**"Think" pragmatically: Children's interpretation of belief reports**

Shevaun Lewis[a*], Valentine Hacquard[b], and Jeffrey Lidz[b]

[a]Maryland Language Science Center, University of Maryland

[b]Department of Linguistics, University of Maryland

*Corresponding author:

Shevaun Lewis

email: shevaun@umd.edu

1102 Francis Scott Key Hall

University of Maryland

College Park, MD 20742

**Abstract:** Children under 4 years of age often evaluate belief reports based on reality instead of beliefs. They tend to reject sentences like, "John thinks that giraffes have stripes" on the grounds that giraffes do not have stripes. Previous accounts have proposed that such judgments reflect immature Theory of Mind or immature syntactic/semantic representations. We argue that the difficulty is actually pragmatic. Adults frequently use belief reports to provide information about reality (e.g., "I think the stove is still hot"). Young children have difficulty determining when the main point is reality (the stove situation) vs. mental states (John's ideas about giraffes). We show that if the context emphasizes beliefs, children are more able to evaluate belief reports appropriately (Experiment 1). The pattern of children's truth value judgments demonstrates that they understand the literal meaning of *think* sentences, despite their pragmatic difficulty grasping the speaker's intention (Experiment 2).

# 1 Introduction

"Mental verbs" like *think*, *want*, and *hope* pose a considerable challenge for children learning their first language. They refer to aspects of the world that are not directly observable and rarely play a salient role in conversation. It is unsurprising, then, that children seem to acquire mental verbs relatively late compared to other verbs. For example, while over 75% of 24-month-olds say the verbs *eat*, *go*, and *kiss*, less than 10% of them say *think*, *pretend*, or *wish* (Jørgensen, Dale, Bleses, & Fenson, 2010). Although most children begin producing mental verbs before their third birthday (Bloom, Rispoli, Gartner, & Hafitz, 1989; Shatz, Wellman, & Silber, 1983), it is uncertain whether their production or comprehension of those verbs reflects adult-like knowledge until a year or more later (de Villiers J. G., 1995; Johnson & Maratsos, 1977).

Here we focus on the belief verb *think*.[1] Children seem to have non-adult-like interpretations of belief reports when the reported belief conflicts with reality, as in (1) (de Villiers J. G., 1995; de Villiers & de Villiers, 2000; de Villiers & Pyers, 2002; Johnson & Maratsos, 1977; Sowalsky, Hacquard, & Roeper, 2009). For example, 3-4 year olds would tend to say that (1) is false, even if John does think that dogs quack.

(1) John thinks that dogs quack.

Children's non-adult-like interpretations of *think* in false belief scenarios align with their notoriously poor performance in traditional "false belief tasks". In the change-of-location false

[1] Although *think* can express non-belief meanings (as in, "John often thinks about global warming"), it was a more appropriate object of study than *believe*, since it is more frequent (especially in child directed speech) and acquired earlier in development. We would expect to observe similar patterns with *believe*, but in older children.

belief task, children are asked how a character with a false belief will behave. In a representative story from Wimmer & Perner (1983), a boy named Maxi is helping his mom put away groceries. He puts some chocolate in the blue cupboard before going out to the playground. While he is gone, Maxi's mom uses the chocolate to make a cake, and puts the leftovers into the green cupboard. Then Maxi returns to eat some chocolate. Children are asked the test question in (2).

(2) Where will Maxi look for the chocolate?

In Wimmer and Perner's study and the dozens of similar studies that followed, 3-year-olds' performance on this question is consistently poor. They appear to be biased by their own knowledge: in response to a question like (2), they would often say that Maxi would look for the chocolate in its actual location (the green cupboard) rather than where he put it (the blue cupboard). In a meta-analysis of 178 studies, Wellman, Cross and Watson (2001) found that children respond based on reality more often than not until about 3;5 (3 years, 5 months). They start responding based on beliefs at about 4;0 on average. Thus, during a certain period of development, the "curse of knowledge" seems to affect children's understanding of false beliefs (Birch & Bloom, 2003), whether the task involves linguistic belief reports or just predictions about a character's behavior.

However, previous studies did not provide a fine-grained assessment of the truth conditions children attribute to *think* sentences, which would allow us to infer exactly how their interpretation of belief reports differs from adults'. We compare three potential sources of non-adult-like behavior: (1) belief concepts, (2) the semantics of belief reports, and (3) the pragmatics of conversational use of belief reports.

### 1.1 The conceptual hypothesis

Under the conceptual hypothesis, children's non-adult-like interpretations of belief reports are directly related to their poor performance on traditional false belief tasks. The explanation of both is that children lack an understanding of false belief at the conceptual level. The change that occurs around 4 years of age, causing improvements on both tasks, is that children become able to attribute false beliefs to others.

This hypothesis encounters two main empirical challenges. The first is that although performance on linguistic and non-linguistic tests of belief understanding are closely correlated in development, children may succeed earlier on the linguistic tasks. In a longitudinal study with 3-4 year-old children, de Villiers and Pyers (2002) found that there were strong correlations at each time point between false belief task performance and language measures, particularly one involving memory for false complements. Children were told a short story with pictures, and then asked a belief question, as in (3).

(3) 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?

Children were much more likely to pass the false complement tasks earlier than the false belief tasks, rather than vice versa. Training studies have provided more direct evidence that understanding of sentential complements, especially false complements, is causally related to success on false belief tasks (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). A strong relationship between language and performance on false belief tasks has been found across several different populations, including children with autism (Tager-Flusberg & Joseph, 2005), children with specific language impairment (de Villiers, Burns, & Pearson, 2003), and deaf children and adults with either delayed or normal language development (de Villiers P. A.,

2005). In a meta-analysis of 107 studies, Milligan and colleagues found large effects for several different kinds of language measure on false belief performance, and early language was more likely to predict later false belief performance than vice versa (Milligan, Astington, & Dack, 2007). Together, this evidence seriously undermines the idea that the development of adult-like comprehension of belief reports is held back by the relatively delayed development of adult-like concepts of belief.

The second main empirical challenge for the conceptual hypothesis is recent evidence that children can reason about false beliefs long before their fourth birthday. Many researchers have argued that traditional false belief tasks fail to find evidence for competence in young children because they require an explicit decision and response. Young children succeed earlier in false belief tasks if the critical question or task is less direct (Buttelmann, Carpenter, & Tomasello, 2009; Knudsen & Liszkowski, 2012; Rubio-Fernández & Geurts, 2013; Scott, He, Baillargeon, & Cummins, 2012; Southgate, Chevallier, & Csibra, 2010). Infants as young as 13 months show understanding of the behavioral consequences of false beliefs when they are tested using more implicit measures in preferential looking or eye-tracking studies (Baillargeon, Scott, & He, 2010; Clements & Perner, 1994; He, Bolz, & Baillargeon, 2012; Onishi & Baillargeon, 2005; Song & Bailargeon, 2008; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007). 7-month-olds are able to at least track the beliefs of others (Kovács, Téglás, & Endress, 2010). These findings align with evidence that even 2-year-olds engage in numerous behaviors that would seem to require them to track people's belief states. For example, they attempt to help people that they know to be ignorant of the location of an object (O'Neill, 1996). They also attempt to deceive, which requires an understanding that it is possible for others to have false beliefs (Chandler, Fritz, & Hala, 1989).

This evidence suggests that the development occurring in the preschool years is not the emergence of a "new" concept of belief. More likely, it involves some more subtle aspect of children's representational or processing abilities, such that they can access their knowledge in some tasks but not in others. One possibility is that attributing false beliefs 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 beliefs. 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.

To summarize, the conceptual hypothesis holds that children are swayed by reality when interpreting belief reports for the same reason that they perform poorly on false belief tasks: they have difficulty representing, tracking, or otherwise reasoning about false beliefs. We have rejected the strongest version of this hypothesis, that young children simply lack a concept of belief. However, the simple fact remains that children struggle with explicit false belief tasks until nearly 4 years of age. Whatever factors explain that delay could also be responsible for the delay in adult-like comprehension of belief reports.

### 1.2 The syntax/semantics hypothesis

Under the various views that we will group under the "syntax/semantics hypothesis", children's non-adult-like responses to belief reports in comprehension tasks are generated by non-adult-like representations of the structure and lexical semantics of belief reports.

One possibility is that although children sometimes produce mental verbs (including *think*) alongside finite clauses, they do not actually represent those clauses as complements of the verb. Instead, the mental verb is stuck onto the clause like an adverb or parenthetical. This view gains support from the fact that children's early productions of *think* and other mental verbs tend to be restricted: *think* occurs overwhelmingly in the present tense with the first person singular subject 'I' (Bloom, Rispoli, Gartner, & Hafitz, 1989; Diessel & Tomasello, 2001). These early uses generally function to direct the conversation or express uncertainty, as in (4)-(8), rather than to comment on someone's belief state.

(4) I thoughted we'd eat some cake. (Bloom et al., 1989)

(5) I think this is a lamb.

(6) I think I'm go in here. [3;1] (Diessel & Tomasello, 2001)

(7) Think some toys over here too. [3;2]

(8) It's a crazy bone I think. [3;5]

Only later do children begin to use *think* with a wider variety of subjects, tenses, and aspects, and with a meaning that unambiguously involves mental states (Diessel & Tomasello, 2001).

Another possibility, proposed by Jill de Villiers, is that the crucial development between the ages of 3 and 4 is the semantic structures necessary to represent attitude reports with false complements (de Villiers J. G., 2005; 2007; de Villiers & de Villiers, 2000; 2009; de Villiers & Pyers, 2002).[2] De Villiers suggests that children's mastery of false complements for

[2] J. de Villiers initially hypothesized that the main difficulty was the syntax of the tensed sentential complements that *think* selects (compared to the tenseless complements that *want* selects). However, cross-linguistic evidence has

communication verbs—which correspond to observable events in the world—allows them to bootstrap their way into an understanding of mental verbs like *think*. These semantic structures for representing mental attitudes are then crucially involved in successful performance on traditional false belief tasks.

Some versions of the syntax/semantics hypothesis (including de Villiers') go so far as to claim that the development of linguistic representations of belief is a prerequisite for the corresponding conceptual representations. This view is of course motivated by the findings (described in the previous section) that belief report understanding seems to precede and perhaps cause success on traditional false belief tasks. This stronger claim, like the strong version of the conceptual hypothesis, is challenged by the recent evidence (discussed above) that very young children and infants have some ability to track and reason about false beliefs. If it were true, as de Villiers argues, that a semantic representation of *think* sentences is a necessary precursor to belief attribution, pre-linguistic infants should not show any understanding of false beliefs. However, linguistic representations could affect the non-linguistic attribution of mental states in

---

demonstrated that belief reports are more difficult than desire reports even if they have the same kind of complements (German: Perner, Sprung, Zauner, & Haider, 2003; Mandarin: Tardif & Wellman, 2000). While the conceptual hypothesis has no problem accounting for these cross-linguistic consistencies, since the relevant conceptual development would presumably unfold in the same way cross-culturally (Callaghan, et al., 2005; Liu, Wellman, Tardif, & Sabbagh, 2008), they are problematic for any version of the syntax/semantics hypothesis that emphasizes tensed complements or any other single syntactic property (but see Hacquard (2014) for a proposal on how the syntactic properties of complements to *think* and *want* converge cross-linguistically at a more abstract level, beyond tensedness).

some way other than by providing access to the concept, as we will explain in the general discussion.

A more fundamental problem with the semantic hypothesis is that, as more basic structural properties like finite clausal complements are ruled out as the key barrier to children's acquisition of false complements, the account becomes less explanatory. For example, it is difficult to distinguish a "point of view" feature on a complement clause (de Villiers & de Villiers, 2009) from the conceptual representations that it is supposed to enable.

To summarize, according to the views encompassed by the syntax/semantics hypothesis, children are swayed by reality when evaluating belief reports because they don't know that *think* maps onto the mental state of belief, or because they are incapable of representing a clausal complement of *think* as false. On this view, they actually do evaluate the truth of the complement clause with respect to reality.

### 1.3 The pragmatic hypothesis

In this paper we provide evidence for a third possibility, that children's difficulty with belief reports is pragmatic. Specifically, 3-year-olds' non-adult-like interpretations reflect the computation of inappropriate *speaker meanings*, rather than incorrect *literal meanings* as proposed by the syntax/semantics hypothesis. Despite understanding the truth conditions of a belief report, young children draw incorrect conclusions about the message that a speaker is trying to convey with it.

Belief reports can be used to express different speaker meanings in different contexts (Hooper, 1975; Simons, 2007; Urmson, 1952). Consider the different implications of the same belief report in (9)-(10).

(9) A: Why didn't Mary invite John to the meeting?

B: She thinks he's working from home.

(10) A: Where is John? It's time to start the meeting.

B: Mary thinks he's working from home.

The exchange in (9) is about explaining Mary's behavior, so her mental states are highly relevant. B's utterance is therefore intended as a straightforward report of Mary's belief, which may or may not be true in the actual world. In this context, B does not have any commitment to the truth of the complement clause, the content of Mary's belief. The exchange in (10) is about John's whereabouts, so Mary's mental states are less relevant. The main contribution of B's utterance is the proposition expressed by the complement clause (that John is working from home). Given that conversational participants are assumed to contribute only what they know to be true (Grice's maxim of Quality), B implicitly endorses the truth of the complement clause in the actual world. However, if B were entirely certain about John's whereabouts, he would have simply said so directly. The inclusion of a source of evidence implies that B only endorses the proposition that John is working from home to the degree that Mary can be trusted. This hint of uncertainty is amplified in first person belief reports like (11).[3]

[3] There has been much discussion of the syntactic properties of so-called "parenthetical' uses of attitude verbs, as in "John's working from home, I think" (Bolinger, 1968; Bresnan, 1968; Hooper, 1975; Rooryck, 2001; Ross, 1973; Urmson, 1952). The attitude verb may occupy a different structural position in these cases—perhaps a sentence adverbial (Bresnan, 1968) or the head of a functional projection for evidential markers (Rooryck, 2001). However, the comparison of (9)-(10) demonstrates that sentences with standard word order can receive either mental state or endorsement interpretations depending on the context. Pragmatic reasoning is therefore required to determine the speaker's intended meaning, regardless of whether there is also syntactic ambiguity (Simons, 2007).

(11) I think he's working from home.

Studies of child directed speech have found that adults' uses of belief reports overwhelmingly occur in the first person and have this endorsement meaning (Diessel & Tomasello, 2001; Howard, Mayeux, & Naigles, 2008; Naigles, 2000), as in (12)-(13).

(12) I think you should try and finish one game. (Howard et al., 2008)

(13) I don't think you'll be working on the roof.

Given the preponderance of endorsement uses in adult speech, it is unsurprising that children's early uses of *think* have the same flavor, as we discussed in the previous section. The pragmatic hypothesis is that children tend to assume endorsement uses of belief reports, even in situations where beliefs are actually relevant. We suggest that although children are capable of computing the literal meaning of belief reports, they often misjudge the discourse context and fail to recognize when beliefs are relevant to the conversation. That is, they fail to infer the correct "Question Under Discussion" (Roberts, 2004; 2012).

To make the pragmatic hypothesis more concrete, let's walk through an example. Suppose we walk into a room and observe my dog frantically sniffing and pawing at an empty treat box. I might utter the belief report in (14).

(14) Jasper thinks that there are still treats in there.

As an adult, you understand that the implicit Question Under Discussion I just introduced is, "Why is Jasper so interested in that empty treat box?" We both already know that there are actually no treats in the box, so it would be irrelevant to comment on that fact.

Now suppose that I'm having this conversation with a 3-year-old child instead of an adult. When we walk into the room and I utter (14), the child has to figure out why I would have said such a thing. One possibility is that I'm commenting on Jasper's erroneous belief state, but

this possibility may not be particularly salient. An alternative possibility is that I don't know whether there are any treats in the box, and I'm citing Jasper's behavior as a source of evidence that there are. That is, the child assumes that I've just introduced a different Question Under Discussion: "Are there any treats in the box?" Under this assumption, the child might deny my statement. This is the behavior that has been observed in previous studies: denying a belief report based on reality.

Thus, the syntax/semantics hypothesis and the pragmatic hypothesis have different interpretations of the same results, which are "parsimonious" in different ways. Proponents of the syntax/semantics hypothesis attribute as little linguistic knowledge as possible to the child, arguing that their early production and comprehension can be explained without recourse to the full adult structure. However, they still have to explain how children eventually acquire the adult-like structure. We take the opposite approach with the pragmatic hypothesis, arguing that young children's restricted productions and non-adult-like responses are actually consistent with adult knowledge. The explanation of children's early errors is more elaborate, but makes it easier to explain how children's performance becomes more adult-like over time.

The pragmatic hypothesis we are proposing is that children often fail to recognize the relevance of belief in context, and therefore assume that the speaker meaning has to do with reality. Why do children so often fail to recognize that beliefs are relevant? Previous work suggests that it is because they do not have the resources to track people's beliefs as accurately or efficiently as adults do. However, an equally explanatory account is that children may not understand when people are likely to be talking about beliefs, even if they do accurately track them under some circumstances. It is by no means trivial to determine that a given conversation is about mental states rather than what is happening in the physical world: even adults—who

presumably have fully developed Theory of Mind skills—need fairly strong cues to figure it out (Gillette, Gleitman, Gleitman, & Lederer, 1999; Papafragou, Cassidy, & Gleitman, 2007). Successful pragmatic reasoning relies on mechanisms that continually update a structured representation of the discourse. Those mechanisms—like comprehension mechanisms more generally—are most likely tuned over time by experience in conversation.

The pragmatic hypothesis makes two clear predictions. First, we should be able to induce children to judge belief reports based on beliefs rather than reality by drawing their attention to the relevance of belief in a particular context. We test this prediction in Experiment 1. Second, if children have access to the literal meaning of belief reports, they should be able to reject belief reports that are literally false, regardless of whether they compute an inappropriate speaker meaning. We test this prediction in Experiment 2.

## 2 Experiment 1: Context sensitivity[4]

The goal of Experiment 1 was to test the prediction that if the relevance of belief is salient in the context, children will show a more adult-like pattern of judgments for belief reports. We presented children with stories about hide-and-seek and asked them to judge belief reports about the seekers' beliefs. In one condition, we enhanced the relevance of belief by introducing a conflict of belief between two "seekers". If children are influenced by this contextual manipulation, they should show more adult-like responses in the critical false belief conditions. Neither the conceptual nor the syntax/semantics hypothesis predicts that children should be influenced by context, so they would not predict any improvement.

[4] These results were reported previously in Lewis, Hacquard, & Lidz (2012).

To determine how much of children's difficulty could be attributed to difficulty inhibiting their own knowledge, we included a condition in which the child was ignorant of the location of the hider. If children's difficulty in the false belief condition is related to inhibition, they should perform better in this "ignorance" condition.

We tested children within a few months of their fourth birthday. At this age, children are beginning to show above-chance performance on traditional false belief tasks (Wellman, Cross, & Watson, 2001), but still fall short of adult-like understanding of belief reports (de Villiers & Pyers, 2002; Sowalsky, Hacquard, & Roeper, 2009).

## 2.1 Methods

### 2.1.1 Participants

Participants for Experiments 1 and 2 were recruited from the Center for Young Children preschool or the Infant and Child Studies Database at the University of Maryland. All were typically developing, monolingual English speakers. 36 children aged 3 years, 10 months (3;10) to 4 years, 5 months (4;5) participated in Experiment 1(mean = 3.9 years, 17 girls). After 4 participants were excluded for low accuracy on fillers (see section 2.2.1), there were 32 participants included in the analysis.

### 2.1.2 Design

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

#### *2.1.2.1 Sample story*

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

In the first scene, the characters (Swiper and Dora) are named and the experimenter confirms that the child can identify them (Figure 1a). 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.*

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

Dora reappears: *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 belief (Figure 1f).

[Figure 1 here]

At this point, the experimenter asks the puppet to say something about what's going on in the story. The puppet delivers a target sentence like (15). After the child responds, the puppet delivers a filler sentence (see below). Once the child has responded to both the target sentence and the filler, the Hider and Distractor emerge from their hiding places.

(15) Dora thinks that Swiper is behind the toy box.

*2.1.2.2 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 Distractor hid in the

scene (as in the sample story). In the *ignorance* condition, the screen was obscured during the hiding phase. After the characters were introduced, the screen went gray and the experimenter explained: *This time it's a secret where Swiper is hiding, so we can't watch either.* In the next scene, the child saw the two "clues" (e.g. the yellow tails behind the toy box and the curtain), but did not know which clue corresponded to the Hider.

We also manipulated the BELIEF TYPE: whether the target sentence referred to a Seeker with a *true belief* or a *false belief*. In the sample story, the target sentences are about a seeker with a *false belief*. Note that in the *ignorance* condition, it is unknown at the point of the target sentence whether the Seeker has a true or false belief. The truth of the target sentences (i.e., the target response) was counterbalanced.[5] Table 1 shows the set of possible *think* target sentences for the sample story.

[Table 1 here]

The most important manipulation for our pragmatic hypothesis was the NUMBER OF SEEKERS. In the *2-seeker* stories, a second seeker guessed the other location marked with a clue (Figure 1g), so that one seeker had a true belief and the other a false belief. The *2-seeker* stories were intended to heighten the relevance of belief in context by introducing an important conflict of belief. The NUMBER OF SEEKERS was a between-subjects factor, so each child was randomly assigned to see either *1-seeker* stories or *2-seeker* stories. Note that even in the *2-seeker* stories, children heard only one target sentence, reporting the belief of one of the seekers.

[5] To confirm our judgments of the target sentences, we recruited 18 adults to complete the study as well (10 in the 1-seeker condition, 8 in the 2-seeker condition). Out of 216 responses, there was only 1 that differed from the target response.

### 2.1.3 Materials

We created 14 stories including a variety of scenes and characters. The identities and locations of the Hider, Seeker, and Distractor were rotated from story to story. The stories were illustrated, animated, and synced with recorded narration using Adobe Flash.

There were two lists of target sentences. 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 (16)-(20). (17)-(18) were only appropriate in *knowledge* stories, and (20) only in *ignorance*.

(16) Dora is looking for Swiper {behind the toy box/behind the curtain}.

(17) Swiper is really hiding {behind the curtain/behind the toy box}.

(18) There's really a squirrel {behind the toy box/behind the curtain}.

(19) We can see a yellow tail {behind the toy box/under the bed}.

(20) 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 × BELIEF TYPE). Each child heard an equal number of critical items with 'yes' and 'no' target responses (i.e., true and false sentences). 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. For example, a child assigned to List A would hear 2 true and 1 false sentence in the *knowledge/false belief* condition, while a child assigned to List B would hear 1 true and 2 false sentences in that condition. Since the number of true and false sentences was not balanced within each condition for each

participant, we did not treat sentence truth as a factor in our analysis. (This was remedied in Experiment 2.)

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. Thus, for most children, the number of actual 'yes' and 'no' responses was roughly balanced across the experiment.

### 2.1.4 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?* This prompt targets the appropriateness of the contribution of the utterance as well as its truth value. Thus, although we refer to this task as "truth value judgment", the child is not constrained to respond based on truth alone. A *yes* response means that the sentence seems like a true and appropriate contribution. A *no* response could indicate that the

sentence is false, but it could also mean that the sentence is underinformative, irrelevant, or in some other way an inadequate contribution, given the goals of the conversation.

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!*

### 2.1.5 Data analysis

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.

Accuracy rates for truth-value judgments were first analyzed separately for children in the *1-seeker* and *2-seeker* conditions, using logistic mixed effects models with fixed effects for BELIEF TYPE and KNOWLEDGE and the maximal by-subject random effects structure (random by-subject intercepts and random by-subject slopes for both main effects and the interaction). Binomial tests were used to compare accuracy (grand totals, not grouping by subjects or items) to chance levels. Accuracy in the *1-seeker* and *2-seeker* conditions was compared using a model

with a fixed effect for NUMBER OF SEEKERS as well as BELIEF TYPE and KNOWLEDGE, and the same by-subject random effects structure as the previous models.

## 2.2 Results

### 2.2.1 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. 4 participants who had accuracy rates below the predetermined cutoff of 65% were excluded from analysis. For the remaining 32 participants, filler accuracy ranged from 67% to 100% (mean = 87%, median = 90%). After exclusions, there were 16 participants each in the *1-seeker* and *2-seeker* conditions.

### 2.2.2 Accuracy on *think* sentences

See Table 2 and Figure 2 for a summary of results.

In the *1-seeker* condition, there was a significant main effect of BELIEF TYPE ($\beta = 0.499$, Wald's $z = 2.45$, $p = .014$): children were more accurate with *true belief* than *false belief*. There was also a significant interaction between BELIEF TYPE and KNOWLEDGE ($\beta = 0.698$, Wald's $z = 2.73$, $p = .0064$): 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, $p = 3.3\times10^{-6}$), and inaccurate in the *false belief* condition (36%: marginally below chance, $p = .079$). In *ignorance*, children were just above chance overall (62%, $p = .022$), with no significant difference in accuracy between *true belief* (69%) and *false belief* (56%)—as expected, since the two types of story are identical in the *ignorance* condition until after the child has responded.

In the *2-seeker* condition, there was again a significant interaction between BELIEF TYPE and KNOWLEDGE ($\beta = 0.826$, Wald's $z = 2.20$, $p = .028$). In *knowledge* stories, children were

highly accurate in the *true belief* condition (90%: above chance, $p = 1.4\times10^{-8}$), but no different from chance in the *false belief* condition (52%, $p = .88$). In *ignorance* stories, children were significantly above chance (82% overall, $p = 7.3\times10^{-10}$), with no significant difference between *true belief* (81%) and *false belief* (82%) sentences.

In the model with NUMBER OF SEEKERS as an additional fixed effect, there was a significant main effect of BELIEF TYPE ($\beta = 0.459$, Wald's $z = 2.61$, $p = .0091$), and an interaction between BELIEF TYPE and KNOWLEDGE ($\beta = 0.621$, Wald's $z = 3.55$, $p = .00039$). There was also a significant main effect of NUMBER OF SEEKERS ($\beta = 0.614$, Wald's $z = 2.66$, $p = .0079$), but no interactions between NUMBER OF SEEKERS and any other factor. Thus, 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.

[Table 2 and Figure 2 here]

### 2.3 Discussion

As predicted by the pragmatic hypothesis, children's accuracy improved across conditions when the story involved two seekers instead of one. We conclude that the heightened relevance of belief in 2-seeker stories helped children access a belief-based rather than a reality-based speaker meaning for the belief report.

Children were highly accurate in the *true belief* condition, inaccurate in the *false belief* condition, and somewhere between in the *ignorance* condition. Although all of the hypotheses predicted the difference in performance between the *true belief* and *false belief* conditions—that is what they were designed to do—the middling performance in the *ignorance* condition is potentially informative. The intention of the *ignorance* condition was to eliminate the conflict between the characters' belief and the child's belief. If the child is as ignorant of the truth as the

character, their judgments about the content of the character's beliefs which should not be affected by the curse of knowledge. However, children showed lower accuracy in this condition than in the *true belief* condition. What made it more difficult? If children were attempting to evaluate the complement clause against reality—as part of a deviant literal meaning for the sentence or an inappropriate pragmatic reading—the *ignorance* condition would be confusing, because it makes that evaluation impossible. Thus, the mild difficulty in the *ignorance* condition might provide some support for either the syntax/semantics or pragmatic hypothesis over the conceptual hypothesis.

The results of Experiment 1 demonstrate that children are sensitive to context when interpreting belief reports, and can provide more adult-like responses in some situations. In Experiment 2, we investigate whether children can in fact compute the correct literal meaning for belief reports.

## 3 Experiment 2: Truth conditions for *think*

The goal of Experiment 2 was to determine whether children are capable of rejecting a belief report based their knowledge of its literal meaning. We used only the 2-seeker stories (to give children the best chance of accessing the literal meaning), and manipulated the literal truth of the belief report as a factor, rather than merely counterbalancing it as in Experiment 1 and other previous studies. If children can evaluate belief reports based on their literal meaning, they should always reject sentences that are literally false. But they may evaluate literally true sentences based on the truth of the complement clause (depending on the speaker meanings they attribute). Thus, children's accuracy should be higher for literally false than literally true sentences.

To illustrate this prediction, let's consider a sample scenario (illustrated in Figure 3), in which Dora has incorrectly guessed that Swiper is behind the toy box—i.e., Dora has a false belief. If children assume that the whole story is about Swiper's whereabouts, they might reject a true sentence about Dora's belief, as in (21), because, while technically true, it does not provide good information about Swiper. However, they should never accept a false sentence about Dora's belief, like (22), even though the complement clause does correctly describe Swiper's location.

[Figure 3 here]

(21) Dora thinks that Swiper is behind the toy box.

(22) Dora thinks that Swiper is behind the curtain.

By contrast, the conceptual and syntax/semantics hypotheses do not predict that children should be sensitive to the truth of the whole sentence. Under the conceptual hypothesis, children's difficulty representing false beliefs should cause poor performance in both false belief conditions. If they are unable to represent Dora's belief separately from their own knowledge of Swiper's true location, they should reject a true sentence like (21), and accept a false sentence like (22). Under the syntax/semantics hypothesis, children respond based on the truth of the complement clause, regardless of whether the belief report expressed by the whole sentence is true or false. Once again, this would lead to poor performance in both false belief conditions, since the complement clause is false when the whole sentence is true (21), and true when the whole sentence is false (22).

To investigate whether success on this task depends on independent conceptual development, we widened the age range to include younger 3-year-olds and added a traditional false belief task.

## 3.1 Methods

### 3.1.1 Participants

66 children aged 3;1 to 4;2 participated in the study (mean = 3.6 years, 32 girls). Data from an additional 8 children (3.1-4.0 years, mean = 3.5, 3 girls) were excluded because they could not complete the task. After 14 participants were excluded for low accuracy on fillers (see section 3.2.1), there were 52 participants included in the analysis.

### 3.1.2 Design and materials

#### *3.1.2.1 Truth-value judgment task*

As in Experiment 1, children were presented with stories about hide-and-seek, and asked to judge target sentences containing *think*. All of the stories contained 2 seekers.

To determine whether children are able to evaluate belief reports against the character's beliefs instead of reality, we manipulated the (literal) SENTENCE TRUTH as a factor, rather than merely counterbalancing it as in Experiment 1.[6] We collapsed BELIEF TYPE and KNOWLEDGE into a single BELIEF TYPE factor with 3 levels: *true belief*, *false belief*, and *unknown*.

We used the same 14 stories from Experiment 1. Rather than presenting the stories in animated videos, we illustrated each story with a series of 8-9 still images.[7]

---

[6] In Experiment 1, since there were 3 trials per condition, sentence truth could not be counterbalanced within participants, but only across participants who heard each list.

[7] We found that this method was more successful for younger children, perhaps because it allowed them to take all their cues from one source (the experimenter), rather than switching their attention between the narrated videos, the puppet, and an experimenter.

We created two lists of target sentences. 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, using the same templates as in Experiment 1. Each participant saw 2 practice trials, followed by 2 trials in each of 6 conditions (SENTENCE TRUTH × BELIEF TYPE).

*3.1.2.2 False belief task*

In addition to the truth-value judgment task, most children also completed two trials of a standard change-of-location false belief task. The story was acted out by the experimenter using toys. The story for one trial was as follows:

> *This story is about Toby and his dad. Toby is playing with his cowboy hat and pretending to be a cowboy. He's having a great time. Then he decides to go outside and play, but he doesn't want his cowboy hat to get dirty. So he puts it under the bucket where he can find it later. [Toby leaves the scene.] While Toby is playing outside, his dad comes in to clean up his room. He finds the cowboy hat under the bucket. He says, "Hey, this doesn't belong here! I'm going to put it in the toy box where it's supposed to be. …There. Much better." [Toby's dad leaves the scene.]*

After the story, children were asked a series of questions:

(23) *Pre-test memory questions:*

a. Where did Toby put the cowboy hat (before he went outside)?

b. Where is the cowboy hat now?

(24) *Test question:* Toby is coming back inside, and he wants to play with his cowboy hat again. He remembers where he put it. Where is Toby going to look for the cowboy hat first?

(25) *Justification:* Why will he look there?

(26) Post-test memory question [depending on child's answer to Test question]:

a. Correct test: Where is the hat really?

b. Incorrect test: Where did Toby put the hat before he went outside?

### 3.1.3 Procedure

Sessions took place in a quiet room. For the truth-value judgment task, the child was seated in front of an iPad operated by an experimenter sitting alongside. Sessions were videotaped so that children's responses could be coded later by an independent viewer.

The experimenter began by explaining the task, introducing a little boy who appeared on the screen next to the story: This little boy is only 2 years old, so he doesn't know how to play hide and seek. We're going to try to help him learn to play. After each story, he's going to try to say what happened in the story, and it will be your job to tell him if he's right or wrong. After obtaining the child's assent to participate, the experimenter continued with the experiment.

In each trial, the experimenter narrated the story, swiping the screen to display each scene, then delivered the target sentence (in the voice of the little boy). The experimenter prompted the child to judge the sentence by asking, *Did the little boy get it right?* As in Experiment 1, this prompt elicits pragmatic as well as semantic judgments.

As in Experiment 1, the experimenter provided feedback for the two practice trials, but not the experimental trials. As in Experiment 1, a true or false filler sentence was chosen based on the child's response to the experimental sentence for that trial. The experimenter recorded the child's responses using buttons on the iPad.

The false belief task always came after the truth-value judgment task. If the child initially provided incorrect answers for the pre-test memory questions, the experimenter retold the story

until the child responded correctly. No feedback was provided for the test question or post-test memory questions.

### 3.1.4 Data analysis

Responses for the truth-value judgment task were coded as *yes*, *no*, *I don't know*, or *unclear*. Unclear responses were excluded from analysis. Video coders excluded responses in cases of experimenter error (2 out of 1396 responses), experimenter cuing of the child's response (28 responses) or when the child was clearly distracted or explicitly committed to a guess about the location of the hider in the *unknown* condition (6 responses). As in Experiment 1, coders were conservative, only rejecting trials in cases of extreme and obvious inattention.

For the false belief task, all children provided correct answers on the pre-test memory questions on the first or second try. Children received a score of 1 for the trial if they provided correct answers for both the test question and the follow-up memory question. Each child was given a total FB score of 0, 1, or 2 for the number of correct trials.

Since the goal of this experiment was to directly assess children's truth conditions for *think*, we analyzed acceptance rates as well as accuracy. The acceptance rate is the proportion *yes* responses out of *yes* or *no* responses, excluding *I don't know* responses.

Accuracy and acceptance rates for truth-value judgments were analyzed using logistic mixed effects models with fixed effects for BELIEF TYPE, SENTENCE TRUTH, and the subject's age, as well as a random by-subject intercept and random by-subject slopes for BELIEF TYPE, SENTENCE TRUTH, and their interaction. All factors were coded orthogonally. The 3-level factor BELIEF TYPE was coded as two contrast variables: the first compared *true belief* to *false belief*; the second compared *false belief* to *unknown*.

To determine whether children's performance on the standard false belief task was predictive of their understanding of *think* sentences, we used two different models. The first added children's FB Score as a fixed effect to the original model in place of the age effect. The second was a model of accuracy in the truth-value judgment task of the *false belief* trials alone. The model had fixed effects for SENTENCE TRUTH and FB Score, as well as a random by-subject intercept and a random by-subject slope for SENTENCE TRUTH.

## 3.2 Results

### 3.2.1 Filler accuracy

14 participants who had accuracy rates below the predetermined cutoff of 65% were excluded from analysis. These participants were distributed over the full age range (3.2-4.0 years, mean = 3.6, 6 girls). For the remaining 52 participants (3.1-4.2 years, mean = 3.6, 26 girls), filler accuracy ranged from 67% to 100% (mean = 86%).

### 3.2.2 Truth-value judgment task

#### *3.2.2.1 Accuracy*

See Table 3 and Figure 4 for summaries of the results.

There was a significant main effect of SENTENCE TRUTH: accuracy was higher overall when the sentence was *false* compared to when it was *true* ($\beta$ = -0.225, Wald's $z$ = -2.02, $p$ = .044). There was a marginal main effect for one of the BELIEF TYPE contrasts: accuracy was higher in the *true belief* compared to the *false belief* condition ($\beta$ = 1.15, Wald's $z$ = 1.68, $p$ = .094), but the difference between the *false belief* and *unknown* conditions was not significant ($\beta$ = -0.557, Wald's $z$ = -1.43, $p$ = .15). There was a significant interaction between the first BELIEF TYPE contrast and SENTENCE TRUTH ($\beta$ = 1.38, Wald's $z$ = 2.05, $p$ = .040): in the *true belief* condition accuracy was higher for *true* (85%) than *false* (68%) sentences, while in the *false belief*

condition accuracy was higher for *false* (82%) than *true* (39%) sentences. There was no significant effect of the participant's age.

[Table 3 and Figure 4 here]

*3.2.2.2 Acceptance rates*

See Table 4 and Figure 5 for summaries of the results.

There was a significant main effect of SENTENCE TRUTH: children were more likely to respond 'yes' when the sentence was *true* ($\beta = 1.68$, Wald's $z = 4.22$, $p = .000024$). There was a significant main effect for one of the BELIEF TYPE contrasts: acceptance rates were significantly higher in the *true belief* compared to the *false belief* condition ($\beta = 1.46$, Wald's $z = 2.03$, $p = .043$), but there was no significant difference between the *false belief* and *unknown* condition. There was also a marginal interaction between the first BELIEF TYPE contrast and SENTENCE TRUTH ($\beta = 1.26$, Wald's $z = 1.71$, $p = .087$): in the *true belief* condition the difference in acceptance rates for *true* and *false* sentences was greater (85% vs. 31%) than in the *false belief* condition (39% vs. 18%). There was no significant effect of age.

Since children accepted the *true* sentences in the *false belief* condition less than half of the time, we used a model of the *false belief* condition alone to directly compare the acceptance rates in *true* vs. *false* sentences. The difference was significant: children accepted *true* sentences more often ($\beta = 0.826$, Wald's $z = 2.76$, $p = .0059$).

[Table 4 and Figure 5 here]

### 3.2.3 False belief task

48 of the 52 participants in the analysis completed the false belief task. Children were grouped by their FB score (see Table 5). There was no significant difference in age between the three groups (one-way ANOVA, $p = .22$). There were no significant effects of FB score on

accuracy in the truth-value judgment task. As shown in Table 5, mean accuracy was extremely similar across the three groups overall (65-68%) and on *false belief* trials alone (58-60%).

[Table 5 here]

### 3.3 Discussion

The pattern of children's responses in this experiment demonstrates that they understand the literal meaning of belief reports. They correctly reject literally *false* sentences and accept *true* ones, with one main exception: they tend to reject *true* sentences in *false belief* scenarios. It is in the *false belief* scenarios that the truth of the sentence and of the complement clause diverge, leading to the possibility that children might judge the sentence based on the truth of the complement. However, they only do so for true sentences, apparently rejecting them on the basis of the complement clause. They are not tempted by the truth of the complement clause in a false sentence.

The contrast in children's performance with *true* and *false* sentences in the *false belief* condition suggests that their well-known difficulty evaluating belief reports in false belief contexts is not due to an incorrect semantic or conceptual representation of belief. While their difficulty with *true* sentences in *false belief* scenarios is consistent with all three hypotheses, only the pragmatic hypothesis explains their success with the *false* sentences. 3-year-olds are sensitive to the literal meaning of belief reports, and are able to evaluate whether the subject holds the stated belief. Nevertheless, with *true* sentences they may often default to a speaker meaning in which the complement clause is the main point of the utterance, and the speaker endorses its truth.

Children's accuracy was not predicted by their age, suggesting that the literal meaning of *think* is in place by around 3 years of age. Furthermore, children's accuracy was not predicted by

their success on a traditional false belief task. This suggests that the acquisition of belief reports is not dependent on the skills needed to pass a false belief task. However, we should be cautious about interpreting the results from the false belief task, since participants completed it after having been exposed to twelve belief reports during the truth-value judgment task. It is possible that this more concentrated exposure to mental state language affected their behavior on the false belief task. An effect of this sort might explain the surprising finding that there was no difference in age between children who failed the false belief task and those who passed it.

In this experiment as in Experiment 1, children's accuracy in the *unknown* condition (the *ignorance* condition of Experiment 1) was above chance but middling compared to accuracy in the *true belief* condition. Accuracy in this condition, like the others, did not change with age. Middling accuracy might be expected if many of the children interpret the complement clause as the main point, since the truth of the complement clause in reality cannot be evaluated. This possibility is supported by the fact that children were more accurate with *false* sentences (70% rejection) compared to *true* sentences (59% acceptance). However, the pragmatic hypothesis predicts that children should be able to reject literally false sentences in this condition just as well as in the *false belief* condition (82%). One possible explanation for the difference in accuracy is that the relevance of beliefs in the story is greater in the knowledge conditions, so children track them more carefully. Anecdotally, in the knowledge conditions children often made comments during the story about which seeker got it right or wrong. In the ignorance condition, on the other hand, children were focused on figuring out where the hider was; their spontaneous comments were mostly about their own guess about the hider's location (although we did try to prevent participants from committing to a guess, and excluded a few trials where they explicitly did so). They didn't care much about what the seekers thought, because they had

the impression that the seekers' beliefs were as arbitrary as their own in this case. Thus, counterintuitively, removing reality from the equation may have reduced children's attention to the beliefs in the story, rather than enhancing it. They are less prepared to evaluate a belief report because they have not tracked the beliefs as carefully.

Another result that deserves more scrutiny is that children's accuracy on *false* sentences in the *true belief* condition was lower than for *true* sentences (68% compared to 88%). Children accepted these *false* sentences 31% of the time. Even some of the oldest children in the study answered incorrectly in this condition. Although children might have more difficult rejecting sentences than accepting them in general, this pattern stands in contrast to the other conditions in this study, where children's performance was as good or better for false sentences. Since this result was not predicted by any of our hypotheses, we can only speculate as to its source. Consider the sample scenario illustrated in Figure 3, where Boots correctly believes that Swiper is hiding behind the curtain. A *false* sentence in the *true belief* condition would be (27).

(27) Boots thinks that Swiper is behind the toy box.

The sentence incorrectly attributes a false belief to Boots: Boots doesn't think that Swiper is behind the toy box, and in fact Swiper is not behind toy box. One would expect the double falsity of the sentence to make it easier for children to reject: there's nothing temptingly right about it. Since the other seeker holds the opposite belief, a contrasting true proposition is readily available regardless of whether the subject or the complement clause is taken to be in focus, as demonstrated in (28)-(29). Perhaps this double falsity confused some children: they might have found it odd that the speaker would be so wrong about what happened in the story, and second-guessed their own knowledge.

(28) No, DORA thinks that Swiper is behind the toy box.

(29) No, Boots thinks that Swiper is behind the CURTAIN.

Although the more unexpected findings from Experiment 2 warrant additional research, the main conclusions are strong. Children are capable of evaluating the literal meaning of belief reports, as evidenced by their rejection of literally false sentences in all conditions, even false belief scenarios. However, they tend to assume that the speaker meaning has to do with reality, as evidenced by their reality-based evaluation of literally true sentences.

# 4 General Discussion

In Experiment 1, we showed that when beliefs are made more relevant by introducing a conflict of beliefs, 3-4 year-olds are better able to evaluate reports about false beliefs. In Experiment 2, we showed that 3-year-olds understand the literal meaning of belief reports, even in false belief contexts. Together, these findings suggest that children's non-adult-like interpretations of belief reports should be attributed to a pragmatic problem, rather than a syntactic/semantic or conceptual deficit. Children are less able to determine when beliefs are under discussion in the discourse. Their default assumption is that beliefs are not relevant—perhaps because of the low frequency of conversations about beliefs.

This conclusion fits well with previous evidence that children are more likely to assume that an expression refers to a mental state if the context makes mental states highly salient and relevant (Papafragou, Cassidy, & Gleitman, 2007). It is also consistent with research suggesting that children tend to acquire mental state language and pass false belief tasks earlier if their exposure to mental state language (from their parents) includes more third-person, explanatory, causal, contrastive uses (Adrian, Clemente, & Villanueva, 2007; Brown, Donelan-McCall, & Dunn, 1996; Gola, 2012; Howard, Mayeux, & Naigles, 2008; Slaughter, Peterson, & Mackintosh, 2007; Taumoepeau & Ruffman, 2008). That is, if children have more experience

with conversations that highlight mental states, they may be more able to recognize when mental states are relevant in conversation.

An alternative explanation for the results is that the pattern of children's judgments is generated by a semantically-encoded endorsement reading, rather than a pragmatic-level speaker meaning. That is, children's semantic representation for *think* is something like *think correctly*. This is a tempting hypothesis: if children constantly hear uses of belief reports in which the speaker endorses the truth of the complement clause, why shouldn't they conclude that *think* means *think correctly*?

There are differences in the truth and felicity conditions of *think correctly* vs. a pragmatically-derived endorsement interpretation, but they are subtle. To explain, we first have to decide what we mean by *think correctly*. There are no verbs in English which simultaneously assert both an attitude attribution and the speaker's perspective on the truth of the content of the attitude. One of these meaning components is always presupposed—taken by the speaker to be shared background information. For a meaning close to *think correctly*, we have two options: *know*, which presupposes the truth of the complement, and *be right*, which presupposes the belief. For example, (30) asserts that Dora believes that Swiper is behind the curtain, and presupposes that he actually is. (31) asserts that Swiper is behind the curtain, and presupposes that Dora believes it (Abusch, 2002; 2010). To assert both parts simultaneously you need two content words—*think correctly*. The truth conditions for these three ways of encoding the meaning are represented in Table 6, along with those of a pragmatic endorsement interpretation.

(30) Dora knows that Swiper is behind the curtain.

(31) Dora is right that Swiper is behind the curtain.

[Table 5 here]

We can test whether children treat part of the meaning as presupposed by seeing how they interpret negated sentences. Dudley, Orita, Hacquard, and Lidz (2015) have demonstrated that three-year-olds do not treat *think* like *know*—they know that *think* does not presuppose its complement. In a setup where a toy is hidden in either a red or a blue box, children understand that in (32), the toy must be in the blue box, while in (33) it is likely to be in the red box.[8]

(32) Lambchop doesn't know that the toy is in the blue box.

(33) Lambchop doesn't think that the toy is in the blue box.

Thus, we can eliminate the possibility that children treat *think* as *know*, but thus far we have no conclusive evidence that they do not treat *think* as either *be right* or *think correctly*. However, there are at least two pieces of evidence that speak against these possibilities. First, children's sensitivity to context—their improved performance in the 2-seeker condition in Experiment 1—is only expected under a pragmatic account. Second, children use *think* with an uncertainty implication quite often in their own production (Bloom, Rispoli, Gartner, & Hafitz, 1989; Diessel & Tomasello, 2001; Shatz, Wellman, & Silber, 1983). This reading would be impossible if children assumed that the literal meaning of *think* was *think correctly* or *be right*. We conclude that the pattern of children's responses is best explained by pragmatic difficulty.

The pragmatic hypothesis may be extended to explain some of children's difficulty in traditional false belief tasks as well. Children show the most difficulty in tasks where they must evaluate a statement or question posed to them directly. This setup requires the child to track the

[8] Not all 3-year-olds seemed to know that *know* presupposes the truth of its complement, and would thus pick the red box after a clue like (32). However, all 3-year-olds seem to understand that *think* doesn't presuppose the truth of its complement: like adults, they consistently chose the red box after a clue like (33).

experimenter's perspective and conversational intentions in addition to the perspectives of the characters in the story (Helming, Strickland, & Jacob, 2014). For example, after hearing the story about Maxi and the chocolate described in the introduction, children under 4 years are very poor at answering the direct question in (34) (Wimmer & Perner, 1983). They are better at acting out a response to a prompt like, "What happens next?" (Rubio-Fernández & Geurts, 2013), or looking in response to the experimenter's "wondering" aloud, as in (35) (He, Bolz, & Baillargeon, 2012). Findings like these suggest that what makes the false belief task difficult is the need to puzzle out the experimenter's intentions in asking the question. When that need is removed, or when the experimenter's intentions are made more clear (e.g. Hansen, 2010), children are more able to demonstrate that they can track false beliefs.

(34) Where will Maxi look for the chocolate?

(35) I wonder where Maxi will look for the chocolate…

We do not claim that all of children's difficulties with false belief tasks are reducible to this particular pragmatic difficulty. The remarkably robust difference in children's behavior on "implicit" vs. "explicit" false belief tasks is compelling evidence for some kind of change in competence (for opinionated reviews, see Apperly & Butterfill, 2009; Perner & Roessler, 2012). Mental state language may play a critical role in explicit tasks, by providing representations of mental states that can be used by general reasoning mechanisms (Carruthers, 2002; 2009). However, we have argued that children have adult-like linguistic representations of mental state language much earlier than that. This opens up new questions about whether and how children's mental state language maps onto the allegedly rudimentary belief concepts available to 2-year-olds.

## 5 Conclusion

We have investigated the "curse of knowledge" in children's interpretation of belief reports—their tendency to be influenced by reality when judging sentences about false beliefs. We argued that children's non-adult-like interpretations derive from their understanding of the pragmatics of belief reports, and thus need not be attributed to non-adult-like conceptual or syntactic/semantic representations, as has been previously proposed. In fact, they seem to know the literal meaning of *think* by the youngest age we tested (37 months): they are able to reject *think* sentences that report a belief incorrectly. Children's difficulty relates to determining the speaker's meaning in context. They assume that speakers are generally talking about reality, not mental states, and therefore sometimes reject true sentences in false belief scenarios.

## 6 Acknowledgements

This work was funded in part by NSF grant #BCS-1124338. Many thanks to our undergraduate research assistants (Amber Frazier, Faina Kostyukovsky, Jessica Lee, Sara McVeigh, Laura Sherry, and Leah Whitehill), and the Attitudes and Theory of Mind research group (Rachel Dudley, Kate Harrigan, Morgan Moyer, Naho Orita, and Aaron Steven White).

## 7 References

Abusch, D. (2002). Lexical alternatives as a source of pragmatic presupposition. In B. Jackson (Ed.), *Proceedings of SALT 12.* CLC Publications.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics, 27*(1), 37-80.

Adrian, J. E., Clemente, R. A., & Villanueva, L. (2007). Mother's use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*(4), 1052-1067.

Apperly, I., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953-970.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110-118.

Birch, S. A., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14*(3), 283-286.

Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language, 16*, 101-120.

Bolinger, D. (1968). Post-posed main phrases: An English rule for the Romance subjunctive. *Canadian Journal of Linguistics, 14*(1), 3-30.

Bresnan, J. (1968). *Remarks on adsententials.* Manuscript, MIT.

Brown, J., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. *Child Development, 67*, 836-849.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337-342.

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

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*, 657-726.

Carruthers, P. (2009). An architecture for dual reasoning. In J. S. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 109-127). Oxford: Oxford University Press.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*(6), 1263-1277.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377-395.

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*, 1858-1878.

de Villiers, J. G. (1995). Questioning minds and answering machines. In D. MacLaughlin, & S. McEwen (Eds.), *Boston University Conference on Language Development (BUCLD) 19* (pp. 20-36). Somerville, MA: Cascadilla Press.

de Villiers, J. G. (2005). Can language acquisition give children a point of view? In J. W. Astington, & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 186-219). New York: Oxford University Press.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell, & K. J. Riggs (Eds.), *Children's Reasoning and the Mind* (pp. 191-228). Hove, U.K.: Psychology Press.

de Villiers, J. G., & de Villiers, P. A. (2009). Complements enable representation of the contents of false beliefs: the evolution of a theory of theory of mind. In S. Foster-Cohen (Ed.), *Language Acquisition.* New York: Palgrave Macmillan.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development, 17*, 1037-1060.

de Villiers, P. A. (2005). The role of language in Theory of Mind development: what deaf children tell us. In J. W. Astington, & J. A. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 266-297). New York: Oxford University Press.

de Villiers, P. A., Burns, F., & Pearson, B. (2003). The role of language in the Theory of Mind development of language-impaired children: complementing theories. In B. Beachley, A. Brown, & F. Conlin (Ed.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 232-242). Somerville, MA: Cascadilla Press.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics, 12*, 97-141.

Dudley, R., Orita, N., Hacquard, V., & Lidz, J. (2015). Three year olds' understanding of 'know' and 'think'. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions* (pp. 241-262). Springer.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabular learning. *Cognition, 73*, 135-176.

Gola, A. A. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development, 27*, 64-76.

Hacquard, V. (2014). Bootstrapping attitudes. *Proceedings of SALT 24.*

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science, 6*, 346-359.

Hansen, M. B. (2010). If you know something, say something: young children's problem with false beliefs. *Frontiers in Psychology, 1*(23).

He, Z., Bolz, M., & Baillargeon, R. (2012). 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology, 30*, 14-29.

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*.

Hooper, J. B. (1975). On assertive predicates. In J. P. Kimball (Ed.), *Syntax and Semantics* (Vol. 4, pp. 91-124). New York: Academic Press.

Howard, A. A., Mayeux, L., & Naigles, L. R. (2008). Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language, 28*(4), 375-402. doi:10.1177/0142723708091044

Jørgensen, R. N., Dale, P. S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language, 37*(2), 419-428.

Johnson, C. N., & Maratsos, M. P. (1977). Early comprehension of mental verbs: Think and know. *Child Development, 48*, 1743-1747.

Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science, 15*(1), 113-122.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830-1834.

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

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of Mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology, 44*(2), 523-531.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130-1144.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-analysis of the relation between language ability and false belief understanding. *Child Development, 78*(2), 622-646.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*, 167-171.

Naigles, L. (2000). Manipulating the input: Studies in mental verb acquisition. In B. Landau, J. Sabini, J. Jonides, & E. L. Newport (Eds.), *Perception, cognition, and language: Essays in honor of Henry and Lila Gleitman* (pp. 245-274). Cambridge, MA: MIT Press.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67*(2), 659-677.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false belief? *Science, 308*, 255-258.

Papafragou, A., Cassidy, K., & Gleitman, L. R. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105*, 125-165.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of beliefs. *Trends in Cognitive Sciences, 16*(10), 519-525.

Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). 'Want that' is understood well before 'say that', 'think that', and false belief: a test of de Villiers's linguistic determinism on German-speaking children. *Child Development, 74*, 179-188.

Roberts, C. (2004). Context in dynamic interpretation. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 197-220). Blackwell.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics, 5*(6), 1-69.

Rooryck, J. (2001). Evidentiality, Part I. *Glot International, 5*, 125-133.

Ross, J. R. (1973). Slifting. In M. Gross, M. Halle, & M.-P. Schützenberger (Ed.), *The formal analysis of natural languages. Proceedings of the First International Conference.* (pp. 133-169). The Hague: Mouton.

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*(1), 27-33.

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental Science, 15*(2), 181-193.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*, 301-321.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua, 117*, 1034-1056.

Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and Theory of Mind in typical children and those on the autism spectrum. *Child Development, 78*(3), 839-858.

Song, H.-j., & Bailargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*, 1789-1795.

Song, H.-j., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition, 109*, 295-315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science, 13*(6), 907-912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 10*, 587-592.

Sowalsky, E., Hacquard, V., & Roeper, T. (2009). Is PP opacity on the path to false belief? In J. Crawford, K. Otaki, & M. Takahashi (Eds.), *Generative Approaches to Language Acquisition North America (GALANA) 3* (pp. 263-261). Somerville, MA: Cascadilla Proceedings Project.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580-586.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false belief in children with autism. In J. W. Astington, & J. A. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 298-318). New York: Oxford University Press.

Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin- and Cantonese-speaking children. *Developmental Psychology, 36*, 25-43.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language adn emotion understanding at 15, 24, 33 months of age. *Child Development, 79*, 284-302.

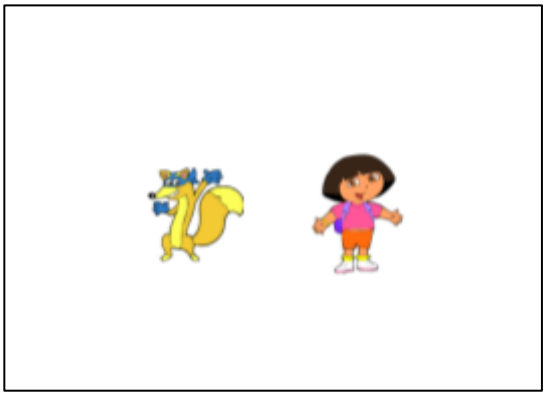
Urmson, J. O. (1952). Parenthetical verbs. *Mind, 61*, 480-490.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*, 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

## INTRODUCTION

**1 seeker** | **2 seekers**

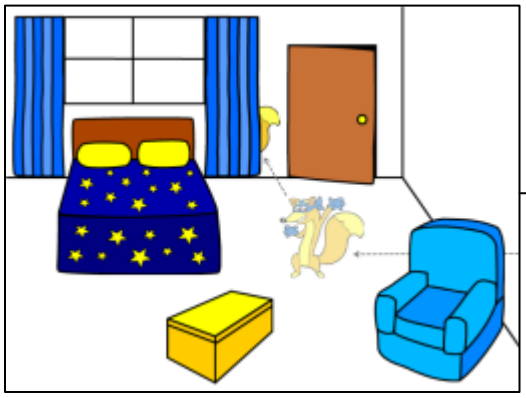


(a)

(b)

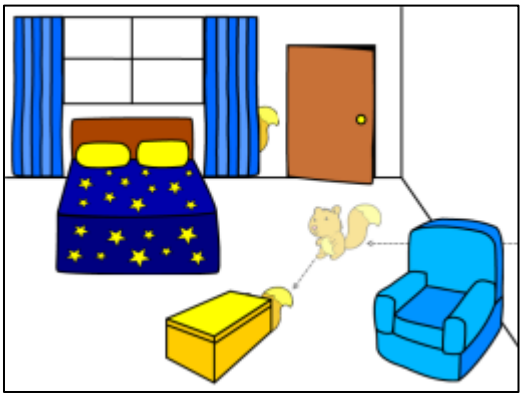
*This time [Boots,] Swiper and Dora are gonna play. Swiper's gonna hide, and Dora [and Boots] are gonna look for him. So she/they'll wait in the other room.*

## HIDING



(c)

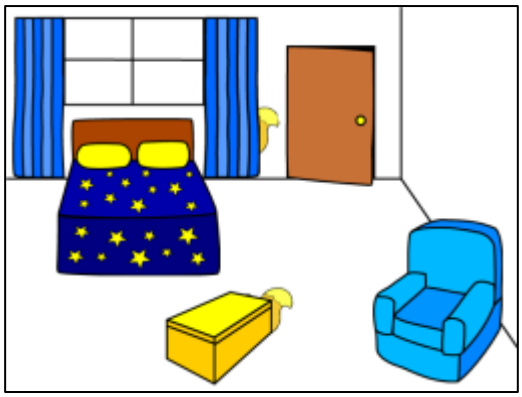
*Here goes Swiper, he's hiding…*



(d)

*And look! That's not Swiper!*

(e)



Experimenter:
*Look, we can see a yellow tail behind the curtain, and a yellow tail behind the toy box!*

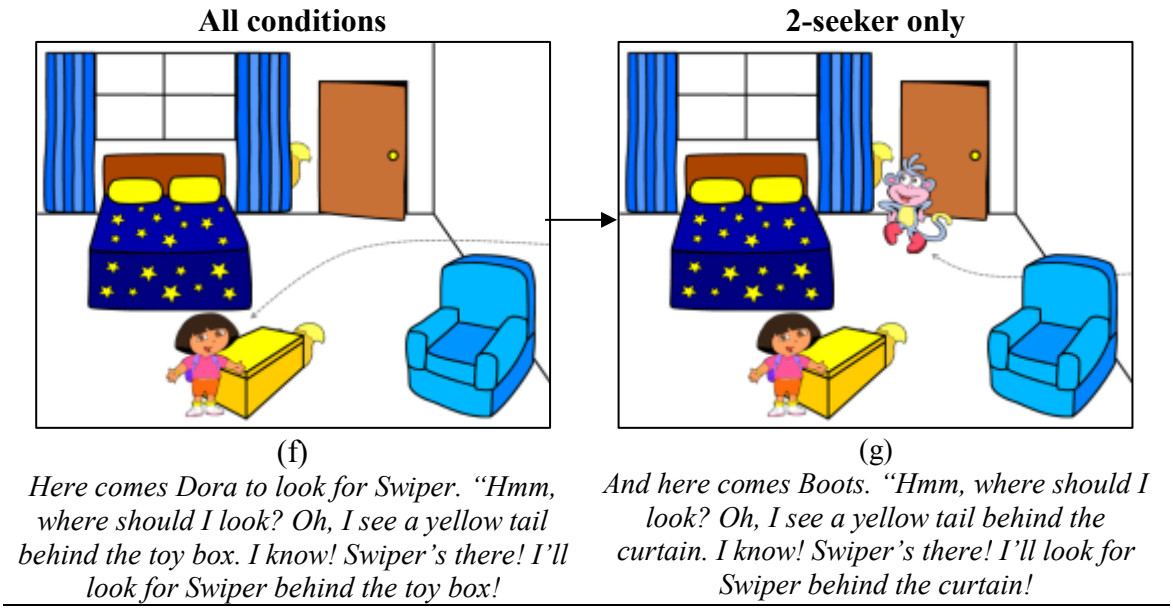



**Figure 1** **Experiment 1: Sample scenes.**

(a-b) Characters are introduced. (c-e) Hider (Swiper, behind the curtain) and Distractor (squirrel, behind the toy box) hide, leaving identical "clues" visible. (f) The first Seeker (Dora) guesses the location of the Hider. (g) In the *2-seeker* condition, a second seeker (Boots) guesses a different location for the hider.

| Sample sentence | Belief Type | Sent. Truth | Comp. Truth |
|---|---|---|---|
| Boots thinks that Swiper is behind the curtain. | TB | T | T |
| Boots thinks that Swiper is behind the toy box. | TB | F | F |
| Dora thinks that Swiper is behind the toy box. | FB | T | F |
| Dora thinks that Swiper is behind the curtain. | FB | F | T |

**Table 1** **Experiment 1: Target sentence types.**

Belief Type (TB = true belief; FB = false belief), truth of the sentence, and truth of the complement clause.

| Knowledge | Belief Type | 1-seeker | 2-seeker |
|---|---|---|---|
| knowledge | true belief | 83%** | 90%** |
| | false belief | 36%° | 52% |
| ignorance | true belief | 69%* | 81%** |
| | false belief | 56% | 82%** |

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

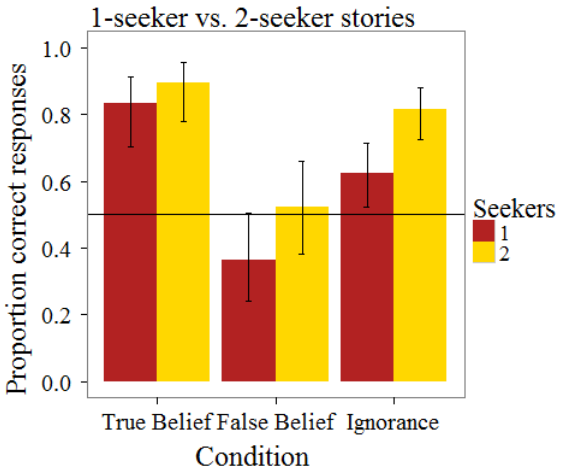



**Figure 2** **Experiment 1: Accuracy rates by condition.**

Error bars represent 95% confidence intervals based on the binomial distribution.

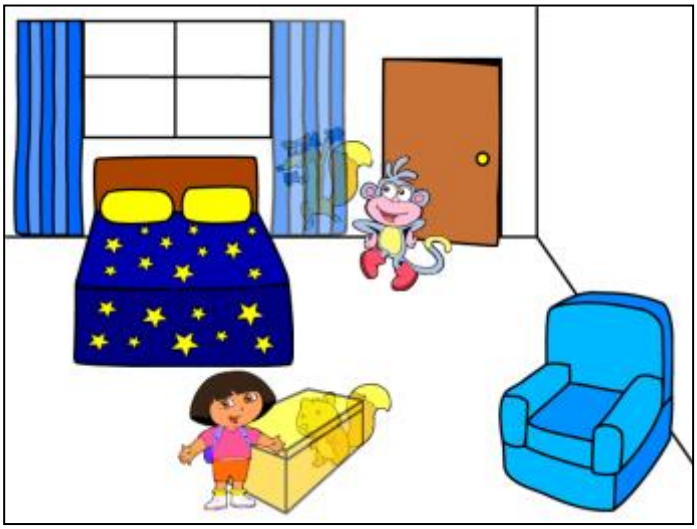


**Figure 3** **Experiment 2: Sample scene.** (Transparency added for expository purposes.)

| | | Sentence Truth | |
|---|---|---|---|
| | | *T* | *F* |
| Belief Type | | 61%*** | 73%*** |
| *true belief* | 77%*** | 85%*** | 68%*** |
| *false belief* | 60%** | 39%* | 82%*** |
| *unknown* | 64% *** | 59% | 70%*** |

**Table 3** **Experiment 2: Accuracy rates by condition.**

Stars indicate difference from chance: * $p < .05$, ** $p < .01$, *** $p < .001$

| | Sentence Truth | |
|---|---|---|
| | T | F |
| Belief Type | 61%*** | 25%*** |
| *true belief* | 85%*** | 31%*** |
| *false belief* | 39%* | 18%*** |
| *unknown* | 60% | 27%*** |

**Table 4** **Experiment 2: Acceptance rates ('yes' responses) by condition.**

Stars indicate difference from chance: * $p < .05$, ** $p < .01$, *** $p < .001$

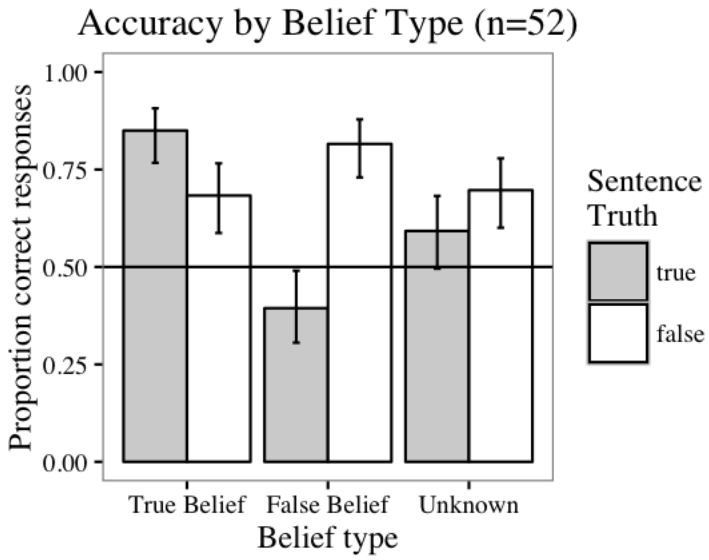



**Figure 4** **Experiment 2: Accuracy rates by condition.**

Error bars represent 95% confidence intervals based on the binomial distribution.

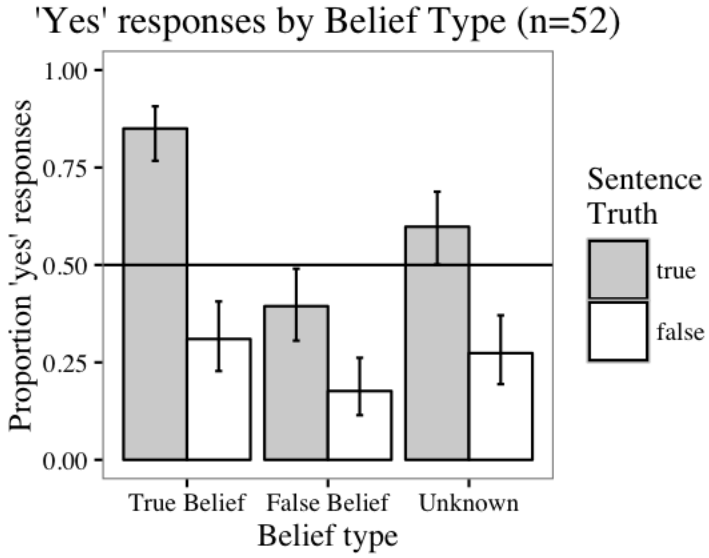



**Figure 5** **Experiment 2: Acceptance rates ('yes' responses) by condition.**

Error bars represent 95% confidence intervals based on the binomial distribution.

| | FB Score | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| Number of children | 16 | 11 | 21 |
| Mean age in months (sd) | 43.3 (3.5) | 43.0 (2.6) | 44.6 (4.0) |
| Mean accuracy: overall | 66% | 68% | 65% |
| Mean accuracy: FB trials | 59% | 58% | 60% |

**Table 5** **Experiment 2: False Belief scores.**

| *x thinks that p* | *p* | *think* | *think correctly* | *know* | *be right* | pragmatic endorsement |
|---|---|---|---|---|---|---|
| T | T | T | T | T | T | T |
| T | F | T | F | # | F | T/F |
| F | T | F | F | F | # | F |
| F | F | F | F | # | # | F |

**Table 6** **Truth table for variations on *think*.** "#" signifies infelicity in the presupposition-violating conditions. The cell marked "T/F" under the pragmatic endorsement is slightly different. In this case, the literal meaning is satisfied but the implicated meaning—that the complement clause is true—is problematic. The natural response is something like "Yes, but…", which seems subtly distinct from the "Wait a minute!" response to a presupposition violation.