J. Child Lang. **44** (2017), 905–923. © Cambridge University Press 2016
doi:10.1017/S0305000916000301

Learning words from speakers with false beliefs*

ANNA PAPAFRAGOU
Department of Psychological & Brain Sciences, University of Delaware

SARAH FAIRCHILD
Department of Psychological & Brain Sciences, University of Delaware

MATTHEW L. COHEN
Department of Physical Therapy, University of Delaware

AND

CARLYN FRIEDBERG
Massachusetts General Hospital Institute of Health Professions

(*Received* 3 *November* 2015 – *Revised* 17 *February* 2016 – *Accepted* 16 *May* 2016 – *First published online 21 June 2016*)

ABSTRACT

During communication, hearers try to infer the speaker's intentions to be able to understand what the speaker means. Nevertheless, whether (and how early) preschoolers track their interlocutors' mental states is still a matter of debate. Furthermore, there is disagreement about how children's ability to consult a speaker's belief in communicative contexts relates to their ability to track someone's belief in non-communicative contexts. Here, we study young children's ability to successfully acquire a word from a speaker with a false belief; we also assess the same children's success on a traditional false belief attribution task. We show that the ability to consult the epistemic state of a speaker during word learning develops between the ages of three and five. We also show that false belief understanding in word-learning contexts proceeds similarly to standard belief-attribution

[*] This research was supported in part by NSF grant BCS-0641105 to A. P. and by Science and Engineering Scholarships awarded by the University of Delaware Undergraduate Research Office to C. F. and M. C. The authors wish to thank Laurie Yarzab, Kendra Goodwin, Taraneh Mojaverian, Alice Ding, Yun Choi, Angela Yung, Dimitris Skordos, and Ann Bunger for their help. Address for correspondence: Anna Papafragou, Department of Psychological and Brain Sciences, University of Delaware, Newark DE 19716. e-mail: sarahcfairchild@gmail.com

contexts when the tasks are equated. Our data offer evidence for the development of mind-reading abilities during language acquisition.

INTRODUCTION

Human communication relies on the ability to track beliefs and other epistemic states, since hearers try to infer the speaker's intentions during conversation (Grice, 1989). The ability to consult epistemic states during communication has been argued to be present early in life: infants take into account a speaker's knowledge when fixing reference for novel expressions (Baldwin, 1993; Bloom, 2000; Koenig & Echols, 2003), and exclude as potential referents for novel labels objects that they already have names for (Diesendruck & Markson, 2001). Other studies show that children trust knowledgeable over uninformed speakers when acquiring novel labels: when a speaker explicitly claimed ignorance about the correct referent of a novel word (e.g. "I don't know what a blicket is. Maybe it's this one."), three- and four-year-olds did not learn the word (Sabbagh & Baldwin, 2001). In another study, four-year-olds, and in some cases three-year-olds, preferred to learn a novel label from a previously reliable speaker than from one who claimed ignorance of the names of familiar objects or who labeled them incorrectly (Koenig & Harris, 2005; cf. also Birch, Vauthier & Bloom, 2008; Jaswal & Neely, 2006).

Nevertheless, whether (and how early) preschoolers track their interlocutors' mental states is still a matter of debate. For instance, some commentators have suggested that young children's ability to match previously unlabeled referents to novel words does not require the use of pragmatic reasoning, since this ability is present in autistic populations who have deficits in social cognition (de Marchena, Eigsti, Worek, Ono & Snedeker, 2011; see also Breheny, 2006). Other work has questioned whether older children and even adults successfully adopt the perspective of their interlocutors during referential communication (Epley, Morewedge & Keysar, 2004; but see Nadig & Sedivy, 2002). Moreover, several studies have shown that five-year-olds have difficulties with pragmatic inferences that rely on calculations of how informative a speaker should be based on her knowledge and the conversational needs (Noveck, 2001; Papafragou & Musolino, 2003; Huang & Snedeker, 2009; among others).

A particularly stringent and straightforward test for preschoolers' ability to consult others' belief states during communication is to ask whether children can adjust their expectations about the speaker's referential intent when learning a word from a speaker with a false belief. False belief paradigms have been used extensively in the developmental literature to study sensitivity to others' epistemic states in non-communicative contexts (i.e. contexts that involve understanding and predicting someone's actions;

see Wimmer & Perner, 1983; Baron-Cohen, Leslie & Frith, 1985). Several studies have adopted these paradigms to ask whether young children can reason about a communicator's referential intent even in circumstances in which the communicator holds a false belief – but results have been somewhat mixed. In the first study of this kind, Happé and Loth (2002) gave three- to four-year-old children a false belief Word Learning task (adapted from Baldwin, 1993). In this task, a puppet called Mary put a novel object in a box, closed the lid, and left the scene. Another puppet called Tom entered, removed Mary's object from the box, and put another novel object in. Mary returned to the scene, picked up the closed box and named the contents ("Do you want to see the modi? There's a modi in this box! Let's see the modi!"). Next, children were presented with the two novel objects and were asked to indicate which one was the modi. Results indicated that 87% of the children who had previously succeeded in control questions passed that task. However, because of the way the data are reported, the precise ages of the passers and any developmental differences between three- and four-year-olds cannot be ascertained from this study. Furthermore, other studies suggest that preschoolers' ability to adjust word learning when the speaker has a false belief may be more limited: Carpenter, Call, and Tomasello (2002, Exp. 2) gave children around their third birthday a Word Learning task similar to the one used by Happé and Loth, except that the original object was removed from its location and not substituted by another novel object. About 40% of these children (after taking into account performance on control trials) passed the Word Learning task. More recent experiments by Houston-Price and colleagues (2011) also lead to higher estimates. For instance, in one of their experiments (Exp. 3), that was similar to Happé and Loth's, only about 55% of three-year-olds and 50% of four-year-olds passed a false belief Word Learning task. In sum, existing evidence does not establish firmly how or when preschoolers develop the ability to use mental states to acquire novel words.

One might assume that a more detailed developmental investigation of the ability to track a speaker's false beliefs is unnecessary given the wealth of research on the development of false belief attribution in standard (non-communicative) contexts. This research has shown that the ability to correctly report false beliefs (and hence to accurately reason about objects of belief as distinct from reality) is typically absent in three-year-olds but emerges by the time children are four or five years of age (see Wellman, Cross & Watson, 2001, for a review; but cf. Clements & Perner, 1994; Garnham & Ruffman, 2001; Onishi & Baillargeon, 2005; Southgate, Senju & Csibra, 2007; Surian, Caldi & Sperber, 2007; Buttelmann, Carpenter & Tomasello, 2009; Rubio-Fernandez & Geurts, 2013, for evidence of earlier success with simpler tasks). However, some of the studies mentioned

above have raised the possibility that the developmental timetable for understanding epistemic states might differ between word-learning and standard belief-attribution contexts. Happé and Loth's (2002) study included a standard false belief Action Prediction task (modified from Baron-Cohen *et al*., 1985). Children were shown an act-out story in which a puppet called Sally put her ball in a basket and went out. Another puppet ('naughty' Ann) came in and decided to play a trick on Sally, moving her ball from the basket to a box. Sally returned and wanted to play with her ball. Children were asked: "Where will Sally look for her ball?" It was found that children were more likely to succeed in tracking false belief in the Word Learning task than in the Action Prediction task. The authors concluded that understanding thoughts behind words might be easier for children than understanding thoughts behind actions, and took their data as tentative evidence supporting the presence of distinct theory of mind mechanisms supporting action prediction vs. communication (in accordance with Roth & Leslie, 1991; Sperber, 2000; Wilson, 2000). Similarly, Carpenter *et al*. (2002) reported that their sample of three-year-olds performed very poorly on the same action prediction task, despite the fact that, as already mentioned, 40% of these children passed word learning trials.

One issue with these studies is that there were several asymmetries between the types of scenarios used to test false belief in action prediction vs. communication contexts. To take a specific example, in both Happé and Loth (2002) and in Carpenter *et al*. (2002), the Word Learning scenario allowed for a tighter association between the protagonist and the target object compared to the Action Prediction scenario, since, in word learning trials, the protagonist only ever came into physical contact with the original object, and never with the object that was used for its substitution. In the Carpenter *et al*., study, the difference was even starker because the target object was not substituted by another object but removed from the experimental scene altogether in the Word Learning scenarios. Thus, when the character who originally hid the object came back and uttered a word for the object while trying to retrieve it, there was no candidate referent in the original location for children to mistakenly label. This manipulation differs markedly from the set-up of the classic Action Prediction task (see Wimmer & Perner, 1983; Koos, Gergeley, Csibra & Biro, 1997; Wellman *et al*., 2001). There are additional task differences in the number of objects, hiding locations, and test questions in Happé and Loth (2002). Even after removing several such asymmetries between the two types of scenarios, more recent studies with three- and four-year-olds by Houston-Price and colleagues (2011, Exp. 1 and 2) replicated the advantage of word learning over action prediction false belief scenarios.

Nevertheless, Houston-Price and colleagues (2011) also showed that the advantage of word learning scenarios could be eliminated. Following Perner, Rendi, and Garnham (2007), they hypothesized that the Action Prediction questions ("Where will X look for Y?") in prior studies and their own initial experiments required greater resistance to a 'reality bias' exerted by the actual location of the target object that the protagonist desired compared to the Word Learning questions ("Which one is the modi?") that involved a not-yet-established referential link. Houston-Price *et al*. (2011, Exp. 3) modified the Action Prediction task used in their previous experiments in a way known to powerfully reduce the 'reality pull' (Wimmer & Perner, 1983; Wellman *et al*., 2001): in the new version, after the target object was switched by a second character, the character would remove both objects and leave. As expected, children performed better in this Action Prediction task compared to the earlier version; furthermore, and crucially, the difference from the Word Learning task disappeared. Houston-Price and colleagues concluded that what appeared to be precocious understanding of false belief in the word-learning domain was driven by a weaker 'reality bias' inherent in the naming task itself.

At present, a synthesis of existing findings is complicated: only one experiment so far has failed to produce the asymmetry in false belief reasoning across domains (Houston-Price *et al*., 2011, Exp. 3), and conclusions drawn from that experiment critically depend on whether one accepts the premise that the newly introduced label in the Word Learning task exerts no or little 'referential pull'. Notice that this idea is not supported by independent evidence. In fact, it seems plausible to assume that the Word Learning task still requires the child to override the tendency to interpret the tricked speaker's sentence (e.g. "The modi is in this box") as if it refers to the here-and-now of the situation (as the child currently represents it), and hence to attach the new label for the desired object to the true contents of the box. On this line of reasoning, simplifying the Action Prediction task by removing the need to inhibit the incorrect response (the single most powerful way to boost children's ability to pass the false belief task, after the effects of age, according to Wellman *et al*.'s (2001) meta-analysis of 178 studies) created a massive imbalance with respect to the Word Learning task (that still involved inhibition) – hence eliminating the previously observed advantage of the latter task.

The alternative explanation sketched above is compatible with the proposal that understanding false labels develops separately from understanding false beliefs (since it assumes that the difference is simply masked in Houston-Price *et al*., 2011, Exp. 3). It is also compatible with the possibility that task demands – not necessarily related to the weaker 'reality bias' in interpreting novel labels – are responsible for the observed false belief advantage in communicative domains. Notice that, despite

efforts to equate task demands across domains, several major differences persisted throughout all studies that have reported the word-learning advantage (see Carpenter *et al*., 2002; Happé & Loth, 2002; Houston-Price *et al*., 2011). For instance, false belief reasoning in non-communicative contexts targeted location changes and was assessed through a location question ("Where … ?"); answers were given by pointing within the experimental scene; and false-belief contents involved familiar objects that belonged to the characters. By contrast, false belief reasoning in communicative contexts targeted object changes and was assessed through a referent-choice question ("Which one … ?"); answers were given by selecting one of two objects placed in front of the child but outside the experimental scene; and false belief contents involved novel objects. One or more of these differences may have made it easier for children to disengage their perspective from their own knowledge for purposes of word learning compared to action prediction. At this point, since several theoretical options are open, further experimentation is needed to settle the question whether false belief reasoning is easier in communicative compared to non-communicative contexts.

In the present experiment, our primary goal was to explore whether young children track a speaker's epistemic state (specifically, a false belief) when learning the meaning of a novel word. We compared three-, four-, and five-year-olds in a simple Label task that drew on Happé and Loth's (2002) Word Learning scenarios. In this task, one character placed a novel object into a box and left. A second character came in and put another novel object inside the box. The first character then came back and referred to the object in the box using a novel label ("blicket"). Children were presented with a choice between the two objects and asked: "Which one is the blicket?" Of interest was whether children would be able to reliably handle false beliefs in word learning contexts and how this ability would change across ages. Success on this task would offer strong support for the presence of mind-reading skills in children's early communication, especially in the face of diverging estimates in the literature (see Carpenter *et al*., 2002; Happé & Loth, 2002; Houston-Price *et al*., 2011).

A secondary goal was to compare preschoolers' understanding of epistemic states in word-learning vs. standard belief-reasoning contexts. We thus tested the same children on a Belief Attribution task that was identical to the Label task in terms of its questions, structure, and availability of a 'reality match' (unlike Carpenter *et al*., 2002; Happé & Loth, 2002; Houston-Price *et al*., 2011). The only difference was that, at the end of the story, the first character came back and, instead of introducing a novel word, simply approached the box. Children were then shown the two objects and asked: "Which one does <the character> think is inside the box?" Unlike earlier studies, we did not ask children to predict the

protagonist's action but to access the protagonist's thoughts. These two methods are equivalent in terms of assessing false belief reasoning (Wellman *et al*., 2001). This design allowed us to test three currently open theoretical possibilities. If the previously demonstrated advantage of communicative contexts was due to distinctly developing theory-of-mind mechanisms (Carpenter *et al*., 2002; Happé & Loth, 2002), then our data should show better performance on the (False) Label compared to the (False) Belief Attribution scenarios. If the advantage of communicative contexts was due to the lack of a strong 'referential pull' towards the incorrect answer during the interpretation of a novel word compared to non-word contexts (Houston-Price *et al*., 2011), the asymmetry should also persist. But if the advantage of communicative contexts was due to task differences in past studies that selectively encouraged perspective-shifting in word-learning contexts, it should disappear.

In line with prior work (Carpenter *et al*., 2002; Happé & Loth, 2002; Houston-Price *et al*., 2011), we also included True Belief trials. These trials preceded the main task and were identical to the false belief scenarios except that the protagonist was present when the object switch occurred. Two such pretest trials were administered, one involving Belief Attribution and the other Labeling. These trials served as a control for the possibility that children might associate a character with the object he/she placed in a box and might thus appear to succeed in the false belief trials without truly computing the character's mental state. If children followed this strategy, they would fail in the true belief trials, since the object that the first character placed in a box is not the object this character thinks is in the box / labels at the end of a true belief story. We therefore sought to ascertain that children could pass at least one, if not both, of these true belief pretests.

EXPERIMENT

METHOD

Participants

Participants fell into three age groups. The three-year-old group consisted of thirty-eight children between the ages of 3;1 and 3;11 (mean age: 3;6). The four-year-old group consisted of sixty-one children between the ages of 4;0 and 4;11 (mean age: 4;6). The five-year-old group consisted of thirty-nine children between the ages of 5;0 and 5;11 (mean age: 5;4). (Four-year-olds were sampled more heavily because their behavior on False Belief computation is variable and thus they might have been more likely than the other two groups to reveal dissociations, if any, between types of False Belief tasks.) Participants were recruited from daycares in Newark, DE.

Additional children were tested but excluded from the final sample for either failing all Memory questions within a single False or True Belief task ($n = 7$) or not answering test questions ($n = 1$).

Materials and procedure

Children were tested individually in a quiet room of their daycare. Materials included videotaped stories with various puppets displayed on a laptop. Each child saw four test stories, all of which involved False Belief scenarios. The test stories involved deception, a feature known to improve false belief performance (Chandler, Fritz & Hala, 1989; Sullivan & Winner, 1993; Wellman *et al*., 2001; see also Happé & Loth, 2002). There were two within-subjects tasks, a Label and a Belief Attribution task, with two test stories assigned to each task type (see the 'Appendix' for a sample story). In a typical test story, a wolf named Lucy appeared on the screen and introduced herself to the children. The wolf then drew the children's attention to a novel object in her hands ("Look what I have! Let's put it away now. Let's put it in this box to keep it safe!"). The wolf then placed the object into a box in the center of the screen. Then she said: "I'm tired now, so I'm going to take a nap. Will you watch this while I'm asleep? Keep it safe while I take a nap!" Then the wolf left. A monkey then came in holding another novel object and introduced himself as David. Then the monkey said: "Look what I have! Let's play a trick on Lucy. Ready? Let's be very quiet so we don't wake Lucy. I take this one out of the box and put this one in." The experimenter next paused the video and asked the children where Lucy went. If a child failed to respond, the experimenter answered that Lucy was taking a nap.

Then the experimenter asked a Perception question ("Can the wolf see what the monkey is doing?"). Children responded correctly 94% of the time. The experimenter showed agreement with correct responses ("You're right! She can't because she's away taking a nap!"). In the rare case that children responded incorrectly, the experimenter corrected them ("No, she can't! She's away taking a nap!").

Next, David switched the objects and left the scene taking Lucy's object with him. Then, Lucy returned. In the version of the story for the Label task, Lucy named the object in the box ("It's time to play with the modi! Remember the modi? There's a modi in this box!"). In the version of the story for the Belief Attribution task, Lucy remained silent. The decision to have the protagonist remain silent (and not, for instance, talk about "the toy" without naming it) was made so as to maximize the difference between a communicative and a non-communicative context: for children, having to identify the correct referent for the noun phrase "the toy" would

have introduced a linguistic interpretation task into the Belief Attribution task (see Robinson & Mitchell, 1992, 1994; and 'General discussion').

At the end of each video, the experimenter presented children with the same two novel objects seen in the video. The objects were placed on the table in front of the children. In the Label task, the experimenter asked children to point to the newly named object ("Which one is the modi?"). In the Belief Attribution task, children were asked to identify which object the main character thought was in the box ("Which one does the wolf think is in the box?"). If a child avoided choosing in either task, the experimenter asked the child to hand over the target object rather than point. The need for this occurred very rarely.

The experimenter went on to ask the children two Memory questions ("Which toy did the wolf put in the box? Which toy did the monkey put in the box?"). The Memory questions served as a control to ensure that children encoded the event correctly (see also Baldwin, 1993; Happé & Loth, 2002). Children answered both of these questions within a trial correctly 84% of the time; those trials were entered into subsequent analyses. Finally, the experimenter asked a Preference question ("Which toy is your favorite?"). The Preference question was included to ensure that children's responses were not driven by a strong preference for one object over the other (following Baldwin, 1993; Happé & Loth, 2002). Responses revealed no reliable preference for one over the other toy within a trial. Furthermore, participants did not choose their preferred objects during the False Belief test trials at levels significantly different from chance ($M = 47\%$, $p > .05$).

Materials also included two True Belief pretest stories (see 'Appendix' for an example). One of these stories involved a Label and the other a Belief Attribution task. The stories, script, and questions in the pretests (including the Perception, Memory, and Preference control questions) were identical to the False Belief test items except that the character was present while the objects were switched. Overall, participants understood that, when the character was present, he/she could see the object substitution, as evidenced by the fact that 97% of responses in the Perception question were correct. Children also remembered the prior event, as shown by the fact that they gave correct responses to both Memory questions 84% of the time. As in the main phase of the experiment, participants in the True Belief trials did not show a consistent bias for one or the other object within a trial in the Preference question, nor did they consistently pick preferred objects in answering True Belief pretest questions ($M = 41\%$, $p > 05$).

Each story used different characters, objects, and labels (for tests: *modi, blicket*; for pretests: *zavin, gogi*). Half of the main characters were male and half female. The characters' statements were prerecorded with a

distinct speaker for each character. The direction (left–right) the character moved in when entering/exiting, and the placement (left–right) of the correct object on the table were counterbalanced across trials. Test stories for each task were administered in blocks, and both block order and story order within blocks were counterbalanced. Pretest story order was fixed, with Label stories administered first. Across subjects, assignment of each test and pretest story to task was counterbalanced. No subject saw versions of the same story across both the Belief Attribution and Label tasks.

RESULTS AND DISCUSSION

True belief pretests

We excluded pretest trials where children failed Memory questions (16% of the data). To ensure that our participants were engaged and understood the task, we also excluded from subsequent analyses children that passed no True Belief pretest trials despite answering the corresponding Memory questions correctly (10 three-year-olds, 13 four-year-olds, and 3 five-year-olds). The total remaining sample thus consisted of 112 participants. Of these, 54% passed all True Belief trials for which the Memory questions were answered correctly (16 three-year-olds, 25 four-year-olds, and 19 five-year-olds). The remaining 46% of participants passed one of the two pretest trials for which they had answered Memory questions correctly (12 three-year-olds, 23 four-year-olds, and 17 five-year-olds). When children succeeded with only one of the two pretests, they were equally likely to succeed on the Label compared to the Belief Attribution pretest.

It might appear puzzling that a substantial number of children had difficulties with True Belief trials. Nevertheless, similar difficulties in True Belief word-learning tests have been reported in both Happé and Loth's (2002) study and across all three experiments in Houston-Price *et al*. (2011) (these studies did not include True Belief Attribution scenarios). The difficulties we (and others) observed in the pretests may have been due in part to the nature of True Belief stories, where the absence of deception could have confused children or made them less interested in the story (see Sullivan & Winner, 1993; Wellman *et al*., 2001, for evidence that children tend to prefer stories that feature deception). Moreover, the need for children to track the actions of two characters and answer our Perception and Memory questions may have increased the cognitive load and led to poorer performance (see also Houston-Price *et al*., 2011).

False belief test trials

We excluded 16% of the trials because of failures in the Memory controls. We defined passing each False Belief task as answering correctly all remaining trials within the task (i.e. trials on which Memory questions

TABLE 1. *Number of 3-, 4- and 5-year-old children passing and failing the False Belief trials in the Label and Belief Attribution task*

		Label Task							
		3s (n = 28)		4s (n = 48)		5s (n = 36)		Total (n = 112)	
		Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail
Belief Attribution Task	Pass	6	5	22	9	26	5	54	19
	Fail	4	13	7	10	4	1	15	24

had been correct). Since children in the majority of cases passed the Memory questions in both trials of each task, this meant that they had to answer both trials within a task to be considered passers. Otherwise, children were considered failers. The number of children within each age group that passed or failed each task is given in Table 1 and will be used for all analyses that follow.

Label task. Beginning with the main point of interest, performance on the Label task changed with age. Only about a third (10/28) of the three-year-olds passed the task, but the vast majority of five-year-olds did so (30/36), with four-year-olds' performance being intermediate (29/48 passed). A chi-square test confirmed that there was an association between age and whether participants passed or failed ($\chi^2 = 18{\cdot}42$, df = 2, $p < {\cdot}001$). A further series of chi-square tests on 2 × 2 contingency tables containing the numbers of passers and failers for pairs of age groups showed that age was associated with performance in the analysis of three- vs. five-year-olds ($\chi^2 = 15{\cdot}24$, df = 1, $p < {\cdot}001$), three- vs. four-year-olds ($\chi^2 = 4{\cdot}32$, df = 1, $p = {\cdot}038$), and four- vs. five-year olds ($\chi^2 = 5{\cdot}17$, df = 1, $p = {\cdot}023$).

Moreover, three-year-olds had equal numbers of passers and failers in the Label task ($n = 10$ vs. 18, $p = {\cdot}185$, two-tailed binomial, n.s.), as did four-year-olds ($n = 29$ vs. 19, $p = {\cdot}193$, n.s), whereas five-year-olds had significantly more passers than failers ($n = 30$ vs. 6, $p < {\cdot}001$, two-tailed binomial).

Belief attribution task. As Table 1 shows, results from the Belief Attribution task are consistent with prior literature on theory-of-mind development. Only a third (11/28) of the three-year-olds passed the Belief Attribution task, but the vast majority of five-year-olds (31/36) did so, with four-year-olds' performance being intermediate (31/48 passed). A chi-square test confirmed that there was an association between age and whether participants passed or failed ($\chi^2 = 18{\cdot}42$, df = 2, $p < {\cdot}001$). A further series of chi-square tests on 2 × 2 contingency tables containing the numbers of passers and failers for pairs of age groups showed that, as with

the Label task, age was associated with performance in all comparisons: three- vs. five-year-olds ($\chi^2 = 15{\cdot}31$, df = 1, $p < {\cdot}001$), four- vs. five-year-olds ($\chi^2 = 4{\cdot}93$, df = 1, $p = {\cdot}026$), and three- vs. four-year-olds ($\chi^2 = 4{\cdot}58$, df = 1, $p = {\cdot}032$).

Moreover, as expected, the extent to which children could reliably pass the false Belief Attribution task differed by age: in both three-year-olds and four-year-olds, the number of passers did not differ from the number of failers (3s: 11 vs. 17, respectively, $p = {\cdot}345$; 4s: 31 vs. 17, $p = {\cdot}060$, two-tailed binomial, n.s.), but in five-year-olds, the number of passers was reliably higher than the number of failers (5s: 31 vs. 5, $p < {\cdot}001$, two-tailed binomial).

We considered the possibility that children's choices in the False Belief test trials could have been based on superficial associations between a character and the object he/she placed in the box: if so, there would be more successes in False Belief compared to True Belief trials. For both pretests and tests, we defined passing as answering correctly all trials on which Memory questions had been correct. The number of children passing pretests was either marginally greater than those passing tests ($n = 16$ vs. 6 in three-year-olds, $p = {\cdot}053$) or very close ($n = 25$ vs. 22 in four-year-olds, $p = {\cdot}771$, n.s.; and $n = 19$ vs. 26 in five-year-olds, $p = {\cdot}371$, n.s.). Thus this possibility is not supported by our data (see also Houston-Price *et al*., 2011, for a similar pattern and conclusion).

Comparison between the label and belief attribution tasks. We compared performance on the Label and Belief Attribution task using the number of children who passed only one of the two tasks as an index of task difficulty. As in Happé and Loth (2002), we conducted McNemar's tests with Yates' continuity correction looking at the number of children across all age groups that passed only one task ($n = 34$). The test revealed no difference between Label-Only and Belief-Attribution-Only passers (15 vs. 19, respectively, $\chi^2 = 0{\cdot}265$, df = 1, $p = {\cdot}607$, n.s.).

We repeated this analysis, focusing only on the subset of children who passed all pretest trials for which they had answered Memory trials correctly and could thus be argued to have the strongest understanding of our task ($n = 60$). Of these children, 27 passed both False Belief tasks and 20 failed both. Again, Label-Only passers were no different from Belief-Attribution-Only passers (5 vs. 8, $\chi^2 = 0{\cdot}308$, df = 1, $p = {\cdot}579$).

To probe further into the relative difficulty of the Belief Attribution and Label tasks, we asked whether the order in which the tasks were received affected children's performance. If reasoning about false labels were easier than reasoning about false beliefs, then performance on the Belief Attribution task might be facilitated in those children who received the Label task first compared to children who received the Label task last; however, performance on the Label task itself should not be facilitated

(and might even be hurt) by prior exposure to the Belief Attribution task. Closer inspection of the data revealed that task order was unrelated to task performance. Thirty-eight out of the 57 children in the final sample (or 67%) who received the Belief Attribution task first passed the task and 35 out of 55 children (or 64%) who received the same task last passed it. Similarly, 33 out of the 55 children (or 60%) who received the Label task first passed the task and 36 out of 57 children (or 63%) who received the same task last passed it. Chi-square analyses confirmed that there was no association between task order and performance on either the Belief Attribution ($\chi^2 = 0{\cdot}11$, df = 1, $p = {\cdot}737$, n.s.) or the Label task ($\chi^2 = 0{\cdot}12$, df = 1, $p = {\cdot}731$, n.s.).

GENERAL DISCUSSION

Our study led to two main results. First, young children consult the epistemic state of a speaker during word learning, and this ability develops between the ages of three and five. Specifically, our data show that five-year-olds successfully acquired a word from a speaker with a false belief by accurately tracing the speaker's referential intent. As shown in Table 1, these children avoided associating a novel word with a novel object inside a box when the speaker had a mistaken belief about what was in the box: the children instead attached the word to the object that accurately corresponded to the speaker's epistemic state. Some three- and four-year-olds also succeeded in these contexts: even though, as a group, these children were not above chance in the Label task, individual data show that about half of the three- and four-year-olds actually passed the task. This finding settles a disagreement in prior reports on children's ability to handle accidentally false labels (Carpenter *et al.*, 2002; Happé & Loth, 2002; Houston-Price *et al.*, 2011) and clarifies the developmental trajectory of this ability (see also Mitchell, Robinson & Thompson, 1999).

The present study adds to the literature documenting the extent to which preschoolers track a speaker's mental state (e.g. whether the speaker is knowledgeable vs. uninformed, or has limited vs. full knowledge) when acquiring information, including novel words, from that speaker (Sabbagh & Baldwin, 2001; Koenig & Harris, 2005; Jaswal & Neely, 2006; Saylor & Troseth, 2006; Birch *et al.*, 2008; Nurmsoo & Bloom, 2008; Rakoczy, Warneken & Tomasello, 2009; Robinson & Nurmsoo, 2009; Bannard & Tomasello, 2012). Our data are consistent with a broadly pragmatic perspective on word learning emerging from this and related literature: according to this perspective, children do not simply associate novel words with the objects contextually tied to the act of naming but use information about the speaker's mental state to attribute reference to newly encountered words.

A second result from this work is that young children's ability to recognize that someone can err because of a false belief, and that someone can offer an erroneous label because of a false belief, develops around the same time (and in accordance with the well-documented transition in false belief reasoning in Wimmer & Perner, 1983; Baron-Cohen *et al*., 1985; Wellman *et al*., 2001; among many others). This result differs from earlier studies (Carpenter *et al*., 2002; Happé & Loth, 2002) that had shown an advantage in mental-state reasoning for communicative over non-communicative contexts: in Happé and Loth's (2002) study, for instance, 88% of the children who passed only one task passed the Label task, whereas in the present data only 58% of children did so. Notice that, compared to those previous studies, the present design included a higher overall number of participants ($n = 112$, compared to less than 50 in the final analyses of H&L) and a higher number of children who passed only one task ($n = 34$, compared to $n = 25$ in H&L). Our design also includes a higher number of participants compared to each of the first two experiments in Houston-Price *et al*. (2011) that also showed the asymmetry. The present data do not support a difference in false belief performance between communicative and non-communicative contexts, nor do they offer a basis for treating these as products of distinct mechanisms, as proposed by Happé and Loth (2002) and Carpenter *et al*. (2002).

Our study is close in spirit to more recent proposals that attributed the divergence in children's false belief performance between word-learning and action-prediction contexts to task parameters (Houston-Price *et al*., 2011; note that our own methods compare word-learning to belief-attribution, not action-prediction, but the latter two tasks have been shown to be equivalent; Wellman *et al*., 2001). Our approach diverges, however, from these earlier proposals in two important respects. First, our study is the first one to date to use truly balanced tasks (in fact, the very same task) to compare children's false belief reasoning across word learning and belief attribution domains. Unlike Houston-Price *et al*. (2011), our design makes no additional assumptions about how false belief is calculated within each domain and is therefore not subject to issues of interpretation arising from the prior paradigm (see 'Introduction'). Therefore, as far as we can tell, the present data offer the first unambiguous evidence so far for the role of methodological imbalances in what appeared to be an advantage of communicative over non-communicative false belief contexts in prior work.

Second, and relatedly, our study offers no support for the role of the specific task parameters identified in prior work as being responsible for children's performance differences. Recall that, according to Houston-Price and colleagues (2011), the advantage of communicative contexts was due to the higher 'pull of the real' exerted by the child's knowledge of the correct response in traditional false belief contexts compared to contexts that asked about the interpretation of a newly introduced word (see also

Perner *et al*., 2007). In our own study, however, other things being equal, the 'pull of the real' (e.g. the child's knowledge of the actual contents of a box following a surreptitious content-change) created no selective deficit in children's performance in the Belief Attribution task. This outcome supports the conclusion that both the Belief Attribution task and the Word Learning task in prior work (as well as our own) involve suppressing the incorrect answer (and that the pattern of results in Houston-Price *et al*., 2011, Exp. 3, rather than demonstrating the role of a 'reality bias', probably resulted from the selective simplification of the Action Prediction task that removed the need to inhibit the incorrect answer).

Our findings are reminiscent of work that examined the relationship between children's false belief understanding and reference assignment for (known) noun phrases. Robinson and Mitchell (1992; see also Robinson & Mitchell, 1994) presented children with scenarios in which a character asked for "the bag in the red drawer" in a situation in which (unbeknownst to the character) the bag had been switched with the bag in the blue drawer. Many children as young as three and four years could correctly pick the bag that the character "really wanted" (children were also accurate in a 'true belief' control task, in which the bags were removed but then returned to their original locations). Furthermore, children appeared to be better in this referential task compared to a standard false belief task involving change of location. However, further experimentation – including tasks that were more carefully matched across the referential and standard false-belief versions (Robinson & Mitchell, 1994) – failed to replicate the asymmetry. Much as in the present experiment, Robinson and Mitchell concluded that both types of false belief task were actually equally difficult for children.

The present results raise several further questions. A first question is what is responsible for the youngest children's failures in mental-state reasoning, and (relatedly) whether different contexts might lead very young children to pass the false label task. Notice that, unlike prior studies in which a speaker's lack of knowledge about the referent of a novel word was established through explicit cues (e.g. "I don't know what a blicket is. Maybe it's this one." Sabbagh & Baldwin, 2001) or past unreliable naming episodes (Koenig & Harris, 2005), in the present study children had to infer a character's false belief from tracking that character's perspective. Furthermore, perspective-tracking was assessed through standard, rather protracted scenarios in which children observed interactions between characters and objects but did not participate in the interactions themselves. This line of reasoning leaves open the possibility that in simpler, more participatory contexts, even younger children might succeed in reading the mind of the speaker (see Nurmsoo & Bloom, 2008; Southgate, Chevallier & Csibra, 2010, for relevant evidence; cf. also Warneken & Tomasello, 2007; Csibra & Gergely, 2009, for discussion).

A second question raised by the present data is whether preschoolers' ability to engage in mental state reasoning might extend to situations beyond word learning. A prominent case in point is the calculation of pragmatic inference, where adults are known to 'take the epistemic step' (Bergen & Grodner, 2012; Breheny, Ferguson & Katsos, 2013) but preschoolers have been reported to face difficulties (e.g. Noveck, 2001; Papafragou & Musolino, 2003; Huang & Snedeker, 2009). This issue bears directly on the current debate about children's use and limits of theory of mind computations during communication more broadly (see, e.g. Diesendruck & Markson, 2001; Epley *et al*., 2004; Breheny, 2006; de Marchena *et al*., 2011) and is ripe for further research.

REFERENCES

Baldwin, D. A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology* **29**, 832–43.
Bannard, C. & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLoS ONE* **7**, e, 49881.
Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985). Does the autistic child have a 'theory of mind?' *Cognition* **21**, 37–46.
Bergen, L. & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **38**, 1450–60.
Birch, S. A. J., Vauthier, S. A. & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition* **107**, 1018–34.
Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.
Breheny, R. (2006). Communication and folk psychology. *Mind and Language* **21**, 74–107.
Breheny, R., Ferguson, H. & Katsos, N. (2013). Taking the epistemic step: toward a model of on-line access to conversational implicatures. *Cognition* **126**, 423–40.
Buttelmann, D., Carpenter, M. & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* **112**, 337–42.
Carpenter, M., Call, J. & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology* **20**, 393–420.
Chandler, M., Fritz, A. S. & Hala, S. (1989). Small scale deceit: deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind. *Child Development* **60**, 1263–77.
Clements, W. & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development* **9**, 377–97.
Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences* **13**(4), 148–53.
de Marchena, A., Eigsti, I. M., Worek, A., Ono, K. E. & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: testing the pragmatic hypothesis. *Cognition* **119**, 96–113.
Diesendruck, G. & Markson, L. (2001). Children's avoidance of lexical overlap: a pragmatic account. *Developmental Psychology* **37**, 630–41.
Epley, N., Morewedge, C. K. & Keysar, B. (2004). Perspective taking in children and adults: equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology* **40**, 760–8.
Garnham, W. A. & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science* **4**, 94–100.
Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
Happé, F. & Loth, E. (2002). 'Theory of mind' and tracking speakers' intentions. *Mind and Language* **17**, 24–36.

Houston-Price, C., Goddard, K., Séclier, C., Grant, S., Reid, C., Boyden, L. & Williams, R. (2011). Tracking speakers' false beliefs: Is theory of mind available earlier for word learning? *Developmental Science* **14**, 623–34.
Huang, Y. T. & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in five-year olds: evidence from real time spoken language comprehension. *Developmental Psychology* **45**, 1723–39.
Jaswal, V. K. & Neely, L. A. (2006). Adults don't always know best: preschoolers use past reliability over age when learning new words. *Psychological Science* **17**, 757–8.
Koenig, M. A. & Echols, C. H. (2003). Infants' understanding of false labeling events: the referential role of words and the speakers who use them. *Cognition* **87**, 179–208.
Koenig, M. A. & Harris, P. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development* **76**, 1261–77.
Koos, O., Gergeley, G., Csibra, G. & Biro, S. (1997). Why eating Smarties makes you smart: understanding false belief at the age of 3. Poster presented at the Meeting of the Society for Research in Child Development, Washington, DC, April.
Mitchell, P., Robinson, E. J. & Thompson, D. E. (1999). Children's understanding that utterances emanate from minds: using speaker belief to aid interpretation. *Cognition* **72**, 45–66.
Nadig, A. S. & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science* **13**, 329–36.
Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* **78**, 165–88.
Nurmsoo, E. & Bloom, P. (2008). Preschoolers' perspective taking in Label: Do they blindly follow eye gaze? *Psychological Science* **19**, 211–5.
Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* **308**, 255–8.
Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* **86**, 253–82.
Perner, J., Rendi, B. & Garnham, A. (2007). Objects of desire, thought, and reality: problems of anchoring discourse referents in development. *Mind and Language* **22**, 475–513.
Rakoczy, H., Warneken, F. & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development* **24**, 61–9.
Robinson, E. J. & Mitchell, P. (1992). Children's interpretation of messages from a speaker with a false belief. *Child Development* **62**, 639–52.
Robinson, E. J. & Mitchell, P. (1994). Young children's false belief reasoning: interpretation of messages is no easier than the classic task. *Developmental Psychology* **30**, 67–72.
Robinson, E. J. & Nurmsoo, E. (2009). When do children learn from unreliable speakers? *Cognitive Development* **24**, 16–22.
Roth, D. & Leslie, A. (1991). The recognition of the attitude conveyed by an utterance: a study of preschool and autistic children. *British Journal of Developmental Psychology* **9**, 315–30.
Rubio-Fernández, P. & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science* **24**, 27–33.
Sabbagh, M. A. & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. *Child Development* **72**, 1054–70.
Saylor, M. M. & Troseth, G. L. (2006). Preschoolers use information about speakers' desires to learn new words. *Cognitive Development* **21**, 214–31.
Southgate, V., Chevallier, C. & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science* **16**, 907–12.
Southgate, V., Senju, A. & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year olds. *Psychological Science* **18**, 587–92.
Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (ed.), *Metarepresentations: a multi-disciplinary perspective* (pp. 116–37). New York: Oxford University Press.

Sullivan, D. & Winner, E. (1993). Three-year-olds' understanding of mental states: the influence of trickery. *Journal of Experimental Child Psychology* **56**, 135–48.
Surian, L., Caldi, S. & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science* **18**, 580–6.
Warneken, F. & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy* **11**, 271–94.
Wellman, H. M., Cross, D. & Watson, J. (2001). A meta-analysis of theory of mind development: the truth about false belief. *Child Development* **72**, 655–84.
Wilson, D. (2000). Metarepresentations in linguistic communication. In D. Sperber (ed.), *Metarepresentations: a multi-disciplinary perspective* (pp. 411–48). New York: Oxford University Press.
Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**, 103–28.

APPENDIX: EXPERIMENTAL PROCEDURE

A. FALSE BELIEF TEST TRIAL

Wolf enters the scene holding an object and says: "*Hi, I'm Lucy! Look what I have! Let's put it away now. Let's put it in this box to keep it safe!*"

Wolf puts object into a box, then says: "*I'm tired now so I'm going to take a nap. Will you watch this while I'm asleep? Keep it safe while I take a nap.*" Lucy leaves.

Monkey enters the scene holding a different object and says: "*Hi, I'm David! Look what I have! Let's play a trick on Lucy! Ready? Let's be very quiet so we don't wake Lucy! I take this one out of the box and put this one in!*" Monkey takes out wolf's toy, places it in view as he puts his own toy in the box and closes the lid.

E: "Where did the wolf go?"
E: "Can the wolf see what the monkey is doing?" (Perception Question)

Monkey: "*I'm going to go away now*". Monkey leaves and takes wolf's toy with him. Wolf returns to the scene and approaches the closed box.

BELIEF ATTRIBUTION TASK	LABEL TASK
---	Wolf*: "It's time to play with the blicket! Remember the blicket? There's a blicket in this box."*

The two objects are presented to the child.

BELIEF ATTRIBUTION TASK	LABEL TASK
E: "Which one does the wolf think is inside the box?" (Target Question)	*E: "Which one is the blicket?"* (Target Question)

E: "Which toy did the wolf put in the box? Which toy did the monkey put in the box?" (Memory Questions)
E: "Which toy is your favorite?" (Preference Question)

B. TRUE BELIEF PRETEST TRIAL

Cat enters the scene holding an object and says "*Hi, I'm Jane! Look what I have! Let's put it away now. Let's put it in this box to keep it safe!*"

Cat puts object into a box, then says*: "I'm going to go outside and play. Will you watch this while I am outside? Keep it safe while I'm playing.*" Cat leaves.

Frog enters the scene holding a different object and says: "*Hi, I'm Larry! Look what I have! Let's get Jane so she can see this! Let's call Jane. Jane, come back and see what I've got!*"

Cat enters.

Frog: "*I take this one out of the box and put this one in*!" Frog takes out cat's toy, places it in view as he puts his own toy in the box and closes the lid. Cat watches.

E: "Can the cat see what the frog is doing?" (Perception Question)

Frog leaves and takes cat's toy with him.

BELIEF ATTRIBUTION TASK	LABEL TASK
---	*Cat: "It's time to play with the zavin! Let's play with the zavin! There's a zavin in this box!*"

The two objects are presented to the child.

BELIEF ATTRIBUTION TASK	LABEL TASK
E: "Which one does the cat think is inside the box?" (Target Question)	*E: "Which one is the zavin?"* (Target Question)

E: "Which toy did the cat put in the box? Which toy did the frog put in the box?" (Memory Questions)

E: "Which toy is your favorite?" (Preference Question)