Cognitive Development 74 (2025) 101579

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Children's understanding of the subjectivity of intentions – Masked by linguistic task demands?

Isa Blomberg[a,b,*,1], Joana Lonquich[a], Marina Proft[a,2], Hannes Rakoczy[a,3]

[a] Department of Developmental Psychology, Institute of Psychology, University of Göttingen, Germany
[b] Leibniz-Science Campus Primate Cognition, Göttingen, Germany

ARTICLE INFO

Keywords:
Intentions
Theory of Mind
Preregistered
Performance factors

ABSTRACT

One crucial form of advanced Theory of Mind that is foundational for our everyday social lives is the understanding of the subjective intentionality behind people's actions. Intentions can be subjective in the following sense: an agent may do many things at once but which of these she does intentionally depends on the description under which she represent these acts (Anscombe, 1979; Searle, 1983). For example, a chef may simultaneously serve an exquisite dinner decorated with nuts, thereby impress the guests, while also triggering an allergic reaction in one guest – and only perform the former two intentionally (serve dinner, impress guests) but not the latter (cause allergic reaction). From a developmental perspective, a crucial question is when such complex action understanding emerges. Previous research indicated that children's understanding of subjective intentions emerges relatively late (Kamawar & Olson, 2011; Proft et al., 2019; Schünemann, Proft, et al., 2021). The present studies investigated whether these difficulties might have been due to linguistic task demands. We developed a new task that did not require subtle linguistic understanding of complex test questions. Instead, children judged whether some behavior was brought about by an intentional action given an agent's (false) beliefs and desires. Four-to-seven-year-olds ($N = 246$) participated in two preregistered studies. Despite the fact that the tasks were linguistically much simplified, we found comparable results to previous studies: Children proficiently ascribed subjective intentions only from around age five to six years. These results provide converging evidence for a protracted development of advanced Theory of Mind and raise questions, for future research, regarding the foundations of such developmental trajectories.

1. Introduction

Action understanding and interpretation are central for how we interact with each other. To understand why agents act the way they do, we need to take into account their subjective perspective (their beliefs, desires, and resulting intentions) from which the action makes rational sense (Astington, 2001; Bratman, 1987; Searle, 1983). From an ontogenetic point of view, the crucial question is how

* Correspondence to: University of Göttingen, Institute of Psychology, Department of Developmental Psychology, Waldweg 26, 37073 Göttingen, Germany.
E-mail address: isa.blomberg@uni-goettingen.de (I. Blomberg).
[1] https://orcid.org/0000-0002-9524-335X.
[2] https://orcid.org/0000-0002-8799-9165.
[3] https://orcid.org/0000-0003-3296-0551.

https://doi.org/10.1016/j.cogdev.2025.101579
Received 26 August 2024; Received in revised form 16 April 2025; Accepted 20 April 2025

this form of action understanding and interpretation develops.

1.1. Development of understanding others' intentional actions

From early on children detect goal-directed qualities in some intentional actions (e.g., Meltzoff, 1995; Woodward, 1998). Infants can distinguish which behavior constitutes an action and which does not. This is evident, for example, in children's imitation behavior (Carpenter et al., 1998; Meltzoff, 1995), and their reactions towards unwilling vs. unable agents (Behne et al., 2005). Children also have a basic grasp of rational agency. They understand ends in relation to means in both imitation studies (Gergely et al., 2002), and looking time studies (Gergely & Csibra, 2003). These capacities of the perception and interpretation of actions are robustly present in prelinguistic infants and other non-linguistic creatures (e.g., Call et al., 2004; Fischer & Price, 2017; Phillips et al., 2009; Schünemann, Keller, et al., 2021; Völter et al., 2023; Wood et al., 2007).

However, clearly human action interpretation goes well beyond these forms of perceiving and interpreting actions. It does not involve appreciating *whether* a given behavior was an intentional act or a mere happening, but much more subtle appreciation *in which ways* behavior may be intentional, *which* intentional act a given behavior constitutes, depending on the subjective intentionality behind it. So, how do children come to develop these capacities for complex, subjective action understanding?

1.2. Understanding the subjectivity of intentions

Mature understanding of intentional actions appreciates the subjective aspectuality of intentions and actions in the following sense: Actions are always performed intentionally under some aspect but not under others (Anscombe, 1979; Bratman, 1987; Searle, 1983). Oedipus can intend to marry Yocasta without intending to marry his mother (because he is not aware of the fact that Yocasta is his mother). Or, to take less dramatic examples: someone eating your last chocolate can be described and represented in various aspectual ways: "eating chocolate" (A), "eating your last chocolate" (B). Now, imagine this person falsely believes that there is more chocolate. She intentionally eats the chocolate (A). She does not, however, act intentionally under the description "eating your last chocolate" (B). Agents can only intentionally act under the aspects or descriptions they represent, and not under aspects they do not represent at all (i.e., that this chocolate is your last).

Intentions can be aspectual because they are based on mis- or partial representations of reality, as illustrated in the example above. In these cases, the cognitive basis (i.e., beliefs and knowledge) of the resulting intention makes the difference: If an agent lacks knowledge of or misrepresents an aspect of an action, they cannot intend them under these aspects or descriptions. But situations can be even more nuanced and complex: The agent may have full knowledge of the action and its effects under various aspects, yet still only intentionally performs it under one aspect (bringing about the main effect) and not under another (foreseeing, but not intending to bring about a side effect). In this paper, we will focus on the first type of aspectual intentions in which an agent only intentionally performs an action under one description but not under another because she is misinformed.

Such forms of subjective action understanding thus build on (but also go beyond) general meta-representational understanding: representing how other agents subjectively represent the world and act accordingly. Explicit forms of general metarepresentational abilities emerge over the preschool years (Perner, 1991; Rakoczy, 2022). Litmus tasks for meta-representing others subjective representations include the false belief task, appearance-reality and related tasks in which children have to represent that another agent has a different (outdated) perspective of reality that differs from their own and from reality and predict the agent's action accordingly (Flavell, 1986; Wimmer & Perner, 1983). Children generally come to solve such litmus tasks and can thus be said to operate with a basic meta-representational framework from around age 4 (Wellman et al., 2001; Wimmer & Perner, 1983). Now, subjective action understanding thus presupposes and builds on such general capacities for meta-representation. So, when after the emergence of basic meta-representation do children make use of this general capacity for complex subjective action understanding?

Initial studies on this question asked children very directly about the intentionality of a given action under various descriptions: if someone performs an action that can be described as A or B, but only performs it intentionally under aspect A, does she thereby also intentionally perform it under aspect B? In Kamawar and Olson (2011)'s study, for example, children were presented with a vignette about a police officer who lost his keys on the street. Anne gives the keys back. The police officer really is also Cathy's dad. But this is unknown to Anne. So, the crucial test question was then whether Anne meant to give the keys to Cathy's dad. These studies found late competence: The majority of children under the age of 8 wrongly answered "yes."

Later studies that reduced relevant performance factors found somewhat earlier competence from 5 to 6 years on (Proft et al., 2019; Schünemann, Proft, et al., 2021). These studies also investigated potential explanations for this protracted development. Schünemann, Proft, et al. (2021) found that false belief understanding was necessary but not sufficient for subjective intention understanding: only those children who passed standard first-order false belief tasks ascribed subjective intentions. But many children at age 4–5 did pass false belief tasks while failing the subjective intention understanding task. Why this lag? Schünemann, Proft, et al. (2021) explored various other cognitive capacities that, together with basic false belief understanding, may be crucial for acquiring a grasp of subjective intentions; but no evidence was found that executive functions, verbal intelligence or second-order false belief understanding played a crucial role. Understanding subjective intentions could be part of more advanced forms of Theory of Mind abilities in general, which develop in tandem with other skills, such as recursive thinking and understanding higher order Theory of Mind (Liddle & Nettle, 2006; O'Grady et al., 2015; Schidelko et al., 2022), as well as understanding complex pragmatic nonliteral speech acts or sarcasm and subtle disclosure (Filippova & Astington, 2010; Happé, 1994). However, it remains thus unclear why there is this considerable gap in development between children's passing basic Theory of Mind (false belief) tasks around age 4, and their proficiency in task that tap understanding subjective intentions not before age 5–6.

1.3. Do linguistic performance factors mask children's competence?

One possibility is that these tasks present false negatives and underestimate children's early competence due to linguistic performance factors. In particular, may children misunderstand the test question (Schünemann, Proft, et al., 2021; Searle, 1983). In particular, they may not yet be able to differentiate so-called "de re" and "de dicto" readings of utterances with mental state expressions (Nelson, 2023). On an extensional - de re - reading of the test question in Kamawar & Olson (2011)'s study "Did she mean to give the keys to Cathy's dad?", the expression "Cathy's dad" merely works referentially. What counts is what the expression refers to; not how anyone represent that referent. The question thus reads "Did she mean to give the key to this person (who is also Cathy's dad)?", and the correct answer would be "yes." However, on an intensional - de dicto - reading of the test question "Did she mean to give the keys to Cathy's dad?", the expression "Cathy's dad" works differently. Now, it is not only about the reference, but the way the agent represents that referent. The question thus reads "Did she mean to give the keys to the person she identified as Cathy's dad?", and the correct answer would be "no." We adult folk mostly read such questions in de dicto ways. But even for adults, there are many grey areas where it is not always clear which reading is the correct one. Suppose Grandmother passed by Taylor Swift in the street without recognizing her (she has not even ever heard of Taylor Swift). Did she see Taylor Swift? Well, in a sense (de re) yes, but in another sense (de dicto) no, since she did not see her *as* Taylor Swift.

Given that these subtle linguistic differences can be challenging even for adults, it may well be that children "failed" previous tasks of subjective intention understanding not because they lack the conceptual competence; this competence may have merely been masked by linguistic performance factors to do with the complex type of question. Evidence that is compatible with this possibility comes from studies on children's understanding of the aspectuality of beliefs: When do children come to understand that someone can believe that Clark Kent is in the house without believing that Superman is in the house even though, of course, Clark Kent = Superman? The first studies on this topic used complex test questions that requires sophisticated linguistic capacities. For example, in some studies there were two objects that were both erasers: one was visibly, whereas the other one was primarily and visibly a dice that was also an eraser (this could not be seen but just found out by using it). The naïve protagonist saw both objects and the test question was which object he would take if he needed an eraser (Apperly & Robinson, 1998, 2003; Perner et al., 2015; Russell, 1987; Sprung et al., 2007). These tasks were very difficult, with children performing competently only from age 6 to 8, much later than on standard false belief tasks. Later studies radically reduced linguistic performance factors and found much earlier competence; in fact competence in these simplified tasks and standard false belief tasks now co-emerged and the two were highly correlated Rakoczy et al. (2015).

Similarly, in the case of aspectual intentions, children might be able to understand the de re/de dicto distinction conceptually, in principle, yet for linguistic or other reasons, may not spontaneously find the de dicto reading obvious. Once relevant linguistic performance factors are reduced, perhaps competence in children's grasp of the subjectivity and aspectuality of intentions can be uncovered much earlier?

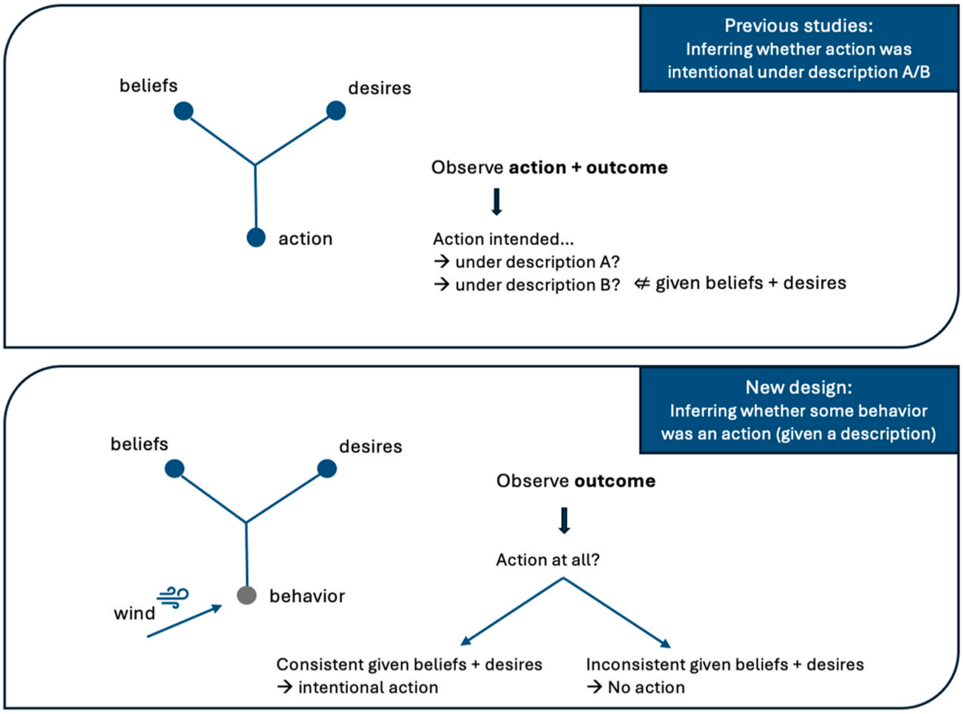


Fig. 1. Logic of the new task compared to logic of previous studies

1.4. Rationale of the present studies

The rationale of the present study was thus to reduce linguistic performance factors in the following way: In previous studies, children heard vignettes about an agent, her beliefs (e.g. "this is the police officer", "those are his keys") and desires (e.g. to give the keys back), and an action she intentionally performed (giving the keys to the person in question); and they then had to infer under which description ("give keys to police officer"/"give keys to Cathy's dad") the action in question was or was not intentional (see Fig. 1 top). That children performed poorly until age 6 or so may have been due to performance factors: in particular, they may have been confused about which reading of the test questions to apply (de re vs. de dicto).

In order to reduce these linguistic performance factors, we developed a new task that does not involve such linguistic subtleties. Rather than asking complex and ambiguous test questions – whether an action was intended under a description A or B – children had to judge whether some behavior was an intentional action at all, given the agent's background beliefs and desires (see Fig. 1 bottom). No handling of subtle linguistic matters (de re/de dicto distinction etc.) was required.

We devised stories of a little pirate who was collecting treasures. The outcome - on which island the little pirate ended up - was shown. It was ambiguous, however, what happened in between. Was the outcome the result of an intentional action (the little pirate actively and intentionally went there by himself) or by a plausible alternative factor (the wind drove the little pirate there)?

The story involved objects with different values to the little pirate that were placed at the island(s). In a control condition, the little pirate was fully informed about the placement of the objects and then ended up at an island in ways consistent or inconsistent with his (fully accurate) background beliefs and preferences. The correct answers (assuming rationality) would thus be that the little pirate went intentionally in the consistent and was driven by the wind in the inconsistent cases. In the test conditions, the little pirate missed some crucial information and thus mis-represented the situation in various ways (falsely believing that the most valuable object was at location B when it was really at location A). He then ended up either at island A (objectively better, but inconsistent with his subjective rational perspective) or at island B (objectively worse, but consistent with his subjective rational perspective).

When children then have to decide whether the little pirate ended up at a given location, they thus have to reason about subjective aspectual intentions, but without the linguistic complications: the question is just "did he go there himself, or was he driven by the wind?". More specifically, they have to reason along the following lines: "He ended up at location A. This is good for him. But he does not know it is good for him. He does not represent island A as the island with the best outcome. Rather, for him, subjectively, island B is the one with the best outcome. So, in his action planning he must have formed the intention (to get to the island with the best outcome) 'I go to island B'. But he ended up at island A. So it must have been the wind."

The general rationale and expectation was thus the following: If children's competence in understanding subjective intentions was masked by linguistic task demands in previous studies, children should show earlier competence in our new task with reduced linguistic demands: They should master test conditions earlier than previously reported and should perform well in control conditions from early on. If, however, previous studies really tapped a conceptual competence deficit in understanding subjective intentions up to age 6, children in the present study should perform similarly as in previous studies: they should master test conditions reliably from around 6 years, but perform well in control conditions from early on.

2. Study 1

2.1. Method

The experimental design, sample size and statistical analyses were pre-registered on OSF (https://osf.io/79n8t). Study materials, data sets, analysis scripts are accessible on OSF (https://osf.io/tbgp4/).

2.1.1. Participants

One hundred twenty-seven 4- to 7-year-old children were included in the final sample (age: $M = 5;11$, $SD = 1;1$, range = 4;0–7;11, 65 female, 62 male; see Table 1). The sample size was determined a priori via data simulation. The calculation was based on the main question whether children's performance on the task varies as a function of age and condition with the goal to obtain .80 power for that effect. Additional thirty-four children took part, but were excluded from the data analysis for the following reasons: preregistered criteria based on error in familiarization trials ($n = 19$) and based on hierarchy checks ($n = 15$). Single trial exclusion (9 % of the data) as preregistered is described in the Appendix. Children were recruited via the department's database of children whose parents previously agreed to the participation in our studies. Participants live in or near a medium-sized city in Germany and had a mixed social-economic status. This data was not systematically collected as it was not part of the research questions.

Table 1
Demographic distribution of age (years;months) and gender per age group in Study 1

Age in years	*M* age	*SD* age	Gender
4 ($n = 32$)	4;6	0;3	59 % female
5 ($n = 33$)	5;4	0;3	48 % female
6 ($n = 32$)	6;5	0;3	38 % female
7 ($n = 30$)	7;6	0;3	60 % female

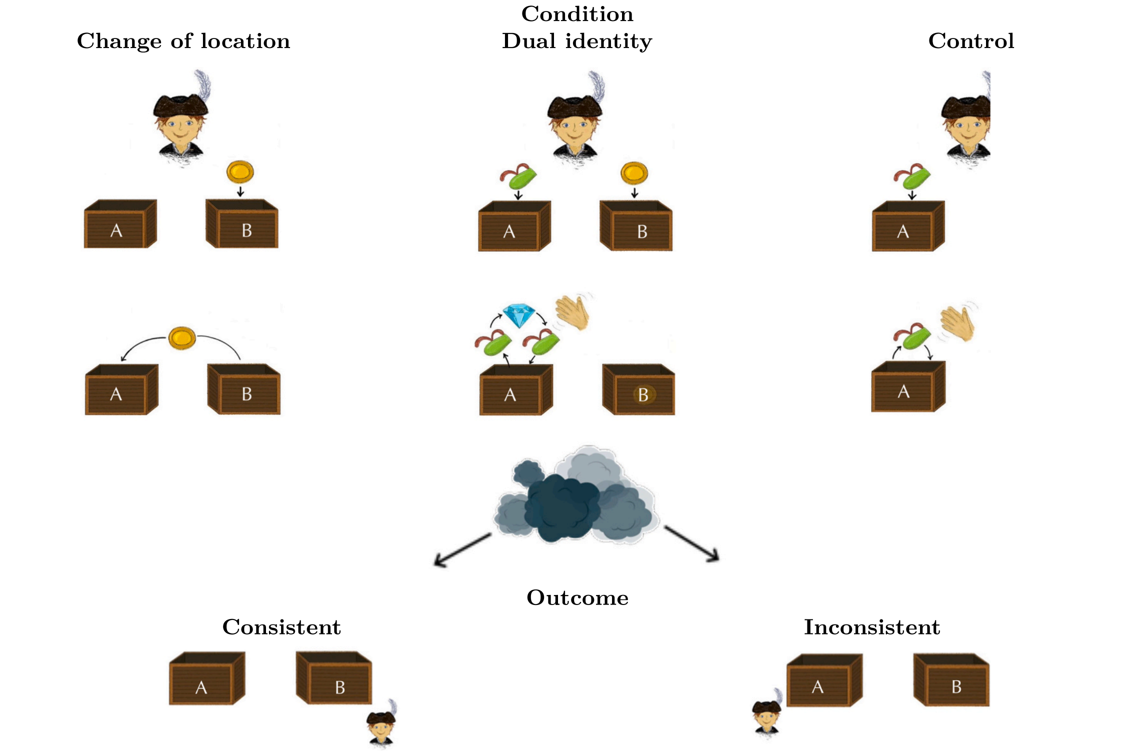
2.1.2. Procedure

The experimental logic of Study 1 is displayed in Fig. 2. Children were first familiarized to the general logic of the study (see Appendix for details). After the familiarization, children saw three types of test trials: change of location, dual identity and control trials. The locations of the objects were counterbalanced. After each outcome, the test question was asked: “Did the little pirate travel to the island himself or did the wind drive him there?”. The first option mentioned in the test question was counterbalanced across children. In general, each story included a second character – a female pirate – who, depending on the condition manipulated the little pirate’s beliefs or not.

Change of location trials. In change of location test trials, the female pirate first hid a coin, for example, at island B. In the little pirate’s absence, the female pirate then changed the location of a coin to island A. After the female pirate left the scene, the little pirate came back and a storm came up. The outcome was shown: the little pirate either ended up at island A or island B.

In order to understand how the little pirate ended up on the respective island, the children should ideally make the following mental state attributions: Children should reason that the little pirate now correctly believed that the coin was the most valuable object in this situation, but he falsely believed that the coin is at island B. They should assume that he wanted to have the coin (desire) and thus he formed the subjective intention to go to island B, as this is the best outcome from his subjective perspective. Children should assume that even though objectively this is not the best outcome, because the coin is in reality now on island A.

If the little pirate now ends up at island B (consistent outcome), children should judge that the outcome as intentionally brought about. Conversely, they should judge that if the little pirate ends up at island A, which is objectively better, this outcome is inconsistent with his beliefs and desires. Children should reason that he did not want to go to island A, because he did not believe that this is the best



Test questions: Did the little pirate travel to the island himself, or did the wind drive him there?

Consistent outcome	*Inconsistent outcome*
[Correct answer: “went there himself”]	[Correct answer: “wind drove him there”]

Desire question: To which island did the little pirate want to go? To this or that?

[Correct answer: “right”]	[Correct answer: “left”]

Knowledge control question:

Change of location	*Dual identity*	*Control*
Did the little pirate see that the female pirate took the coin to the other island? [Correct answer: “no”]	The [fantasy object] is also a diamond. Does the little pirate know that? [Correct answer: “no”]	-

Fig. 2. Experimental logic of study 1.

outcome for him. In this case, children should then judge that the outcome was not intentionally brought about.

Dual identity trials. In dual identity trials, the female pirate first hid some (not valuable) fantasy object, for example, on island A and a coin on island B. In the little pirate's absence, she then revealed a second, more valuable identity: the fantasy object is also a diamond. Children learnt in the familiarization that the little pirate preferred the coin over the not valuable fantasy object, but preferred the diamond over the coin. After the female pirate left the scene again, the little pirate came back, a storm came up, and children saw the outcome: the little pirate ended in either island A or island B.

Children should reason that the little pirate now correctly believes that the coin is at island B. They should reason that he does not know that the fantasy object (= diamond) is in fact more valuable to the little pirate than the coin. Children should therefore conclude that he incorrectly believes that the coin is the most valuable object, because of his false belief about the fantasy object's identity. Children should reason that he has the desire to get the coin and thinks this is the best outcome for him (island B), however, objectively, island A is the best outcome for him (diamond).

If the little pirate now ended up at island B, children should judge that this is consistent with his subjective beliefs and desires, and thus, this outcome was brought about by his subjective intention. Conversely, when he ends up at island A (inconsistent outcome), children should judge the outcome as not brought about intentionally, even though it is objectively better for the little pirate.

Control trials. In control trials, the female pirate hid again some (not valuable) fantasy object at, for example, island A and a coin at island B. In the absence of the little pirate, the female pirate then showed that a fantasy object did not have a second identity. After the female pirate left the scene, the little pirate came back, a big storm came up and children saw the outcome.

Thus in this case, children should reason that the little pirate correctly believes that the coin is at location B and that the coin is the most valuable treasure to the little pirate, because he has a true belief about the fantasy object's identity. Children should reason that if he then ends up at location B, this is objectively and subjectively the best outcome for the little pirate. Children should judge that he traveled there himself. If he ends up at island A, children should reason that this is both objectively and subjectively not good for him, thus this outcome should be judged as not intentionally brought about.

2.1.3. Design

In a 3 (Condition: "change of location", "dual identity", "control") × 2 (outcome type: "consistent", "inconsistent") × Age mixed effects design, each child received six test trials in counterbalanced order. Condition and outcome type were repeated measures within-subject factors and age was a between-subjects factor. The study was tested in-lab with animated videos presented via PowerPoint. Videos were created using Vyond (GoAnimate, 2024). The experimenter asked follow-up questions to make sure they understood the preference hierarchy of the objects involved.

2.1.4. Coding

Test session were video-recorded and coded from tape. All trials were coded by a second, independent coder. In case of disagreement the first author checked the video tape again and made a decision. If a case was still unclear it was coded as a missing value. Agreement between coders was 99.6 %. Children answered either verbally (e.g., "drove himself", "wind") or by pointing on the printed versions of the choice options. A paddle represented that the little pirate traveled to the island himself. Clouds represented that the wind drove him there.

2.1.5. Statistical analyses

Analyses followed the preregistered analysis protocol unless noted otherwise. All analyses were conducted using R 4.3.0 (R Core Team, 2023). For each model, we used Likelihood Ratio Tests to compare the fit of the full model to that of a null model which was identical but lacked the predictors of interest. This way, we tested the overall effect of our fixed effects and avoided "cryptical multiple testing" (Forstmeier & Schielzeth, 2011), that is, the risk of inflated type I error rates due to testing multiple predictors without proper correction. If not stated otherwise, the full-null model comparison was significant.

To test whether children's probability to answer the intention questions correctly varied as a function of condition, age, and their interaction, we fitted a binomial mixed effects model. Condition was included as a categorical variable with three levels using dummy coding. To ensure generalizability of the results, we ran two models with different reference groups. In the first model, the dual identity condition served as the reference group. This allowed us to test two contrasts: dual identity vs. control condition (the contrast of primary interest), and dual-identity vs. change of location. In the second model, we recoded the reference group to change of location to directly test the contrast between control and change of location. Note that we only expected a difference in the first contrast (i.e., dual identity vs. control or change of location vs. control) and no difference in the second contrast (i.e., dual identity vs. change of location). All models included trial and outcome type as fixed effects, a random intercept for experimenter and participant, and a random slope for the effect of condition and outcome type within participant. By adding trial and the random effects, we deviate from our preregistration. However, we follow the recommended maximum random effects approach as justified by the design (Barr, 2013). We expected an interaction effect between condition (first contrast) and age if younger children fail in subjective intention task due to a competence and not a performance deficit. That means that all children (regardless of age) should have a high probability, above chance level (0.50), of correct answers in the control condition. Whereas children's probability of correct answers in dual identity and change of location conditions increases with age. The difference between conditions involving misrepresentations (dual identity and change of location) and conditions that did not involve misrepresentations (control) should get smaller with age.

To test from which age onwards children perform significantly above chance we calculated 95 % bootstrapped confidence intervals based on the fitted models. The lowest age (x-axis intercept) was calculated for which the bootstrapped lower confidence interval limits no longer included the chance level of 0.50.

2.2. Results

Children's responses to the intentional action question are shown in Fig. 3 and Table 2. Consistent outcome types were consistent with the little pirate's desires and beliefs. That means he landed on the island where he wanted to go and where he believed the treasure to be. The correct answer was always that he drove there himself. Inconsistent outcome types were inconsistent with the little pirate's desires and belief. The correct answer was that the wind drove him there. Children from 4 years on gave these correct responses when no misrepresentation was involved (Fig. 3, control condition). In conditions that did involve misrepresentation (dual identity and change of location) the correct response pattern developed with age. In the dual identity condition, 4-year-old's seemed to answer randomly. In the change of location condition, however, 4-year-olds showed the reverse pattern to the correct one. Four-year-old's said, for example, that the agent intentionally drove to the objectively better island (the one with the treasure about which the agent did not know). Vice versa, when the agent landed at the objectively worse island (the one where he mistakenly believed the treasure to be located), children claimed the wind drove him there.

In order to analyze children's responses, we modeled the probability of correct answers as a function of age, condition and their interaction. As expected by a competence deficit account, we found a significant interaction between condition (dual identity vs. control) and age ($b = -1.80$, $z = -2.79$, bootstrapped 95 % CI: [-14.65, -0.72], $p = .005$; Fig. 4). In a recoded model, this interaction was also significant for the contrast change of location vs. control condition ($b = -1.64$, $z = -3.94$, bootstrapped 95 % CI: [-17.50, -1.10], $p < .001$). That means that younger but not older children were more likely to make errors in dual identity or change of location conditions than in control condition. The older children got, the smaller the difference between the conditions became.

Once predictors are involved in an interaction, the respective main effects are conditional on the reference group of the other predictor in the interaction. The reference group is 0 for continuous variables, which makes no sense for age, so it was z-transformed. Thus the conditional main effect of condition holds for mean age: Children were more likely to answer correctly in dual identity than change of location conditions ($b = -2.72$, $z = -3.30$, bootstrapped 95 % CI: [-18.28, -1.80], $p = .001$). Children were slightly better in inconsistent outcome types (76.7 % correct) than in consistent ones (68.8 % correct). However, this difference was not significant ($b =$

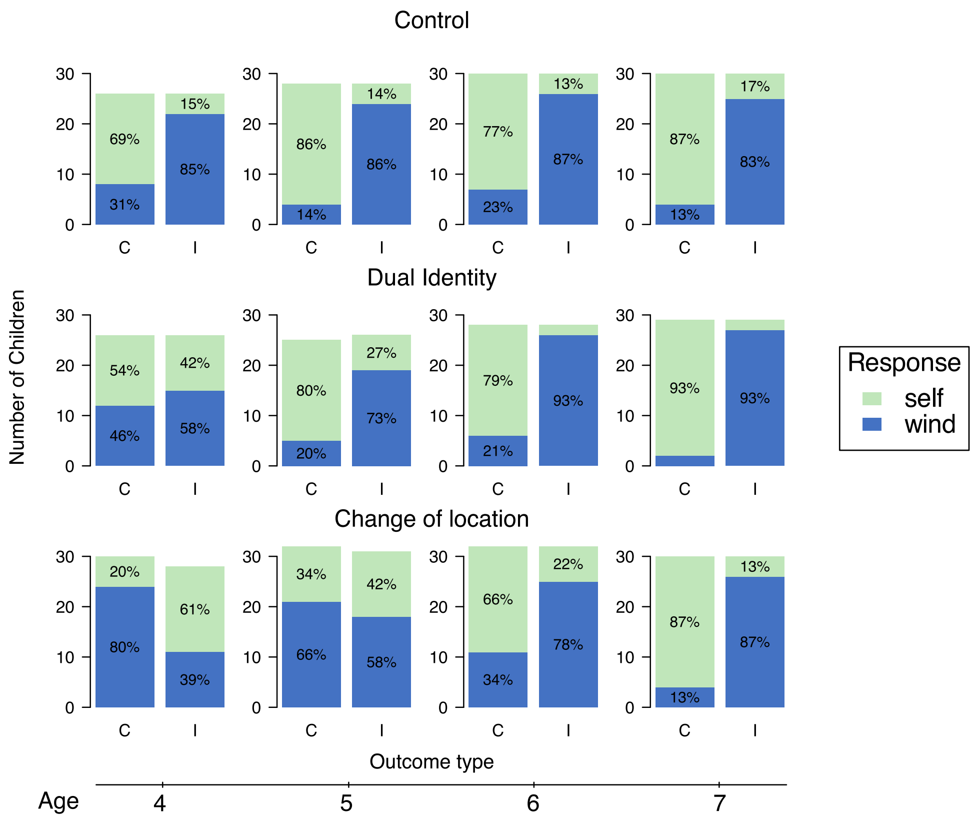


Fig. 3. Children's responses to the test question. *Note.* Children answered either "self" or "wind" to indicate how the little pirate ended up on a given island. The figure shows these responses as a function of age groups, condition (Control, Dual identity, Change of location), and Outcome types (C = consistent, I = inconsistent).

Table 2
Percentage of correct responses per age group and condition in Study 1

Age in years	Condition		
	Control	Dual identity	Change of location
4 ($n = 32$)	77 %	56 %	29 %
5 ($n = 33$)	86 %	76 %	46 %
6 ($n = 32$)	82 %	86 %	72 %
7 ($n = 30$)	85 %	93 %	87 %

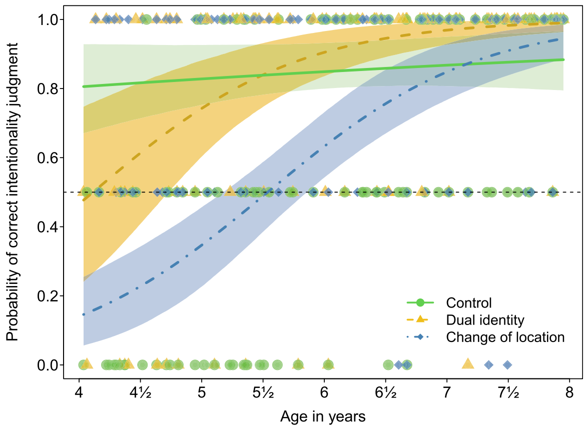


Fig. 4. Probability of correct responses across age by condition. *Note.* The horizontal dashed line represents the chance level of 0.50. Each filled dot represents the response proportion of one participant per condition. Each fitted line indicates the predicted probability of a correct response per condition by the binomial mixed effects model. Areas around the fitted lines are 95 % confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 10,000 boots. Each participant is represented by one dot per condition. The dashed line indicates chance level.

0.86, $z = 1.96$, bootstrapped 95 % CI: [0.13, 12.80], $p = .050$). We found a main effect of trial: Children were better in the 2nd trial per condition than in the 1st trial ($b = 0.68$, $z = 2.28$, bootstrapped 95 % CI: [0.11, 5.98], $p = .023$).

As can be seen in Fig. 4, children performed well in control conditions from 4 years on. From age of 4;8 years on they performed significantly above chance level in the dual identity condition. From the age of 5;10 years, children performed significantly above chance level in the change of location condition.

2.3. Discussion

The research question of this study was when children begin to reliably understand subjective aspectual intentions. The rationale was to test whether previous studies that found proficiency not before age 6 underestimated early competence due to linguistic performance factors. Therefore, a new task was devised without complex linguistic requirements. Rather, children had to infer, given the agent's background beliefs and desires, which subjective intentions he would form and whether a given state of affairs was thus the result of such an intentional act, or brought about externally (by the wind).

The results were the following: Children mastered the control condition (without misrepresentation) from 4 years on. However, conditions that involved misrepresentation were solved only from 5 to 6 years on. These findings thus converge with those of previous studies (Proft et al., 2019; Schünemann, Proft, et al., 2021). They thus do not provide any evidence for earlier competence and for the suspicion that the results of previous studies may have masked competence due to linguistic performance factors.

However, while we have reduced linguistic performance factors in the present task, we may have introduced other performance factors in terms of inferential complexity and working memory. Generally, children had to reason through a complex inferential chain in these tasks. And they had to keep track of much information regarding the agent's preference hierarchies and the locations and identities of the objects.

In addition, we found unexpected differences between conditions. Children performed better in the dual identity condition than in the change of location condition. We know from many studies that dual identity/function tasks are more difficult (Apperly & Robinson, 1998, 2003; Perner et al., 2015; Russell, 1987; Sprung et al., 2007) or that there is no difference between a standard and aspectual false belief understanding with reduced performance factors (Rakoczy et al., 2015). In the change of location version younger children tended to respond incorrectly, whereas in the dual identity condition they were more likely to guess. The pattern in the change of location condition is reminiscent of the error pattern that children make before solving the classic false belief task. However, this pattern may have been due to some superficial differences between the present conditions: Even though we tried to match the two conditions as closely as possible, we acknowledge that there may have been different working memory demands between conditions. In particular, both in the control and the dual identity condition the objects were not transferred to another location whereas they were in the change of location condition. To overcome these limitations, we devised with a simplified version of our new task in Study 2.

3. Study 2

The rationale of Study 2 was to test if children revealed earlier competence in a subjective intention ascription task with even more reduced task demands in terms of inferential complexity and working memory. In addition, the structure of the different conditions were more closely matched and aligned. The general structure of the new task was the same as in Study 1, with some crucial modifications that we describe in detail below.

3.1. Method

As in Study 1, the experimental design, sample size, and statistical analyses were preregistered on OSF (https://osf.io/umg36). Study materials, sample size calculations, data sets, analysis script and results are accessible on OSF (https://osf.io/tbgp4/).

3.1.1. Participants

One hundred nineteen 4- to 7-year-old children were included in the final sample (age: $M = 6;0$, $SD = 1;2$, range = 4;0–8;0, 61 female, 58 male; see Table 3). The sample size was determined a priori via data simulation. The calculation was based on the main question whether children's performance on the task varies as a function of age and condition with the goal to obtain .80 power for that effect based on results of Study 1. Additional sixteen children took part, but were excluded from the data analysis as preregistered, due to errors in familiarization trials ($n = 15$), and previous participation in Study 1 ($n = 1$). Children were recruited from the same database as in Study 1. Children who participated in Study 1 did not participate in Study 2.

3.1.2. Design and procedure

The tasks were conceptually analogous to those in Study 1, but the following changes were made to reduce the task demands compared to Study 1.

- First, the objects used were no longer subject to a preference hierarchy. Instead, the little pirate said which object he wanted in each case.
- Second, children were scaffolded through the necessary inferential steps (adapted from Schünemann, Proft, et al., 2021). Children were asked relevant knowledge and belief questions prior to the subjective intention test question. That gave us the chance to correct children if they ascribed wrong knowledge or belief states. Schünemann, Proft, et al. (2021) reported 2 years earlier competence when a primer of the agent's false belief was added before the intention ascription question.
- Third, the dual identity condition was matched more closely to previous studies (e.g., Proft et al., 2019; Schünemann, Proft, et al., 2021). In the little pirate's absence, the second identity was revealed: for example, a ball (identity A) that was also a rattle (identity B). In the little pirate's presence, the objects were transferred to another location under its second identity B (i.e., making rattling noises).
- Fourth, we implemented actual true belief control conditions. In true belief conditions, the little pirate either observed the transfer of the object to another island (change of location condition). Or he witnessed the uncovering of the second identity and the transfer of the object under its second identity to another island (dual identity condition).

In a 2 (Condition: "change of location", "dual identity") x 2 (belief: "true belief", "false belief") x 2 (outcome type: "consistent", "inconsistent") x Age mixed effects design, each child received eight test trials in counterbalanced order. Condition, belief, and outcome type were repeated measures within-subject factors and age was a between-subjects factor. The study was tested in-lab with animated videos (GoAnimate, 2024) presented via jsPsych (De Leeuw et al., 2023). The general logic of the pirate story was the same as in Study 1. Changes in control questions are described in the Appendix.

Table 3
Demographic distribution of age (years;months) and gender per age group in Study 2

Age in years	*M* age	*SD* age	Gender
4 ($n = 32$)	4;6	0;3	47 % female
5 ($n = 29$)	5;6	0;3	52 % female
6 ($n = 28$)	6;5	0;3	54 % female
7 ($n = 29$)	7;7	0;2	52 % female
8 ($n = 1$)	8;0		100 % female

Note. Due to an error in our database, a child took part in the study who was already 8 years old. However, this child was not excluded from the analysis. For a simplified visualization, the 8-year-old has been added to the group of 7-year-olds for the following illustrations.

3.1.3. Coding

Experimenters coded live during the experiment. Whether an answer was correct depended on the condition and outcome type. The assignment to correct/incorrect answers was implemented in the jsPsych experiment. If a child's response was unclear during the study, the video recording was reviewed by the experimenter after the test session and then a decision was made on coding. Reliability coding by a second blinded coder for 30 video recording (25 %) revealed 99.9 % agreement.

3.1.4. Statistical analyses

As in Study 1, we fitted a binomial mixed effects model predicting children's probability of correct answers by belief (true belief vs. false belief), z-standardized age, and their interaction. We added condition (dual identity vs. change of location), trial, outcome type (consistent vs. inconsistent) as fixed effects. We added a random slope for belief, condition and outcome type nested in participant, and a random intercept for participant and experimenter as random effects. By adding trial and the random slope effects for condition and outcome type, we deviate from our preregistration. However, we follow the recommended maximum random effects approach as justified by the design (Barr, 2013). As in Study 1, we expected an interaction effect of belief. Across all ages and conditions, we expected that children should perform well, and above chance of 0.50, in all true belief trials. As in Study 1, we tested from which age onwards children perform significantly above chance. We calculated 95 % bootstrapped confidence intervals based on the fitted models. The lowest age (x-axis intercept) was calculated for which the bootstrapped lower confidence interval limits no longer included the random level of 0.50.

3.2. Results and discussion

Children's responses to the intentional action question by belief manipulation (True vs. False belief), and condition (Change of location vs. Dual identity), and age groups are displayed in Fig. 5 and Table 4. As expected, children performed similarly in the change

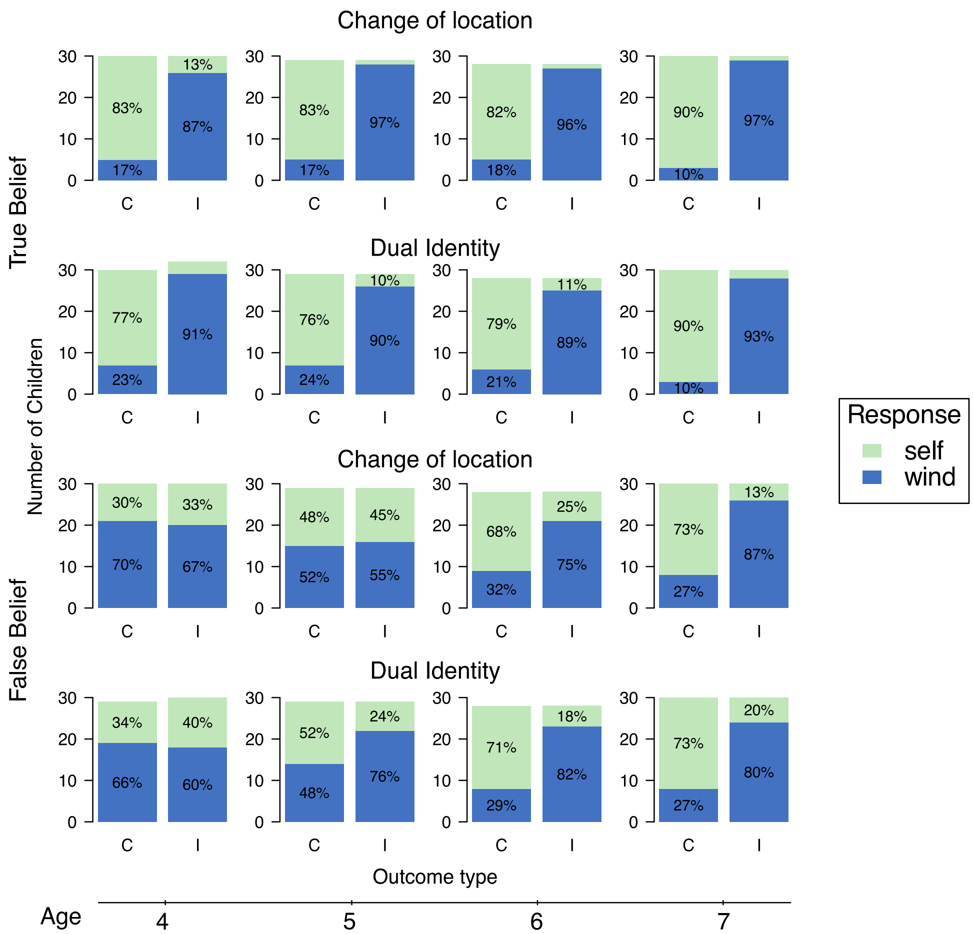


Fig. 5. Children's responses to the test question. *Note.* Children answered either "self" or "wind" to indicate how the little pirate ended up on a given island. The figure displays these responses as a function of age groups, belief manipulation (True vs. false belief), condition (Control, Dual identity, Change of location), and Outcome types (C = consistent, I = inconsistent).

of location and dual identity condition. As in Study 1, from 4 years on the vast majority of children answered correctly when no misrepresentation was involved (first two rows, True belief). They said correctly that the little pirate drove to the island himself in consistent trials and that the wind drove him there in inconsistent trials. As in Study 1, this correct response pattern developed only with age when misrepresentations (3rd and 4th row, False belief) were involved. It is already visible in Fig. 5 that children answered very similarly to the two conditions (Dual identity and Change of location). As in Study 1, children were better in inconsistent than consistent outcome type trials (see Fig. 5). However, this difference was particularly extreme in false belief trials.

In order to analyze children's responses we modeled the probability of correct responses. As in Study 1, we found a significant interaction between belief type (true vs. false belief) and age (b = -0.58, z = -2.07, bootstrapped 95 % CI: [-1.35, -0.03], p = .038; Fig. 6). Additionally, we found conditional main effects. Once predictors are involved in an interaction, the respective main effects are conditional on the reference group of the other predictor in the interaction. In our case this means that the conditional main effect of age holds for the reference group of condition - false belief. The likelihood of correct responses in false belief trials increased with age (b = 0.98, z = 4.54, bootstrapped 95 % CI: [0.58, 1.68], $p < .001$). The conditional main effect of belief types holds for mean age: children performed better in true belief than false belief trials (b = 1.84, z = 5.66, bootstrapped 95 % CI: [1.31, 3.04], $p < .001$). We found a main effect of outcome type: children were significantly better in inconsistent outcome types (82.6 % correct) than in consistent ones (69.2 % correct; b = 0.96, z = 2.64, bootstrapped 95 % CI: [0.31, 2.10], p = .008). As expected, we no longer found a main effect of condition: children were as likely to answer correctly in dual identity and change of location conditions (b = -0.05, z = -0.23, bootstrapped 95 % CI: [-0.53, 0.45], p = .819). We found no main effect of trial (b = -0.30, z = -1.44, bootstrapped 95 % CI: [-0.76, 0.10], p = .150). As can be seen in Fig. 6, children performed at ceiling from four years on in true belief trials. From age of 5;1 years on they performed significantly above chance level in false belief trials.

Overall, the present results largely replicate those of Study 1: As in Study 1's control condition, children mastered our task in true belief versions from early on. However, trials that involved misrepresentations were only reliably solved from 5 to 6 years on. The (unexpected) better performance in dual identity conditions found in Study 1 disappeared in Study 2. There was no longer a difference between dual identity and change of location condition once they were adjusted in line with previous studies (Proft et al., 2019; Rakoczy et al., 2015; Schünemann, Proft, et al., 2021).

4. General discussion

The guiding question of the present work was when children develop a solid grasp of the subjectivity and aspectuality of intentions: agents may intentionally perform a given action under one description, but not necessarily under another. Previous studies found competence not before age 6. The rationale of the present study was to test whether these results may have masked earlier competence due to linguistic performance demands posed by complex and ambiguous test questions. To this end, we developed a new task to investigate children's understanding of the subjectivity of intentions. Instead of judging whether an action was intentional under a description A or B, children had to infer whether some behavior was an action at all given relevant background information about the agent's beliefs and desires. Children saw outcomes that could have been the result of intentional actions or brought about by other external factors (the wind). These outcomes were (or were not) consistent with the subjective rational perspective of the agent, and it was thus plausible to infer that they were brought about by an intentional action (or by some external event like the wind). The new task had the advantage of eliminating the need to understand the linguistic subtleties of the test question (such as the de re/de dicto distinction).

4.1. *Summary of main findings*

The results were the following: First, children mastered our new task in critical conditions involving misrepresentations only from 5 to 6 years on. Conditions that did not involve any misrepresentations were solved from 4 years on. Second, children performed similarly in conditions with various forms of subjective beliefs (false beliefs about locations or identities) once conditions were matched in complexity and task demands.

4.2. *Differences between consistent and inconsistent outcomes*

In both studies, we noted, as a third set of findings, a surprising trend: children performed better in inconsistent trials than in consistent ones. This trend was especially strong in false belief versions of the task. In both outcome types children knew about the treasure's actual location which differed from the agent's belief (in false belief versions), and had to keep track of the relation of the

Table 4
Percentage of correct responses per age group, belief manipulation and condition in Study 2

Age in years	True belief		False belief	
	Change of location	Dual identity	Change of location	Dual identity
4 (n = 32)	85 %	84 %	48 %	47 %
5 (n = 29)	90 %	83 %	52 %	64 %
6 (n = 28)	89 %	84 %	71 %	77 %
7 (n = 30)	93 %	92 %	80 %	77 %

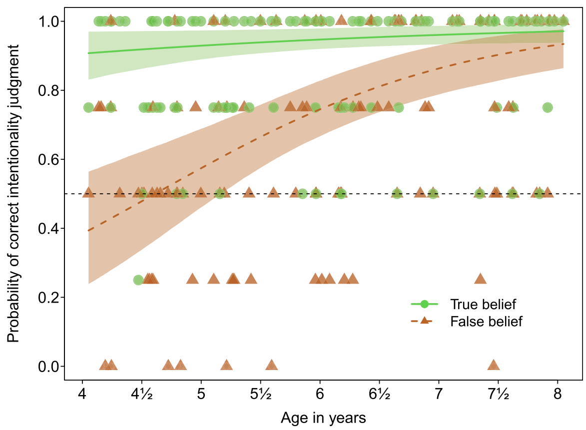


Fig. 6. Probability of correct responses across age by belief. *Note.* Since the performance no longer differed between conditions (Change of location and Dual identity), they are summarized in the plot. The horizontal dashed line represents the chance level of 0.50. Each filled dot represents the response proportion of one participant per belief type. Each fitted lines indicates the predicted probability of a correct response per condition by the binomial mixed effects model. Areas around the fitted lines are 95 % confidence intervals. Predicted values and their confidence intervals have been obtained via bootstrapping with 10,000 boots.

agent's subjective belief and objective reality. Why then the asymmetry in performance between inconsistent and consistent trials? One possibility is that children, for whatever reasons, have a baseline preference for saying it was the wind (correct in inconsistent/incorrect in consistent) rather than the agent (vice versa), and that this preference underlies the asymmetry. While this is possible in principle, the results of the control condition (where children solve both consistent and inconsistent trials proficiently) seem to speak against it. Another possibility is that the difference in performance between outcome types may stem from the fact that the consistent trials were more ambiguous than the inconsistent trials. In consistent trials, it was plausible and possible that the outcome was the product of an intentional action. But strictly speaking, the outcome could have been caused both by the agent and the external force, the wind. In inconsistent trials, in contrast, given the agent's beliefs and desires and assuming rationality, it *had* to be the wind. This difference in ambiguity was a necessary byproduct of the current design. However, the different levels of ambiguity are not a serious problem for the existing results. On the one hand, it is clear from the children's responses in the control conditions that they make the correct matching (i.e., consistent outcome = intentional action; inconsistent outcome = wind). Furthermore, our research question focused on whether this matching tendency is different across age in conditions when the manipulation of the protagonist's beliefs is added. Future studies may investigate whether performance differences in consistent vs. inconsistent conditions persist in less asymmetrically ambiguous implementations. Furthermore, future research should aim at disentangling possible performance factors and further minimizing task complexity.

4.3. Theoretical implications and further directions

Our results, in line with previous studies (Proft et al., 2019; Schünemann, Proft, et al., 2021) further support a competence deficit account: Children's late successful performance in subjective intention tasks was unlikely a performance deficit due to linguistic task demands. Our results suggest that children's struggle to ascribe intentionality in previous studies is probably not due to a confusion of what kind of reading – de re vs. de dicto – has to be applied to a question. This suggests that children probably do not find subjective intentions complicated because they do not understand the linguistic subtleties in previous test questions. To refer to the Taylor Swift example, what children found difficult was probably not a merely linguistic distinction between sentences ("Can we say 'Grandmother saw Taylor Swift'?"). It is, however, still possible that linguistic performance factors play some role. The absence of evidence in these two studies is no evidence of absence. Future studies could also look at this relationship in even more ecologically valid settings.

The present studies further suggest that advanced forms of subjective action understanding does develop later in childhood. A more nuanced picture of the developing understanding of intentions is necessary. Children do not yet understand intentions in their fully-fledged metarepresentational, subjective sense before the age of 5 or 6 years. Claims that young children or even infants understand other's intentions need to be qualified. Younger children understand acting intentionally vs. accidentally, but not necessarily acting with a fine-grained intention: when it depends on the description whether the action was intentional under this specific aspect (see Searle, 1983).

In broader context, intentions can be subject and aspectual in various ways: Here, we introduced the belief manipulation because it is a clear and probably the simplest example of how to create subjective, aspectual intentions (e.g., if Oedipus intentionally arranges a date with Iocasta, he does not intentionally arrange a date with his mother). There are other forms in which intentions can be subjective and aspectual such as in the distinction between intended main and foreseen side effects (one performs action A in order reach goal G while knowing, but not caring, that A will also lead to side effect S). However, these cases are more difficult for children (and adults) to understand, which is why in this study we have limited ourselves to understanding subjective intentions based on misrepresentation. However, for a holistic understanding of the development of subjective intentions, more studies are needed that examine other forms of subjective, aspectual intentions.

Still, it is unclear why children's understanding of the subjectivity of intentions shows such a protracted development. Although meta-representational skills, particularly belief understanding, are necessary, they are not sufficient for subjective intention understanding (Schünemann, Proft, et al., 2021). These skills typically emerge already around age 4 (Wellman et al., 2001; Wimmer & Perner, 1983), so the question remains why there is this developmental gap. What is missing? Apart from linguistic performance factors masking the competence, what are potential explanations? Plausible candidates include executive functioning, higher order Theory of Mind or recursive thinking all of which develop in protracted ways and may be necessary to acquire mature subjective action understanding. So far, however, existing studies have not found conclusive evidence that second-order false belief understanding or executive functions predicted subjective intention understanding (Schünemann, Proft, et al., 2021). Here, we can only speculate what other potential explanations there may be.

Other potentially relevant factors include counterfactual or causal reasoning. One possibility is that subjective intention understanding involves an understanding of counterfactual dependency between intentions and actions (e.g., Byrne, 2016; Halpern & Kleiman-Weiner, 2018; Paul, 2021). Consider the following example: "An agent performs an action that is both X (going to this island) and Y (going to an island without treasure)". To say that the agent performed X intentionally, but did not intentionally perform Y, means, among other things, this: had the agent known that Y applies to his action, he would not have performed it (but would have gone to the other island). Or take the distinction between intended main and merely foreseen side-effects. To say that an agent performed an action A that has effect P and Q in such a way that she intended to bring about P while merely foreseeing and not caring about Q means, among other things this: had the agent not believed that A leads to P, she would not have performed A; but if she had not believed that A leads to Q, that would have made no difference to her intentionally performing A.

Perhaps, thus, counterfactual reasoning is a more general, crucial developmental ingredient in the acquisition of understanding subjective and aspectual intentions. We know from many studies that mature counterfactual reasoning also develops in protracted ways (Byrne, 2016; Kominsky et al., 2021) and in fact, often in close parallel with mature social cognition (Rafetseder et al., 2021; Redshaw & Suddendorf, 2020).

Another (perhaps complementary) possibility is the following: It may be that complex causal reasoning in general is necessary for understanding subjective intentions. In their recent suggestion for a "simple" definition of intentionality Quillien and German (2021) argue that intentionality judgments are basically causal judgments: An agent did X intentionally to the extent that X was causally dependent on how much the agent wanted X to happen (or not to happen). One question is whether this account could be extended to make room for an understanding of aspectual intentions as investigated in the present studies. For example, for simple intentionality judgements, it is clear that agents act intentionally under certain descriptions. For instance, an agent went intentionally to the cupboard, because she wanted to get the chocolate (and the chocolate is in fact there). Her intentional action was caused by her wanting to go there. For more complex aspectual cases, however, one might need to make more complex causal judgement that also involve counterfactual simulations. If the chocolate is not in the cupboard, but the agent believes so, she intentionally goes to the cupboard. However, she does not intentionally go to the empty cupboard. Her desire to get the chocolate caused her action, but not under the description "empty cupboard." For this judgment the following counterfactual simulation might be necessary: had she known that the chocolate was somewhere else she would not have gone there. So her desire caused only some part of the action. Which part of the action was caused becomes clear if we think about relevant counterfactual alternatives. Future research should explore whether and how causal and counterfactual reasoning might explain children's difficulties in tasks involving subjective intentionality more systematically.

4.4. Conclusion

The present studies investigated when children begin to grasp subjective aspectual intentions. We examined whether previous studies, indicating proficiency only from age 6 (Proft et al., 2019; Schünemann, Proft, et al., 2021), underestimated competence due to linguistic performance factors. To this end, we developed a new task that avoided linguistic subtleties of the test question. Children judged whether an outcome was likely be caused by an intentional action or not given an agent's beliefs and desires. Our results showed that children solved our new task without misrepresentation reliably from 4 years on. However, once misrepresentations were involved, children solved our task only from 5 to 6 years on. These findings converge with previous results (Proft et al., 2019; Schünemann, Proft, et al., 2021). Therefore, these and prior results taken together corroborate the general picture that children's problems to ascribe subjective intentions are more likely a competence than a performance deficit issue due to linguistic task demands. It remains unclear why children struggle so long with an understanding of subjective intentions. Future work should address potential explanations (e.g., maturing causal or counterfactual reasoning) for the developmental gap.

CRediT authorship contribution statement

Hannes Rakoczy: Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Marina Proft:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Joana Lonquich:** Methodology, Investigation, Conceptualization. **Isa Blomberg:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Data availability

Data is shared and cited in the manuscript (anoymized).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation); SFB 1528: Cognition of Interaction. We thank Marlen Kaufmann for the organization of data collection, Davida Patzak for fruitful ideas on the study design and implementation in study material, Eva Wellmer, Melissa Kulp, Elena Lorenz, and Merle Dalmer for help with data collection and reliability coding.

Appendix A. Procedure

Testing site/general

All studies were conducted in a laboratory setting. The children took part individually. Before the study, the children played with the experimenter (all were female) for about 5 to 10 minutes. In total, the study lasted between 18 and 30 minutes.

Study 1

Participants were first introduced to the task through printed versions of the scenery and objects (see also OSF). The experimenter played one consistent and one inconsistent trial with the children. Children learnt that the little pirate's goal was to collect the most valuable objects. The story established that the little pirate aimed to travel to a specific island, but his success was variable due to an upcoming storm. At times, he managed to reach his intended island (intentional action); however, when the winds were too strong, he would end up on the other island where he did not want to go.

During four familiarization trials, two different outcome types (consistent and inconsistent) were paired with two preference hierarchy constellations ("diamond more valuable than golden coin" and "fantasy object less valuable than coin"). Printed animations of a paddle and a storm were used to guide children with a test question. In familiarization trials the retrospective desire question was asked before the test question.

- Retrospective desire question: "To which island did the little pirate want to go? To this or that?" [E points to both locations, first location counterbalanced]
- Test question: "Did the little pirate travel to the island himself, or did the wind drive him there?" [E points at printed animation of options, first option counterbalanced]

Feedback was provided in response to the child's answer. The little pirate's emotional reaction, either happy or sad, was shown in the video and reiterated by the experimenter.

Test trials were similar to the familiarization trials, but included a twist where the little pirate would disappear in between. During test trials, no feedback or reaction was displayed. In test trials the knowledge-seeing question was asked after the test question:

- Change of location condition: "Did the little pirate see that the female pirate took the coin to the other island?" [correct answer: "no"]
- Dual identity condition: "The [fantasy object] is also a diamond. Does the little pirate know that?" [correct answer: "no"]

A pause after three test trials allowed for a search for stickers to maintain children's attention and motivation.

After the completion of all test trials, a preference hierarchy check was conducted. An arrow pointing upwards was used to indicate the value the little pirate placed on different items: at the top, what he found most desirable; in the middle, what he found of average interest; and at the bottom, what he found least appealing and somewhat boring. Children were then asked to place the diamond and the coin in the preference hierarchy as understood from the story. The expectation was that the diamond would be placed at the top or at least higher than the coin. They were also presented with a dual identity fantasy object and asked to consider its value if the little pirate had been aware of its second identity (diamond). Lastly, they were given a fantasy object (without a second identity) and expected to place it below the coin in terms of desirability.

Study 2

Study 2's procedure was identical to Study 1 with the following exceptions. In all trials objects were used that matched in value. Therefore, no preference hierarchy had to be learnt or checked afterwards. We used an adaptive number of familiarization trials. If children answered all control and test questions correctly, they received 2 trials. If they made a mistake they received another familiarization trial of the same outcome type. That means if a child made an error in a consistent trial, she would receive another consistent trial, but not another inconsistent trial. Children were also asked for each trial what object the little pirate wanted:

- Object control question: "What does the little pirate want to get now?"

Other than in Study 1, the knowledge control question was asked before the storm came up. Children were first asked again, if they answered wrong, and finally corrected if they insisted.

- Knowledge question - change of location condition: “The female pirate put [desired object] on the other island. Does the little pirate know that?” [correct answer true belief: “yes”/correct answer false belief: “no”]
- Knowledge question - dual identity condition: “The [desired object] is also a B. Does the little pirate know that?” [correct answer true belief: “yes”/correct answer false belief: “no”]

Additionally children were asked a belief question. Children could point to the island or verbally indicate their answer. Children were immediately correct when they answered incorrectly.

- Belief question: “Where does the little pirate think the [desired object/identity A] is?”

Appendix B. A priori sample size, and exclusion criteria

Study 1

The sample size of $N = 120$ (complete data) was determined a priori via data simulation. Due to scheduling errors and initially incorrect calculation of the exclusion, we exceed our sample by 7 children. However, we decided to include these children in the analysis. This was particularly useful given that the sample size calculation is based on complete data and we have partial exclusion of the data.

Additional 34 children took part, but were excluded from the data. Based on the preregistration, $n = 19$ children who made errors in test or control questions the last 2 familiarization trials, or made overall less than 75 % correct in familiarization trials, were excluded. Additionally, beyond the preregistration, we excluded ($n = 15$) children who made a crucial error in the hierarchy checks. These children falsely claimed that the little pirate valued the coin more than the diamond.

Beyond the preregistration, we excluded $n = 71$ (9 % of the data) single trials. $N = 34$ responses in dual identity trials were excluded, because children falsely claimed that the fantasy object that was also a diamond would be regarded as less valuable or equally valuable as the coin. $N = 26$ responses in control trials were excluded, because children falsely claimed that the fantasy object that without a second identity would be regarded as more valuable or equally valuable as the coin. $N = 11$ responses in dual identity or change of location trials were excluded respectively, because children falsely claimed that the little pirate know about the second identity of the object or had seen the coins transfer.

Study 2

Two children were included in the final sample but aborted the study after one test trial. Differently from preregistration, we decided not to exclude children when they answered the control question about which item the little pirate wanted. Only a few children answered this question incorrectly or could not remember. If children did not answer anything or stated that they could not remember, they were reminded.

Appendix C. Counterbalancing

Study 1

The following factors were counterbalanced between children: first option named in the test question (“drove there himself” vs. “wind drove him there”) and respectively printed out option visualizations, and the island that was pointed to first (left vs. right). The following factors were counterbalanced with restrictions: order of outcome type trials – consistent (C) vs. inconsistent (I) (familiarization trials: “ICCI” vs. “CIIC”; test trials: “CICICI”, “ICICIC”), order of test trials (“dual identity, control, change of location” vs. “change of location, control, dual identity”), outcome – left (L) vs. right (R) (test trials: “RRLLLR”, “LLRRRL”, “LRRRLL”, “RLLLRR”, “RRLRRL”, “LLRLLR”, “LRRLRR”, “RLLRLL”). This resulted in a total of 32 orders. Children were randomly assigned to a counterbalancing order.

Study 2

The counterbalancing was preregistered and can be found on OSF. With the same logic, we counterbalanced the first option named in test question, the island that was pointed to first, the order of belief trials (4x true belief first, 4x false belief first vs. vice versa), the order of condition trials that came first (dual identity vs. change of location). We counterbalanced the following factors with restriction: the objects used, the order of outcome type trials – consistent (C) vs. inconsistent (I) (familiarization trials: “ICCI” vs. “CIIC”; test trials: “CIICCIIC” vs. “ICCIICCI”) and the outcome – left (L) vs. right (R) (“RRLLLRRL”, “LLRRRLLR”, “LRRRLLRL”, “RLLLRRRL”, “RRLRRLRL”, “LLRLLRLR”, “LRRLRRLR”, “RLLRLLLR”). That resulted in a total of 64 orders. Children were assigned in a balanced way to the orders per age group. All age groups received the dual identity or change of location conditions first and the true or false belief trials first, in roughly equal parts.

References

Anscombe, G. E. M. (1979). Under a description. *Noûs, 13*(2), 219. https://doi.org/10.2307/2214398

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

Apperly, I. A., & Robinson, E. J. (2003). When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology, 85*(4), 297–311. https://doi.org/10.1016/S0022-0965(03)00099-7

Astington, J. W. (2001). The paradox of intention: Assessing children's metarepresentational understanding. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 85–103). Massachusetts Institute of Technology.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41*(2), 328–337. https://doi.org/10.1037/0012-1649.41.2.328

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology, 67*(1), 135–157. https://doi.org/10.1146/annurev-psych-122414-033249

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). "Unwilling" versus "unable": Chimpanzees' understanding of human intentional action. *Developmental Science, 7*(4), 488–498. https://doi.org/10.1111/j.1467-7687.2004.00368.x

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315–330. https://doi.org/10.1016/S0163-6383(98)90009-1

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborativeecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351. https://doi.org/10.21105/joss.05351

Filippova, E., & Astington, J. W. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development, 81*(3), 913–928. https://doi.org/10.1111/j.1467-8624.2010.01442.x

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience Biobehavioral Reviews, 82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Flavell, J. H. (1986). The development of children's knowledge about the appearance–reality distinction. *American Psychologist, 41*(4), 418–425. https://doi.org/10.1037/0003-066X.41.4.418

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), 755. https://doi.org/10.1038/415755a

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

GoAnimate Inc. (2024). *Vyond*.⟨https://www.vyond.com/⟩.

Halpern, J., & Kleiman-Weiner, M. (2018). Towards formal definitions of blameworthiness, intention, and moral responsibility. *Proceedings of the AAAI Conference on Artificial Intelligence, 32*(1). https://doi.org/10.1609/aaai.v32i1.11557

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

Kamawar, D., & Olson, D. R. (2011). Thinking about representations: The case of opaque contexts. *Journal of Experimental Child Psychology, 108*(4), 734–746. https://doi.org/10.1016/j.jecp.2010.10.005

Kominsky, J. F., Gerstenberg, T., Pelz, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology, 57*(2), 253.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology, 4*(3-4), 231–244. https://doi.org/10.1556/JCEP.4.2006.3-4.3

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Nelson, M. (2023). Propositional attitude reports. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. ⟨https://plato.stanford.edu/archives/spr2023/entries/prop-attitude-reports/⟩.

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior, 36*(4), 313–322. https://doi.org/10.1016/j.evolhumbehav.2015.01.004

Paul, S. K. (2021). *Philiosophy of action*. Routledge.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

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

Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). "Unwilling" versus "unable": Capuchin monkeys' (Cebus Apella) understanding of human intentional action. *Developmental Science, 12*(6), 938–945. https://doi.org/10.1111/j.1467-7687.2009.00840.x

Proft, M., Schünemann, B., & Rakoczy, H. (2019). Children's understanding of the aspectuality of intentions. *Journal of Experimental Child Psychology, 181*, 17–33. https://doi.org/10.1016/j.jecp.2018.12.001

Quillien, T., & German, T. C. (2021). A simple definition of "intentionally". *Cognition, 214*, Article 104806. https://doi.org/10.1016/j.cognition.2021.104806

R Core Team. (2023). *A language and environment for statistical computing*. R Foundation for Statistical Computing. ⟨https://www.R-project.org/⟩.

Rafetseder, E., O'Brien, C., Leahy, B., & Perner, J. (2021). Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective. *Journal of Experimental Child Psychology, 204*, Article 105058. https://doi.org/10.1016/j.jecp.2020.105058

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology, 1*(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z

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

Redshaw, J., & Suddendorf, T. (2020). Temporal Junctures in the Mind. *Trends in Cognitive Sciences, 24*(1), 52–64. https://doi.org/10.1016/j.tics.2019.10.009

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

Schidelko, L. P., Proft, M., & Rakoczy, H. (2022). How do children overcome their pragmatic performance problems in the true belief task? The role of advanced pragmatics and higher-order theory of mind. *PLoS One, 17*(4), Article e0266959. https://doi.org/10.1371/journal.pone.0266959

Schünemann, B., Keller, J., Rakoczy, H., Behne, T., & Bräuer, J. (2021a). Dogs distinguish human intentional and unintentional action. *Scientific Reports, 11*(1), 14967. https://doi.org/10.1038/s41598-021-94374-3

Schünemann, B., Proft, M., & Rakoczy, H. (2021b). Children's developing understanding of the subjectivity of intentions – A case of "advanced theory of mind". *Journal of Cognition and Development*, 1–23. https://doi.org/10.1080/15248372.2021.2003366

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

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

Völter, C. J., Lonardo, L., Steinmann, M. G. G. M., Ramos, C. F., Gerwisch, K., Schranz, M.-T., Dobernig, I., & Huber, L. (2023). Unwilling or unable? Using three-dimensional tracking to evaluate dogs' reactions to differing human intentions. *Proceedings of the Royal Society B: Biological Sciences, 290*(1991), Article 20221621. https://doi.org/10.1098/rspb.2022.1621

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

Wimmer, H., & Perner, J. (1983). Belief about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science, 317*(5843), 1402–1405. https://doi.org/10.1126/science.1144663

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4