pierre-jacobs-blog/another-look-at-the-two-systems-model-of-mindreading>>.
- Kaufman, A., Kaufman, N., 1999. *Kaufman assessment battery for children*, 4th ed. Frankfurt am Main, Germany.
- Knudsen, B., Liszkowski, U., 2012. 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy* 17 (6), 672–691. <https://doi.org/10.1111/j.1532-7078.2011.00105.x>.
- 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>.
- 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>.
- Low, J., Apperly, I.A., Butterfill, S.A., Rakoczy, H., 2016. Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives* 10 (3), 184–189. <https://doi.org/10.1111/cdep.12183>.
- 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., Edwards, K., 2017. The curious case of adults' interpretations of violation-of-expectation false belief scenarios. *Cognitive Development* <https://doi.org/10.1016/j.cogdev.2017.07.004>.
- 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>.
- Onishi, K.H., Baillargeon, R., 2005. Do 15-month-old infants understand false beliefs? *Science* 308 (5719), 255–258. <https://doi.org/10.1126/science.1107621>.
- Perner, J., 1991. *Understanding the representational mind*. MIT Press, Cambridge, MA.
- Perner, J., Howes, D., 1992. 'He thinks he knows': And more developmental evidence against the simulation (role taking) theory. *Mind & Language* 7 (1–2), 72–86. <https://doi.org/10.1111/j.1468-0017.1992.tb00197.x>.
- 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>.
- Phillips, J., Ong, D.C., Surtees, A.D.R., Xin, Y., Williams, S., Saxe, R., Frank, M.C., and Endress (2010). *Psychological Science* 26 (9), 1353–1367. <https://doi.org/10.1177/0956797614558717>.
- Rakoczy, H., Bergfeld, D., Schwarz, I., Fiske, 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>.
- Santiesteban, I., Catmur, C., Hopkins, S.C., Bird, G., Heyes, C., 2014. Avatars and arrows: Implicit mentalizing or domain-general processing?. *Journal of Experimental Psychology: Human Perception and Performance* 40 (3), 929–937. <https://doi.org/10.1037/a0035175>.
- 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., Richman, J.C., Baillargeon, R., 2015. Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology* 82, 32–56. <https://doi.org/10.1016/j.cogpsych.2015.08.003>.
- 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>.
- 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>.
- 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>.
- Wang, B., Hadi, N.S.A., Low, J., 2015. Limits on efficient human mindreading: Convergence across Chinese adults and Semai children. *British Journal of Psychology* (London, England: 1953) 106 (4), 724–740. <https://doi.org/10.1111/bjop.12121>.
- 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., 2011. Developing a theory of mind. In: Goswami, U. (Ed.), *Wiley-Blackwell handbooks of developmental psychology*, 2nd ed. Wiley-Blackwell, Malden, Mass., pp. 258–284.
- 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](https://doi.org/10.1016/0010-0277(83)90004-5).
- Yott, Poulin-Dubois, 2016. Are infants' Theory of Mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development* 17 (5), 683–698. <https://doi.org/10.1080/15248372.2015.1086771>.