

## The semantics and pragmatics of belief reports in preschoolers\*

Shevaun Lewis  
*University of Maryland*

Valentine Hacquard  
*University of Maryland*

Jeffrey Lidz  
*University of Maryland*

**Abstract** Children under 4 years have been claimed to lack adult-like semantic representations of belief verbs like ‘think’. Based on two experiments involving a truth-value judgment task, we argue that 4-year olds’ apparently deviant interpretations arise from pragmatic difficulty understanding the *relevance* of belief, rather than from conceptual or semantic immaturity.

**Keywords:** attitudes, belief reports, parentheticals, Theory of Mind, false belief

### 1 Introduction

Suppose John saved a cupcake in the fridge, but unbeknownst to him, Mary stole it and hid it in a drawer. If we asked a 3-year-old where John will look for it when he gets hungry, she would unhesitatingly respond that he will look in the drawer, because that’s where it is. This result has been replicated in hundreds of studies, with thousands of children from different linguistic and cultural backgrounds (for review: Wellman, Cross & Watson 2001). Although some task factors can affect children’s success to a limited degree, a robust generalization stands uncontested: children do not consistently respond correctly until age 4, and young 3-year-olds usually respond incorrectly—worse than if they guessed at random.

Until recently, this generalization was taken as evidence that children under 4 years cannot attribute false beliefs to others. Given this hypothesis, it would not be surprising if children’s interpretations of sentences with belief verbs like ‘think’ were not adult-like, particularly in contexts involving false belief. And indeed, if we asked the 3-year-old where John *thinks* the cupcake is, she would also answer “in the drawer” (Wellman, Cross & Watson 2001; de Villiers 1995;

---

\* Many thanks to our undergraduate research assistants (Sam Blitzstein, Faina Kostyukovsky, Jessica Lee, and Leah Whitehill), and the Attitudes and Theory of Mind research group (Rachel Dudley, Kate Harrigan, Naho Orita, Carolina Peterson, and Aaron Steven White). This work was funded in part by NSF grant #BCS-1124338.

de Villiers & Pyers 2002). The statement in (1) would be rejected on the grounds that the cupcake is not in the fridge (Sowalsky, Hacquard & Roeper 2009).

(1) John thinks that his cupcake is in the fridge.

We consider three competing hypotheses to account for young children's responses to 'think' sentences and for the development that occurs around age 4. Under the Non-linguistic Development Hypothesis, the main change that occurs is the development of the ability to attribute mental states to others. Under the Semantic Development Hypothesis, the conceptual basis for understanding belief reports is already in place. What children must acquire are the syntactic and semantic structures that allow them to represent mental states linguistically. Under the Pragmatic Development Hypothesis, children have already acquired the conceptual and syntactic/semantic representations, and what develops is their understanding of the relevance of beliefs in certain contexts.

## 2 The semantics and pragmatics of 'think'

A hallmark of attitude constructions is that they can be true even if their complement is false in the world of evaluation. All that is required for (1) to be true in world  $w$  is that 'the cupcake is in the fridge' be true in John's "belief worlds" (Hintikka 1971), not in  $w$ . Yet in certain contexts, the speaker endorses the truth of the complement in  $w$ : in so-called "parenthetical" uses (Urmson 1952), the complement clause carries the main point of the utterance (and is thus endorsed by the speaker) and the main clause with the attitude verb is demoted to parenthetical status, serving an evidential-like function (Rooryck 2001a).

(2) Mary is at the movies, {I think/John thinks}.

Parenthetical interpretations are often associated with syntactic phenomena like preposing, as in (2) (Ross 1973; Bolinger 1968; Hooper 1975). This has led to syntactic accounts in which the parenthetical verb sits at the head of a functional projection for evidential markers (Rooryck 2001a; 2001b) or a sentence adverbial (Bresnan 1968). Yet parenthetical uses do not require (overt) parenthetical syntax: sentences with standard word order can receive either mental state or parenthetical interpretations, depending on whether it is the mental state itself that is most relevant in context, as in the exchange in (3), or the content of the mental state, as in (4).

- (3) A: Why is John mad at Bill?  
B: John thinks Bill was the thief.
- (4) A: Who stole the cupcake?  
B: John thinks Bill was the thief.

Under a pragmatic account (Simons 2007), parenthetical interpretations arise

from the way attitudes are used in context. In (4), the literal content of B's utterance is not a felicitous answer to A's question, but its complement is. A hearer will reason that B is being cooperative and intending to answer A's question. Given that the complement clause is a felicitous answer, B must endorse it (though, the hearer further reasons, B hedges somewhat; otherwise, she would have uttered the complement clause alone). The main point of B's utterance, then, is the complement clause; the attitude serves a kind of evidential function, indicating B's source of evidence. Note, finally, that although B's utterance has a different main point in (3) and (4), both instances still involve an assertion about John's belief, as evidenced by the fact that it can be subsequently denied in both cases ('No he doesn't. He thinks Mary was').

Whether the syntactic or the pragmatic analysis of parentheticals is correct—or if the two coexist—the learner must be able to recognize which contexts support a parenthetical use, and which do not. Although it is an interesting question whether and when children acquire parenthetical syntactic structures, here we are primarily concerned with pragmatic competence.

We ask whether both parenthetical and non-parenthetical uses are accessible to children, and to what extent.

### **3 Belief attribution in development**

#### **3.1 Belief attribution at the conceptual (non-linguistic) level**

Although it is difficult to define what a "concept" of belief is, we can enumerate observable consequences of having acquired it. A person with a mature concept of belief should understand how events can give rise to belief states, and how belief states in turn give rise to actions or emotions. The standard false belief task requires children to do both: they must infer a character's belief state based on his past actions, and then predict his future actions based on that belief state.

In Wimmer and Perner's (1983) now classic change-of-location paradigm, children are told a story with two characters. In a representative story, Maxi helps his mom put away groceries. He puts some chocolate in the blue cupboard before going out. While he is away, his mom uses the chocolate to make a cake. In the "Standard" condition, she puts the leftover chocolate in the green cupboard instead of back in the blue cupboard. In the "Disappear" condition, she uses all the chocolate. Maxi returns to eat some chocolate. At this point, the child is asked where he will look for it. The correct response is that he will look in the blue cupboard, since that's where he left it. Wimmer and Perner tested 3-, 4-, and 5-year-olds. In the "Standard" condition, none of the 3-year-olds' and only around half of the 4- and 5-year-olds responded correctly. The "Disappear" condition was somewhat better: most 3-year-olds and nearly all 4- and 5-year-olds were correct.

A host of additional studies followed with various manipulations of the task. Wellman, Cross and Watson's (2001) meta-analysis of 178 studies found that the age trend reported in Wimmer & Perner 1983 is robust across studies: children perform significantly below chance until about 41 months, and above chance beginning around 48 months. The type of question (what the protagonist thinks, knows, or will do) has no effect on children's performance. Four factors affect performance: children are more accurate when (i) the false belief is created with a deceptive motive, (ii) the participant actively helps create the false belief (e.g. by moving the object), (iii) the object is not present at the end (as in the "Disappear" condition above), or (iv) the protagonist's false belief is explicitly stated or pictured. However, since these factors do not interact with age (they help all children equally, regardless of age), the authors conclude that the improvement from 3 to 4 years of age is due to real conceptual development rather than an increased ability to deal with the challenges of the task.

Nevertheless, many researchers have argued that since traditional false belief tasks require an explicit decision and response, they are fundamentally flawed as measures of young children's underlying competence. Clements and Perner (1994) used both "implicit" and "explicit" measures to test understanding of false belief. Children watched a change-of-location story. Their gaze direction (to the true or false location of the object) was recorded as they heard the prompt in (5). A few seconds later they were asked the standard direct question, as in (6).

(5) I wonder where he's going to look?

(6) Sam wants to get the cheese. Which box will he open first?

Looks to the false location after the first prompt (the "implicit" task) and points to it after the question (the "explicit" task) were considered correct. 3-year-olds (2;11-3;2) performed much better on the implicit task (>75%) than the explicit task (<25%). The authors speculate that an explicit judgment requires comparing multiple representations of the situation, and this comparison is what is difficult for young children.

More recent work suggests that even young infants have some understanding of false belief. Onishi and Baillargeon (2005) found that 15-month-olds show "surprise" in a looking-time paradigm when an agent's actions conflict with her beliefs, even when the beliefs are false. Subsequent studies have shown that children as young as 24 months (Southgate, Senju & Csibra 2007), 13 months (Surian, Caldi & Sperber 2007), and even 7 months (Kovács, Téglás & Endress 2010) are sensitive to false beliefs in implicit tasks (for review: Baillargeon, Scott & He 2010). In light of these findings, it is unlikely that the development in the preschool years is the emergence of a new ability to attribute false beliefs.

### 3.2 Linguistic belief reports

### 3.2.1 Spontaneous production

Shatz, Wellman and Silber (1983) collected spontaneous speech samples from 2-3 year-olds in a longitudinal study. Most children produced at least one mental verb, but very few of their utterances unambiguously referred to mental states. Over half were instances of 'I don't know'. The rest were mainly parenthetical, functioning to direct the conversation, as in (7), or modulate the degree of certainty, as in (8).

(7) I thoughted we'd eat some cake.

(8) I think this is a lamb.

Bloom, Rispoli, Gartner, and Hafitz (1989) found that children's (aged 2;0-3;2) use of mental verbs in spontaneous speech was restricted compared to the input. For example, almost all instances of 'think' had a first person singular subject. Like Shatz et al. (1983), they concluded that in the majority of such utterances, the main clause verb is parenthetical. 'Think' and 'know' were usually used to modulate the degree of certainty.

Diessel and Tomasello (2001) found that the "vast majority" of early uses of complement-taking verbs by 2-4 year-olds were "formulaic", with other "performative" (our "parenthetical") and "assertive" uses emerging around age 4.

In summary, studies of children's spontaneous production suggest that before age 4, they mainly use attitude verbs for parenthetical functions. However, this pattern may not be representative of their underlying competence: production often lags behind comprehension in development. Furthermore, parenthetical uses of attitude verbs are very common in adult speech: Diessel and Tomasello (2001) note that in parents' speech in the transcripts they studied, formulaic uses were several times more frequent than other uses. Children's usage may simply reflect the distribution in the input.

### 3.2.2 Comprehension

Moore, Bryant, and Furrow (1989) investigated children's understanding of 'think' and 'know' in cases where they are used to express the speaker's degree of certainty. In a hiding game, children looked for a piece of candy in one of two locations based on the advice of two puppets. One puppet would say, *I know it's in the blue box*, the other *I think it's in the red box*. A child's response was considered correct if they looked in the box suggested by the puppet who said he *knew*. 3-year-olds were at chance, while 4-year-olds mostly chose the correct location. These results suggest that 3-year-olds' understanding of the parenthetical uses of belief verbs may be more tenuous in comprehension than in production.

Johnson and Maratsos (1977) investigated 3- and 4-year-olds' comprehension of *think* and *know* in situations involving true and false beliefs. In a representative

story, “Sally played a trick on John. While John wasn’t looking, Sally took his toy duck and hid it under this box. But she played a trick on John. She told him it was under this [other] box, and he believed her.” Children were asked a series of questions about the beliefs of Sally and John. Only children who correctly answered the question *Where will John look for the toy?* on the first or second attempt were asked the remaining questions. All 4-year-olds passed this criterion and performed fairly well on the belief questions. By contrast, less than half of the 3-year-olds answered this question correctly on the first try. They answered ‘yes’ to all belief questions (e.g. *Does John/Sally think/know the toy is under this box?*) as though both characters were in states of true and false belief simultaneously. The authors concluded that a “sophisticated understanding of mental verbs” is emerging in 4-year-olds, but is quite limited in 3-year-olds.

De Villiers designed a test of children’s ability to represent false complements that minimizes conceptual demands as much as possible (de Villiers 1995; de Villiers & de Villiers 2000; de Villiers & Pyers 2002). Children are presented with a story in which a character made a mistake, tells a lie, or has a false belief. The sentence with an attitude verb and false complement is provided directly, as in (9); children only need to remember the complement to succeed.

- (9) This girl saw something funny at a tag sale and paid a dollar for it. She thought it was a toy bird but it was really a funny hat. What did she think she bought?

De Villiers & Pyers (2002) tested 3-4 year-olds three times over about 7 months. In the first round, fewer than 30% passed; by the last round, about 90% passed. These results converge with the previous studies, suggesting that children are not able to represent false complements until around age 4.

Sowalsky, Hacquard, and Roeper (2009) used a truth-value judgment task to test 2-5 year-olds’ understanding of sentences with ‘think’ and ‘according to’, like those in the examples in (10) and (11). In some stories a character had a false belief; in others it was unknown whether the belief was true or false.

- (10) Puppy thinks that it is raining outside.  
 (11) According to Turtle, it is snowing outside.

2-3 year-olds had difficulty with both sentence types, but were better with ‘according to’ (66% correct) than ‘think’ (35%). 4-5 year-olds were near ceiling with ‘according to’ (90%), but still had difficulty with ‘think’ (4-year-olds: 56%; 5-year-olds: 67%). All age groups were less accurate with false belief stories (52% overall) than when the reality was unspecified (66%). The relatively poor performance of 4-5 year-olds on ‘think’ is somewhat surprising given their success in previous studies (Johnson & Maratsos 1977; de Villiers 1995), suggesting that the truth-value judgment task might be a more rigorous test.

### **3.3 Summary and hypotheses**

The overall picture that emerges from previous studies is that while children have some implicit understanding of belief states very early on, they are not able to make use of that knowledge in explicit reasoning tasks until around age 4. Similarly, although children are producing ‘think’ by age 3, they are not adult-like in their production or comprehension until 4 years, if not older. Their early productions are generally parenthetical in function. In comprehension, they show consistent difficulty with false complements. We can now return to our three broad hypotheses about the nature of the development that occurs around age 4.

#### **3.3.1 The Non-linguistic Development Hypothesis**

Under the Non-linguistic Development Hypothesis, the primary change that occurs around 4 years is the ability to attribute false beliefs to others. This conceptual development is directly responsible for children’s improving performance on explicit false belief tasks, including those that test ‘think’.

The recent evidence concerning young infants’ implicit understanding of false belief undermines the strong claim that children lack any concept of belief until age 4. We will consider a weaker version of this hypothesis: although children are able to attribute false beliefs to others, it is a cognitively demanding task that easily breaks down under stress. Standard false belief tasks prompt children to engage in an explicit reasoning process that contrasts someone else’s (false) belief with their own (true) belief. Considering both belief states simultaneously may overwhelm children’s limited processing capacity, causing them to respond based on the most salient and available representation—their own belief state. This hypothesis gains support from the fact that children’s performance on false belief tasks is improved when the conflict with their own beliefs is reduced (Wellman, Cross & Watson 2001)—for example, when the object is removed from the scene in Wimmer and Perner’s (1983) “Disappear” condition.

Under this version of the hypothesis, the relevant developments are in general processing capacity and the ability to resolve representational conflicts. We need not be specific about which non-linguistic capacities are important: we will simply assume that whatever factors make an explicit false belief task difficult should equally affect tasks testing ‘think’ sentences in false belief scenarios.

#### **3.3.2 The Semantic Development Hypothesis**

According to the Semantic Development Hypothesis, what children acquire between the ages of 3 and 4 are the syntactic and semantic structures necessary to represent propositional attitudes. Based on the production data, it seems that 3-

year-olds can use ‘think’ parenthetically, perhaps by lexicalizing frequent expressions like ‘I think’. Given this assumption, there are at least two possible explanations for their difficulty with false complements in comprehension.

One possibility is that children use their parenthetical representation to interpret all sentences with ‘think’, and thus always judge such sentences based on the truth of the complement in the actual world. In this case, children must learn a new syntactic structure—a finite clause embedded under an attitude verb—as well as how to evaluate complement clauses with respect to worlds other than the actual one. This version of the hypothesis is similar to what has been proposed by Jill de Villiers (2000; 2005; 2007; de Villiers & Pyers 2002; de Villiers & de Villiers 2009). De Villiers goes further to say that the acquisition of these linguistic representations is what allows children to begin succeeding on standard explicit false belief tasks, perhaps by un-encapsulating belief representations in the social cognition system, making them available to higher-level reasoning processes. This hypothesis gains support from the apparent tight connection between linguistic and non-linguistic measures of false belief understanding observed in longitudinal and training studies (Astington & Jenkins 1999; de Villiers & Pyers 2002; Hale & Tager-Flusberg 2003).

A second possibility is that children do have a representation for non-parenthetical uses, but their frequent experience with parenthetical ‘think’ has led them to assume that the complement of ‘think’ is always true. Their semantics for ‘think’ would be comparable to factive ‘know’ or a non-factive that means ‘think correctly’. In this case, children would have to unlearn this semantics for ‘think’ based on positive evidence that it can be used with false complements.

### 3.3.3 The Pragmatic Development Hypothesis

Under the Pragmatic Development Hypothesis, children have acquired an adult-like semantic representation of ‘think’ earlier than their performance on comprehension tasks would suggest. They have difficulty in tasks involving sentences with false complements because they tend to interpret ‘think’ parenthetically in situations where adults would not.<sup>1</sup> That is, although they have both parenthetical and non-parenthetical interpretations of ‘think’ available, they do not have an adult-like understanding of which contexts license a parenthetical use, and which do not. They tend to underestimate the relevance of belief states, and the probability that a speaker is intending to comment on someone’s belief

---

<sup>1</sup> A variant of this pragmatic hypothesis would be that children have an adult semantics for *think*, but pragmatically strengthen *think* to *think correctly* under the assumption that people’s beliefs are generally correct (Philippe Schlenker, p.c.). Under either variant, we expect performance to be affected by contextual manipulations like the one used in Experiment 2.

state rather than the state of the actual world.

This hypothesis is consistent with children's performance on standard false belief tasks. In a change-of-location task, children could easily misconstrue *Where will Maxi look for the chocolate?* as a question about where he ought to look, or as an indirect request for the child to help him find the chocolate. Children are more likely to understand that the question is about Maxi's belief state if their attention is drawn to the false belief during the story, as in variations of the task where the child helps create the false belief or when deception is otherwise emphasized.

## **4 Experiment 1**

At 4 years, children are beginning to show adult-like competence with false belief and false complements, but their performance on many tasks is still vulnerable. The goal of Experiment 1 was to test the hypothesis that children's non-adult-like interpretations of 'think' arise from non-linguistic difficulty understanding false belief situations. We did this in two ways. First, we compared children's performance with 'think' vs. 'say', as well as with 'according to', to see if we would replicate previous results that children perform better with 'according to' than with 'think' (Sowalsky, Hacquard & Roeper 2009). Second, we tested whether children's performance improves when their own (true) beliefs do not conflict with those of the attitude holder, as has been found in non-linguistic false belief tasks as well as Sowalsky and colleagues' truth-value judgment task.

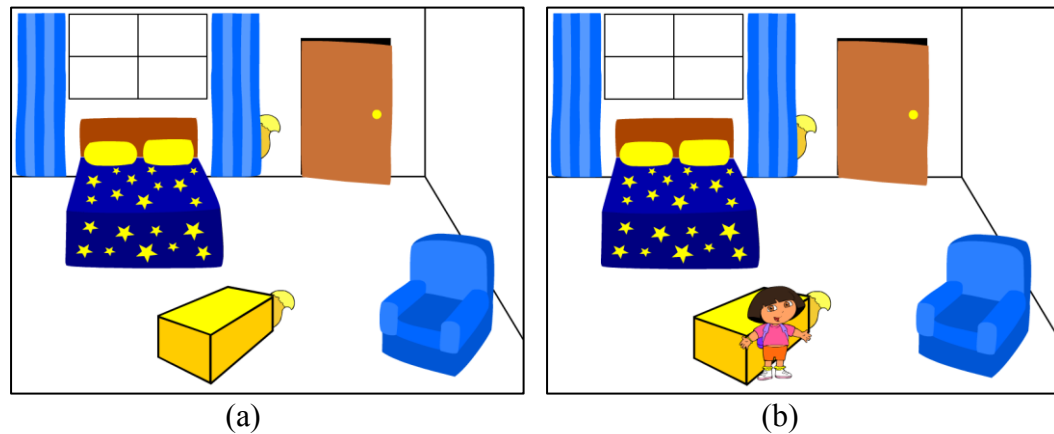
### **4.1 Subjects**

47 4-year-old children (3;9-4;3, mean 4.0; 16 boys) participated in the study. Participants were recruited from the Center for Young Children or the Infant Studies Database at the University of Maryland. All participants were typically-developing monolingual English-speakers. Data from 4 additional children were excluded because they could not finish the task (n=2), their native language was not English (n=1), or because of experimenter error (n=1).

### **4.2 Design**

Children were presented with stories about hide-and-seek. After each story, a puppet uttered a target sentence with 'think', 'say', or 'according to', and the child was asked to judge whether the puppet was "right" about what happened.

#### **4.2.1 Sample story**



**Figure 1** (a) Identical clues for the Hider (Swiper, behind the curtain) and Distracter (squirrel, behind the toy box). (b) The Seeker (Dora) guesses the location of the Hider.

All stories followed the same template, illustrated in the following sample story.

**Introduce characters** In the first scene, the characters (Swiper and Dora) are named and the experimenter confirms that the child can identify them. Swiper is identified as the Hider, Dora as the Seeker: *Swiper is gonna hide, and Dora will look for him. So she'll wait in the other room where she can't see.*

**Hiding** The child watches as Swiper hides behind the curtain. His yellow tail remains visible, protruding from behind the curtain. Then a squirrel (the Distracter) hides behind the toy box, leaving an identical yellow tail visible (**Figure 1a**). The experimenter points out the two clues to ensure that the child knows what evidence Dora will be using to guess Swiper's location.

**Seeking** Dora reappears to state a guess about Swiper's location based on one of the clues: *Hmm, where should I look? Oh! I see a yellow tail behind the toy box! I know--Swiper is there! I'll look for Swiper behind the toy box.* The Seeker's script is intended to establish that she is just guessing based on the first clue she noticed, but she is nevertheless confident—she believes what she's saying. Dora moves toward the toy box as she speaks and remains there for the rest of the story as a cue to her stated belief (**Figure 1b**).

**Target sentences** At this point, the experimenter asks the puppet to say something about what's going on in the story. The puppet delivers the target sentence, which contains either 'think', 'say', or 'according to':

- (12) a. Dora thinks that Swiper is behind the toy box.  
 b. Dora said that Swiper is behind the toy box.  
 c. According to Dora, Swiper is behind the toy box.

<b>Sample sentence</b>	<b>Belief Type</b>	<b>Sent. Truth</b>	<b>Comp. Truth</b>
Boots thinks that Swiper is behind the curtain.	TB	T	T
Boots thinks that Swiper is behind the toybox.	TB	F	F
Dora thinks that Swiper is behind the toybox.	FB	T	F
Dora thinks that Swiper is behind the curtain.	FB	F	T

**Table 1** Target sentence types for Experiment 1, with BELIEF TYPE (TB = *true belief*; FB = *false belief*), truth of the sentence, and truth of the complement clause.

After the child responds, the puppet delivers a filler sentence.

**Reveal hiders** Once the child has responded to both the target sentence and the filler, the Hider and Distracter emerge from their hiding places.

#### 4.2.2 Manipulations

Within the stories, we manipulated whether the child had KNOWLEDGE of the Hider’s true location. In the *knowledge* condition, the child watched as the Hider and Distracter hid in the scene (as in the sample story). In the *ignorance* condition, the screen was obscured during the hiding, so the child did not know which clue corresponded to the Hider until after responding to the sentences.

The main VERB of the target sentence was manipulated between subjects: children heard sentences with either ‘think’, ‘say’, or ‘according to’. We manipulated whether the target sentence referred to a Seeker with a *true belief* or a *false belief* (BELIEF TYPE). In the sample story, the target sentences are about a *false belief*. In the *ignorance* condition, it is unknown at the point of the target sentence whether the Seeker in question has a true or false belief. The truth of the target sentences (i.e., the target response) was counterbalanced. **Table 1** shows the set of possible ‘think’ target sentences for the sample story. Note that the *false belief* sentences are those where the truth of the complement clause in the actual world conflicts with the truth of the whole sentence.

#### 4.3 Predictions

We expected 4-year-olds to do well overall, since by that age they are beginning to pass tests involving false beliefs and false complements. However, to the extent that they are non-adult-like, our hypotheses make different predictions about which conditions will be the most difficult.

The Non-linguistic Development Hypothesis predicts that children should

have difficulty with sentences involving a *false belief*. They should have less difficulty in the *ignorance* condition, since there is no possible conflict between their own beliefs and the Seeker's. The VERB should have no effect on their performance, since the need to resolve a representational conflict arises regardless of whether the ultimate judgment is being made about thoughts or statements.

Under the Semantic and Pragmatic Development Hypotheses, children respond primarily based on the complement clause. Thus, they should look most non-adult-like in the conditions where the truth of the complement clause in the actual world conflicts with the truth of the whole sentence: the *false belief* conditions. They should have difficulty in the *ignorance* conditions, because they will have no basis for judging the truth of the complement clause. They may show different patterns of responses depending on the VERB.

#### 4.4 Materials

A list of 14 stories was created including a variety of scenes and characters. The locations of the Hider and Distracter were spread across the different hiding spots across trials, and the characters playing the Hider and Seeker rotated from story to story. Stories were illustrated and animated in Adobe Flash. Narration for each story was recorded and added to the videos.

Two lists of target sentences were created. In each list, the order of sentences with respect to BELIEF TYPE and sentence truth was pseudo-randomized. Two filler sentences, one true and one false, were created for each story. The fillers did not involve belief. They were created using templates exemplified by (13)-(17). (15)-(16) were only appropriate in *knowledge* stories, and (17) in *ignorance*.

- (13) Dora is looking for Swiper {behind the toy box/behind the curtain}.
- (14) Swiper is really hiding {behind the curtain/behind the toy box}.
- (15) There's really a squirrel {behind the toy box/behind the curtain}.
- (16) We can see a yellow tail {behind the toy box/under the bed}.
- (17) Swiper is {behind the curtain or behind the toy box/behind the door or under the bed}.

Each participant saw 2 practice trials, followed by 3 trials in each of 4 conditions (KNOWLEDGE  $\times$  BELIEF TYPE). The distribution of true and false sentences was counterbalanced across conditions. Since there were an odd number of trials per condition, the distribution is only fully balanced when both lists are taken together.

#### 4.5 Procedure

Sessions took place in a quiet room with the child seated in front of a laptop. The

experimenter sat alongside the child, operating the puppet with one hand and coding responses with the other. Sessions were videotaped so that children's responses could be coded later by an independent viewer.

The experimenter began by explaining the task, introducing the puppet ("Drog", a baby dragon who wants to learn how to play hide-and-seek), and obtaining the child's assent to participate. To ensure that the child was comfortable telling the puppet whether he was right or wrong, the experimenter asked the puppet to label a few objects, and prompted the child to say whether the puppet was correct. Once the child had produced at least two *yes* and two *no* responses, the experimenter continued with the experiment.

In each trial, the child watched the animated video alongside the puppet. After the story, the puppet uttered the target sentence. The experimenter prompted the child to judge the sentence by asking, *Is Drog right?* For the two practice trials (which included only filler sentences), the experimenter provided feedback if the child responded incorrectly. The form of the feedback was flexible, but often involved pointing out relevant parts of the scene, repeating parts of the story, or modeling the correct response. After the practice trials, the experimenter did not provide feedback. In general, the experimenter reacted to the child's response by giving feedback to Drog: *Good job, Drog—you got it right!* or *Silly Drog, you got that one mixed up!*

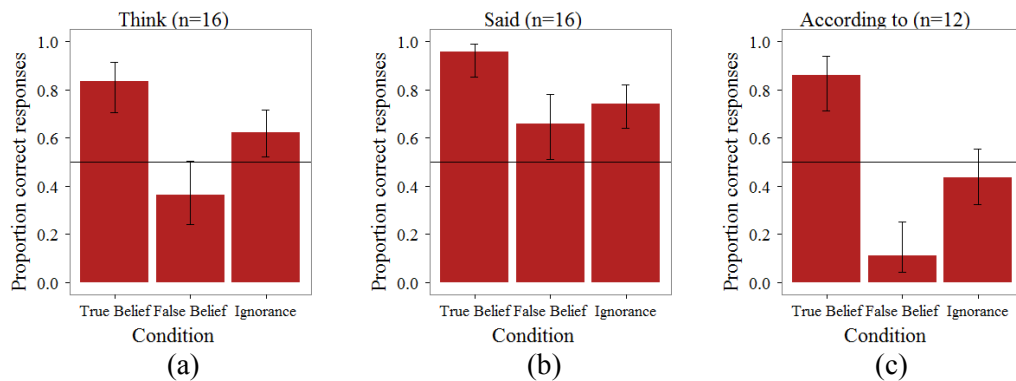
The filler sentence for each trial was chosen based on the child's response to the experimental sentence. If the child accepted the sentence, a false filler was chosen; if the child rejected it, a true filler was chosen.

## 4.6 Results

Children's responses were coded online by the experimenter and again from the video recording by a different person. Responses were coded as *yes*, *no*, *I don't know*, or *unclear*. Only clear *yes* or *no* responses that were never revised were counted in accuracy rates. Video coders rejected trials in cases of experimenter error (3 out of 1420 trials), or when the child was clearly not attending or distracted (28 trials). Since most of the 4-year-old participants had fragile attention spans, coders only rejected trials in extreme cases where the child was out of her chair or talking over the story.

### 4.6.1 Filler accuracy

The fillers were designed to be easy to judge so they could be used as a criterion to exclude participants who could not understand or attend to the task. 3 subjects who had accuracy rates below the predetermined cutoff of 65% were excluded from analysis. For the remaining 44 subjects, filler accuracy ranged from 67% to



**Figure 2** Results of Exp 1, by VERB: (a) think, (b) say and (c) according to. Error bars represent 95% confidence intervals based on the binomial distribution.

Knowledge	Belief Type	<i>Think</i>	<i>Say</i>	<i>According to</i>
<i>knowledge</i>	<i>true belief</i>	83%**	95%**	86%**
	<i>false belief</i>	36% <sup>°</sup>	66%*	11%**
<i>ignorance</i>	<i>true belief</i>	69%*	75%*	43%
	<i>false belief</i>	56%	73%*	44%

**Table 2** Accuracy rates by condition for Experiment 1. Stars indicate that the accuracy rate was different from chance: <sup>°</sup>  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.001$ .

100% (mean = 87%, median = 90%). After exclusions, there were 16 subjects in each of the *think* and *say* conditions, and 12 in the *according to* condition.

#### 4.6.2 Truth-value judgments

Accuracy rates for truth-value judgments were analyzed using logistic mixed effects models with fixed effects for BELIEF TYPE and KNOWLEDGE and a random effect for subjects. Binomial tests were used to compare accuracy to chance levels. See **Figure 2** and **Table 2** for a summary of results.

**Think** There was a significant main effect of BELIEF TYPE ( $p = 0.005$ ): children were more accurate with *true belief* than *false belief*. There was also a significant interaction between BELIEF TYPE and KNOWLEDGE ( $p << 0.001$ ): the asymmetry based on BELIEF TYPE only held in the *knowledge* condition (as expected, since in the *ignorance* condition the belief type is in fact unknown).

In *knowledge* stories, children were highly accurate in the *true belief* condition

(83%: above chance,  $p \ll 0.001$ ), and inaccurate in the *false belief* condition (36%: marginally below chance,  $p = 0.08$ ). In *ignorance* (collapsing across BELIEF TYPE), children were just above chance (62%,  $p = 0.02$ ). There was no significant difference in accuracy between *true belief* (69%) and *false belief* (56%) sentences.

**Say** As with *think*, there was a significant main effect of BELIEF TYPE ( $p = 0.01$ ) and a significant interaction between BELIEF TYPE and KNOWLEDGE ( $p = 0.007$ ). In *knowledge* stories, children were near ceiling in the *true belief* condition (95%: above chance,  $p \ll 0.001$ ). Their accuracy in the *false belief* condition was lower, but still above chance (66%,  $p < 0.05$ ). In the *ignorance* stories, children were above chance regardless of Belief Type (*true belief*: 75%, *false belief*: 73%,  $ps < 0.01$ ).

When the data from the *think* and *say* conditions were analyzed together with VERB as a fixed effect, there was a significant main effect of VERB ( $p = 0.02$ ) and no interactions with other factors ( $ps > 0.1$ ). Children's performance in the *say* condition was better across the board without changing the general pattern.

**According to** Again, there was a highly significant main effect of BELIEF TYPE ( $p \ll 0.001$ ) and a significant interaction between BELIEF TYPE and KNOWLEDGE ( $p \ll 0.001$ ). In *knowledge* stories, children were highly accurate in the *true belief* condition (86%: above chance,  $p \ll 0.001$ ), and near floor in the *false belief* condition (11%: below chance,  $p \ll 0.001$ ). In *ignorance* stories, children's accuracy was no different from chance (*true belief*: 43%, *false belief*: 44%,  $ps > 0.5$ ).

### 4.6.3 Discussion

Overall the results suggest that 4-year-olds are near adult-like in their understanding of 'say', but not 'think' or 'according to'.

With *think*, children's accuracy was lower in the *ignorance* conditions than in the *knowledge/true belief* condition. This finding is surprising under the Non-linguistic Development Hypothesis, since the *ignorance* condition reduces the complexity and representational conflict in the story. It is more consistent with the linguistic hypotheses, which predict difficulty in the *ignorance* condition where children are not able to evaluate the truth of the complement with respect to the actual world. Children may simply respond at random in this case. Although there were not enough such responses to analyze, children did answer "maybe" or "I don't know" more often in the *ignorance* condition than in other conditions.

Children's responses to 'according to' suggest that they only evaluated the complement clause. They are near ceiling when the truth of the sentence aligns with the complement (*true belief*), and near floor when they conflict (*false belief*). They seem to respond at random in the *ignorance* condition, where they have no

way of evaluating the complement. This result is a striking non-replication of Sowalsky et al.'s (2009) finding that children were more accurate with 'according to' than 'think'. In any case, the results provide a useful baseline for evaluating children's responses to 'think': although they clearly are not adult-like, at least some children are doing more than responding to the complement alone.

The Non-linguistic Development Hypothesis is not consistent with the difficulty in the *ignorance* condition. The pattern of responses with *think* sentences is consistent with a parenthetical interpretation, but it is not clear what sort of representation is giving rise to that interpretation.

## 5 Experiment 2

In Experiment 2 we manipulated the context in order to increase the relevance of beliefs. If children interpret 'think' parenthetically because they lack an adult-like syntactic/semantic representation for 'think', then no pragmatic manipulations should affect their responses. However, if parenthetical interpretations are due to a misunderstanding of the relevance of belief in context, it should be possible to improve performance by highlighting the beliefs.

The procedure was identical to Experiment 1, and thus is omitted here.

### 5.1 Subjects

17 4-year-old children (3;10-4;4, mean 4.1; 11 boys) participated in the study. All participants were typically-developing monolingual English-speakers.

### 5.2 Design and materials

The design was identical to Experiment 1, except that we only tested 'think' sentences. We made only one change to the stories: a second Seeker came out after the first Seeker and guessed the alternative location. The target sentences remained the same as those for Experiment 1; we counterbalanced whether the mentioned Seeker guessed first or second. Adding a second Seeker to the story creates a conflict of belief—only one of the two Seekers can have a true belief. This conflict was intended to highlight the relevance of beliefs in the story.

The design was identical to Experiment 1, except that we only tested 'think' sentences. We made only one change to the stories: a second Seeker came out after the first Seeker and guessed the alternative location. The target sentences remained the same as those for Experiment 1; we counterbalanced whether the mentioned Seeker guessed first or second. Adding a second Seeker to the story creates a conflict of belief—only one of the two Seekers can have a true belief. This conflict was intended to highlight the relevance of beliefs in the story.

### 5.3 Predictions

If children have an adult-like representation of ‘think’ available (the Pragmatic Development Hypothesis), then they should give more adult-like responses when the relevance of belief is highlighted. If children do not have an adult-like representation of ‘think’ (the Semantic Development Hypothesis), then this pragmatic manipulation should have no effect.

### 5.4 Results

See **Table 3** for a summary of the results. As in Experiment 1, there was a significant main effect of Belief Type ( $p = 0.006$ ) and an interaction with Knowledge ( $p = 0.003$ ). In *knowledge* stories, children were highly accurate in the *true belief* condition (89%: above chance,  $p \ll 0.001$ ), but no different from chance in the *false belief* condition (56%,  $p = 0.9$ ). In *ignorance* stories, children were significantly above chance (82% overall,  $p \ll 0.001$ ), with no significant difference between *true belief* (81%) and *false belief* (82%) sentences.

In a model with the number of seekers added as an additional fixed effect, there was a significant main effect of NUMBER OF SEEKERS ( $p = 0.009$ ), and no interactions with any other factors. That is, the overall pattern of responses was similar across the 1-seeker and 2-seeker stories, but children were more accurate overall with the 2-seeker stories (see **Figure 3**).

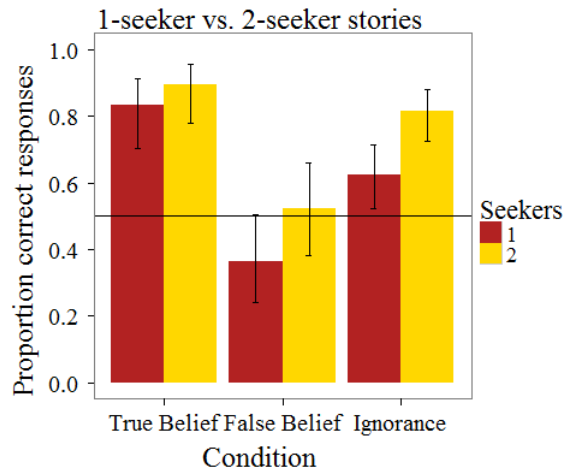
### 5.5 Discussion

The improved performance with 2-seeker stories suggests that 4-year-olds’ non-adult-like interpretations of ‘think’ are not due to a deviant syntactic or semantic representation. In the 1-seeker stories, children may take the “Question Under Discussion” (QUD) to be about the location of the Hider—a natural assumption in a story about Hide and Seek. Thus, they treat the target sentence as though its complement carries the main point of the utterance, and thus respond to the truth of the complement in reality, as demonstrated in (18). In 2-seeker stories, on the other hand, the Seeker’s beliefs are more obviously central, and children are more often able to take them into account when judging the target sentences.

- (18) A: (Where is Swiper?) (QUD)  
B: (Dora thinks that) Swiper is behind the toy box. (Target)  
C: No—he’s behind the curtain! (Child)

Knowledge	Belief Type	Accuracy
<i>knowledge</i>	<i>true belief</i>	89%**
	<i>false belief</i>	52%
<i>ignorance</i>	<i>true belief</i>	81%**
	<i>false belief</i>	82%**

**Table 3** Accuracy rates by condition for Experiment 2. Stars indicate that the accuracy rate was different from chance: \*\*  $p < 0.001$ .



**Figure 3** Accuracy rates by condition for Experiments 1 (1-seeker stories) and 2 (2-seeker stories). Error bars represent 95% confidence intervals based on the binomial distribution.

## 6 Conclusion

Our results indicate that 4-year olds’ difficulty with verbs like ‘think’ isn’t due to conceptual, syntactic or semantic immaturity: in contexts that emphasize the relevance of belief, they give more adult-like responses. We hypothesize that their difficulty is primarily pragmatic: children tend to underestimate the relevance of mental states to be what is being discussed, and therefore interpret ‘think’ parenthetically. Thus, even children who have adult-like conceptual and semantic representations can seem non-adult like in their comprehension of belief reports.

What does this tell us about children’s pragmatic competence? In some aspects, they are quite sophisticated: they can construct a parenthetical interpretation for an attitude report based on (sometimes erroneous) guesses about speakers’ intentions and the underlying QUD. What they might have trouble with

is accommodating a QUD different from the one they have initially hypothesized.

Preliminary results using similar methodology suggest that even younger children are able to give more adult-like responses in the right pragmatic contexts, providing further evidence that children have the relevant conceptual and semantic representations in place earlier than previously supposed.

## References

- Astington, Janet W. & Jennifer M. Jenkins. 1999. A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology* 35. 1311-1320.
- Baillargeon, Renée, Rose M. Scott & Zijing He. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14. 110-118.
- Bloom, Lois, Matthew Rispoli, Barbara Gartner & Jeremie Hafitz. 1989. Acquisition of complementation. *Journal of Child Language* 16. 101-120.
- Bolinger, Dwight. 1968. Post-posed main phrases: an English rule for the Romance subjunctive. *Canadian Journal of Linguistics* 14. 3-30.
- Bresnan, Joan. 1968. Remarks on adsententials. Manuscript, MIT.
- Clements, Wendy A. & Josef Perner. 1994. Implicit understanding of belief. *Cognitive Development* 9. 377-395.
- de Villiers, Jill G. 1995. Questioning minds and answering machines. In Dawn MacLaughlin & Susan McEwen (eds.), *Boston University Conference on Language Development (BUCLD) 19*, 20-36. Somerville, MA: Cascadilla Press.
- de Villiers, Jill G. 2005. Can language acquisition give children a point of view? In Janet W. Astington & Jodie A. Baird (eds.), *Why language matters for theory of mind*, 186-219. New York: Oxford University Press.
- de Villiers, Jill. 2007. The interface of language and theory of mind. *Lingua* 117. 1858-1878.
- de Villiers, Jill G. & Peter A. de Villiers. 2000. Linguistic determinism and the understanding of false beliefs. In Peter Mitchell & Kevin J. Riggs (eds.), *Children's Reasoning and the Mind*, 191-228. Hove, U.K.: Psychology Press.
- de Villiers, Jill G. & Peter A. de Villiers. 2009. Complements enable representation of the contents of false beliefs: the evolution of a theory of theory of mind. In Susan H. Foster-Cohen (ed.), *Language Acquisition*. New York: Palgrave Macmillan.
- de Villiers, Jill G. & Jennie E. Pyers. 2002. Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development* 17. 1037-1060.
- Diessel, Holger & Michael Tomasello. 2001. The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive*

- Linguistics* 12. 97-141.
- Hale, Courtney M. & Helen Tager-Flusberg. 2003. The influence of language on theory of mind: a training study. *Developmental Science* 6. 346-359.
- Hintikka, Jaakko. 1971. Semantics for propositional attitudes. In Leonard Linsky (ed.), *Reference and Modality*, 145-167. London: Oxford University Press.
- Hooper, Joan B. 1975. On assertive predicates. In John P. Kimball (ed.), *Syntax and Semantics*, 91-124. New York: Academic Press.
- Johnson, Carl N. & Michael P. Maratsos. 1977. Early comprehension of mental verbs: Think and know. *Child Development* 48. 1743-1747.
- Kovács, Ágnes M., Emő Téglaś & Ansgar D. Endress. 2010. The social sense: Susceptibility to others' beliefs in human infants and adults. *Science* 330. 1830-1834.
- Moore, Chris, Dana Bryant & David Furrow. 1989. Mental terms and the development of certainty. *Child Development* 60. 167-171.
- Onishi, Kristine H. & Renée Baillargeon. 2005. Do 15-month-old infants understand false belief? *Science* 308. 255-258.
- Rooryck, Johan. 2001a. Evidentiality, Part I. *Glott International* 5. 125-133.
- Rooryck, Johan. 2001b. Evidentiality, Part II. *Glott International* 5. 161-168.
- Ross, John R. 1973. Slifting. In Maurice Gross, Morris Halle & Marcel-Paul Schützenberger (eds.), *The formal analysis of natural languages. Proceedings of the First International Conference*. The Hague: Mouton.
- Shatz, Marilyn, Henry M. Wellman & Sharon Silber. 1983. The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition* 14. 301-321.
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117. 1034-1056.
- Southgate, V., A. Senju & G. Csibra. 2007. Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science* 18. 587-592.
- Sowalsky, Emily, Valentine Hacquard & Tom Roeper. 2009. Is PP opacity on the path to false belief? In Jean Crawford, Koichi Otaki & Masahiko Takahashi (eds.), *Generative Approaches to Language Acquisition North America (GALANA) 3*, 263-261. Somerville, MA: Cascadilla Proceedings Project.
- Surian, Luca, Stefania Caldi & Dan Sperber. 2007. Attribution of beliefs by 13-month-old infants. *Psychological Science* 18. 580-586.
- Urmson, J. O. 1952. Parenthetical verbs. *Mind* 61. 480-490.
- Wellman, Henry M., David Cross & Julianne Watson. 2001. Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development* 72. 655-684.
- Wimmer, Heinz & Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13. 103-128.

The semantics and pragmatics of belief reports in development

Shevaun Lewis  
1401 Marie Mount Hall  
College Park, MD 20742  
shevaun@umd.edu

Valentine Hacquard  
1401 Marie Mount Hall  
College Park, MD 20742  
hacquard@umd.edu

Jeffrey Lidz  
1401 Marie Mount Hall  
College Park, MD 20742  
jlidz@umd.edu