CHILD DEVELOPMENT

Child Development, September/October 2018, Volume 89, Number 5, Pages 1642–1656

The Role of Speaker Knowledge in Children's Pragmatic Inferences

Anna Papafragou
University of Delaware

Carlyn Friedberg
Lexia Learning

Matthew L. Cohen
University of Delaware

During communication, conversational partners should offer as much information as is required and relevant. For instance, the statement "Some Xs Y" is infelicitous if one knows that all Xs Y. Do children understand the link between speaker knowledge and utterance strength? In Experiment 1, 5-year-olds (N = 32) but not 4-year-olds (N = 32) reliably connected statements of different logical strength (e.g., "The girl colored all/some of the star") to observers who were fully or partially informed. Four-year-olds' performance improved when observer knowledge could be assessed more easily (Experiment 2a, N = 25) but remained the same in a non-linguistic version of Experiment 1 that preserved the epistemic requirements of the original study (Experiment 2b, N = 26). These findings have implications for the development of early communicative abilities.

A foundational principle of human communication is that speakers should offer as much information as is required and relevant during conversation (Grice, 1989). When a speaker fails to be as informative as is needed, the hearer is generally justified in assuming that this is because of lack of knowledge on the speaker's part. Consider a statement such as "Some dinosaurs had long tails." Assuming that the speaker could have used a stronger, more informative term (*all*), and because it would have been relevant to use *all* if it were true, the hearer is entitled to infer that the speaker is not in a position to offer a statement containing *all*—either because he or she does not know for a fact that such a statement is true or (if the speaker is assumed to have relevant knowledge about dinosaurs) because he or she knows that it is *not* true (Franke, 2011; Geurts, 2011; Horn, 1972, 1989; Sauerland, 2004; Spector, 2006; van Rooij & Schulz, 2004). The inference from the use of a logically weaker term to the conclusion that the speaker is epistemically uncommitted to/rejects a stronger term is known as a *scalar implicature* (for discussion and different perspectives, see Carston, 1998; Chierchia, 2004; Fox & Spector, 2018; Gazdar, 1979; Grice, 1989; Horn, 1972; Levinson, 2000; Noveck & Sperber, 2007; Sauerland, 2004, 2012; Sperber & Wilson, 1986/1995; van Rooij & Schulz, 2004).

Scalar implicatures take their name from the fact that quantifiers (*some*, *all*) and other logical terms fall on a scale defined in terms of logical entailment (Horn, 1972). For instance, in the quantifier scale <*all*, *most*, . . . *some*>, *All Xs Y* entails *Some Xs Y* (but not vice versa), so the former statement is logically stronger/more informative than the latter. The notion of strength has also been applied to context-based scales that are based simply on part–whole or temporal relations (Hirschberg, 1985): In this sense, *X and Y are married* is stronger/more informative than *X and Y are engaged*. In both logical and ad hoc/contextual scales, the weak scalar item is semantically compatible with the strong scalar item but can be used to communicate that (for all the speaker knows) the stronger item is excluded.

There is evidence that adults' expectation that the speaker should be informative (to the degree required by relevance) leads them to regularly compute scalar implicatures when this pragmatic expectation appears to be violated in conversation (Bott

We acknowledge support from Grant #R01-HD055498 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to Anna Papafragou, as well as support from Undergraduate Science and Engineering Scholar grants from the University of Delaware to both Carlyn Friedberg and Matthew L. Cohen. We thank Paul Quinn for his helpful comments, and Myrto Grigoroglou, Sarah Fairchild, Kyle Schneck, Dimitris Skordos, Megan Johanson, Solveig Bosse, Sarah Lorch, Taraneh Mojaverian, and Ozge Ozturk for their time and contributions. The bulk of the data for this article were collected between August 2013 and May 2015.

Correspondence concerning this article should be addressed to Anna Papafragou, Department of Psychological and Brain Sciences, University of Delaware, Newark, DE 19716. Electronic mail may be sent to apapafragou@psych.udel.edu.

 0009-3920/2018/8905-0016
DOI: 10.1111/cdev.12841

& Noveck, 2004; Breheny, Katsos, & Williams, 2006; Grodner, Klein, Carbary, & Tanenhaus, 2010; cf. Degen & Tanenhaus, 2015). Young children, however, seem to have difficulties with these subtle aspects of communicated meaning. Noveck (2001) found that 7- to 9-year-old children are more likely than adults to accept underinformative statements such as "Some birds have wings," presumably because they fail to generate the implicature that (the speaker knows that) *not all birds have wings* (cf. also Braine & Rumain, 1981; Smith, 1980). A related set of studies discovered that, in scenarios which made a statement containing a strong logical term true (e.g., "Every boy chose a skateboard *and* a bike"), 5-year-old children—unlike adults—failed to reject a statement containing a weaker term (e.g., "Every boy chose a skateboard *or* a bike"; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001). More recently, studies monitoring listeners' eye movements have shown that young children, unlike adults, fail to exclude *all* scenes from consideration when they hear utterances with the quantifier *some* (Huang & Snedeker, 2009).

Other work has suggested that children's pragmatic performance improves under certain conditions. Papafragou and Musolino (2003, experiment 2) found that children are more likely to disagree with a logically true but underinformative statement if task demands clearly set up the expectation of a stronger statement (cf. also Foppolo, Guasti, & Chierchia, 2012; Guasti et al., 2005; Skordos & Papafragou, 2016). Moreover, children are more likely to infer that weak statements exclude a stronger alternative in tasks that do not involve explicit, binary acceptability judgments or in tasks that make scalar alternatives more easily accessible (Barner, Brooks, & Bale, 2011; Katsos & Bishop, 2011; Papafragou, 2006; Papafragou & Skordos, 2016; Papafragou & Tantalou, 2004; Pouscoulous, Noveck, Politzer, & Bastide, 2007; Stiller, Goodman, & Frank, 2015). Furthermore, 5-year-olds seem to distinguish between natural language expressions in terms of their informational strength: When given a choice between a weak (e.g., *or*) and a strong (e.g., *and*) logical term to describe the outcome of a story, children prefer the strong, more informative one when they know it to be true (Chierchia et al., 2001; cf. also Katsos & Bishop, 2011; Ozturk & Papafragou, 2015).

An issue that has not been explored systematically is whether children understand the link between the strength of an utterance and the speaker's state of knowledge—specifically, the link between more versus less informative statements and more versus less knowledgeable speakers. In several prior studies, children's sensitivity to pragmatic principles was measured by the ability to reject underinformative utterances offered by a (presumably fully knowledgeable) experimenter (Guasti et al., 2005; Noveck, 2001) or by a speaker known to be incompetent (e.g., a "silly puppet"; Katsos & Bishop, 2011; Papafragou & Musolino, 2003). For instance, in one study (Papafragou & Musolino, 2003, experiment 1), children were presented with an acted out story in which all three horses in the scene jumped over a fence and then heard a silly puppet say, "I know what happened! Some of the horses jumped over the fence." The idea was that children should be able to recognize that use of a weak utterance is inappropriate if the speaker knows that a stronger statement is true, because the weak statement falsely implies lack of more definitive speaker knowledge; nevertheless, children rarely rejected the puppet's underinformative statements. However, it is possible that children would be more successful in inferring lack of speaker knowledge from an underinformative statement in situations where the statement were *genuinely* used to convey this inference (i.e., in situations in which a puppet using a statement with *some* truly intended to communicate that *all* does not hold). After all, in actual conversations, scalar implicatures are part of what the speaker intends to communicate with the use of a weak assertion (see Papafragou & Tantalou, 2004).

The issue of whether children consult speaker knowledge when computing scalar inferences has great theoretical interest for accounts of conversational implicature. Under broadly Gricean accounts of scalar implicature of the sort we sketched earlier, the hearer needs to engage in a rich computational process that calculates the speaker's perspective and what else the speaker might have said given his or her goals and knowledge state (see Carston, 1998; Geurts, 2011; Horn, 1972; Sauerland, 2004, 2012; Sperber & Wilson, 1986/1995). Other types of accounts, however, propose that scalar implicatures are often computed as part of the conventional content of an utterance and might be recoverable on the basis of lexical–grammatical mechanisms alone (e.g., stored lexical scales such as <*all, some*>) without appeal to speaker knowledge (Chierchia, 2004; Chierchia, Fox, & Spector, 2012; Levinson, 2000). For adults, experimental evidence argues against such lean accounts of implicature computation and in favor of rich, inferential accounts, because speaker perspective seems to affect the earliest stages of pragmatic processing (Bergen & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013). For children, however, this issue remains open.

The question whether children are sensitive to speaker knowledge when computing pragmatic inferences connects to broader debates about whether children engage in rich, Gricean types of pragmatic reasoning during early communication. In the domain of word learning, some authors have claimed that children's early-emerging ability to make use of lexical contrast/mutual exclusivity in acquiring the meaning of novel words has a characteristically Gricean signature (Clark, 1990; Diesendruck & Markson, 2001), whereas other authors have challenged this proposal (de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Preissler & Carey, 2005; Regier, 2005). Relatedly, there is considerable disagreement about the extent to which children take into account the mental perspective of others in communication: Some authors have argued that children have difficulty evaluating and editing communicative content (Conti & Camras, 1984; Glucksberg, Krauss, & Higgins, 1975) and others have offered evidence for children's flexible adjustments to other people's perspectives in both speech production and comprehension (Matthews, Lieven, Theakston, & Tomasello, 2006; Nadig & Sedivy, 2002; O'Neill, 1996; Shatz & Gelman, 1973), with some sensitivity to the mental states of one's interlocutors present even in infants (Baldwin, 1991; Liszkowski, Carpenter, & Tomasello, 2008; Southgate, Chevallier, & Csibra, 2010).

In this article, we wish to contribute to these debates by examining whether children can use information about a speaker's knowledge to attribute a weak or strong scalar utterance to that speaker, in accordance with Grice's principles. Specifically, we ask whether 4- and 5-year-old children can link speakers with differential access to knowledge (full/updated vs. partial/outdated) to statements with different levels of informational strength (strong vs. weak). Unlike previous experiments on scalar implicature, weak scalar statements in our scenarios are motivated by lack of speaker knowledge (e.g., speakers use *some* when they are not in a position to assert *all*, thereby conveying a scalar implicature). If our studies provide evidence that children consult the speaker's epistemic state in interpreting a weak, implicature-carrying scalar statement, they would support a rich, mentalistic model of implicature computation and argue against the view that children are insensitive to speaker perspective in their communicative interactions.

The choice of age groups was based on the fact that the ability to track an agent's mental state has been argued to develop between 4 and 5 years of age. Many tests of the ability to correctly report an agent's false belief (and hence to accurately reason about objects of belief as distinct from reality) have shown that 5-year-olds but only some 4-year-olds understand that people can hold false beliefs if access to updated information is blocked. For instance, in a classic false belief task (Baron-Cohen, Leslie, & Frith, 1985; cf. Wimmer & Perner, 1983), 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?" Typically, 5-year-olds picked the correct answer, indicating the location that Sally would have in mind (i.e., the basket), whereas 4-year-olds' performed more variably, sometimes choosing instead the last location of the ball (and 3-year-olds failed completely; for a review of such tasks, see Wellman, Cross, & Watson, 2001; but see Rubio-Fernández & Geurts, 2013, among others). Given this background, it was of particular interest to see whether there would be differences between 4- and 5-year-olds' ability to consult speaker knowledge in deriving scalar inferences. Existing tasks in which scalar implicatures were not clearly motivated by lack of speaker knowledge (e.g., Barner et al., 2011; Katsos & Bishop, 2011; Noveck, 2001; Papafragou & Musolino, 2003) have shown that the ability to derive implicatures is still developing in 5-year-olds.

Experiment 1

In Experiment 1, children were asked to match scalar statements to speakers who had different levels of information. Children saw two tokens of the same event (e.g., coloring a star) performed by identical-looking agents ("twins"). For each token of the event, the agent was watched by an observer: In one case, the observer had full access to the whole event, but in the other, the observer fell asleep halfway through the event. Children were then presented with either a strong or a weak scalar statement and had to match it to the right observer. To perform correctly in this task, children had to take into account what each observer knew about the corresponding token event, as from the children's own perspective both tokens of the event were identical (e.g., in both, a star was colored). Thus, to pass the task, children had to engage in genuine mental state tracking (see the false belief paradigms reviewed earlier for a similar logic).

Once the observers' perspective was adopted, strong statements could be unambiguously matched to the fully informed observer because of the requirement that speakers should not say something they lack adequate evidence for (cf. Grice's quality maxim; Grice, 1989). But taking the observers' stance (and adopting the quality maxim) was not enough for attributing weak statements unequivocally because, from a strictly logical perspective, such statements could belong to either observer. Participants needed to determine on pragmatic grounds that such statements would violate Grice's quantity maxim by being underinformative if uttered by the fully informed observer but could be used by the partially informed observer to convey lack of knowledge of the complete event (through a scalar implicature).

We included two types of scale: a traditional quantificational scale that was defined through logical entailment (<*all, some*>; quantity task) and nonlogical, ad hoc scales that relied on some sort of part–whole structure set up by context (transformation task). Some developmental studies have reported stronger performance with logical compared to ad hoc scales (Barner et al., 2011; Stiller et al., 2015), but other studies have found no difference between scale types (Katsos & Bishop, 2011; Papafragou & Tantalou, 2004), so the issue offers itself for further investigation.

Method

Participants

Thirty-two 4-year-olds (M_{age} = 4;8, range = 4;2–4;11; 17 females and thirty-two 5-year-olds (M_{age} = 5;5, range = 5;0–5;11; 16 female) participated. Children were recruited from Newark, Delaware preschools. We also tested a control group of 12 adults (8 females) that were undergraduate students at the University of Delaware attending Introductory Psychology classes. Adults were 97% correct, and their results will not be discussed further.

Materials and Procedure

Participants saw eight pairs of videos depicting events (see Table 1). The two videos in each pair were displayed sequentially in two windows placed side by side on a computer screen. Before the videos began, participants saw the initial (frozen) scene of each video. Each scene showed two girls and an identical object (e.g., a star). One of the girls was in fact the same in both scenes, differing only in the color of her T-shirt. The experimenter pointed to the two identical girls across the two scenes and said,

Table 1
Stimulus List for Experiment 1

Quantity task	Transformation task
1. Coloring some/all of the star	1. Drawing a cloud/an ice cream
2. Drinking some/all of the juice	2. Drawing a banana/a face
3. Eating some/all of the cookie	3. Drawing a beach ball/a pumpkin
4. Filling some/all of the jar with candy	4. Drawing a triangle/a tree

Note. Each numbered stimulus corresponds to a pair of videos.

> These are two twins. One of them always wears a pink shirt, and one of them always wears a blue shirt. We are going to watch some videos about the twins and their friends. In one video, the twin in the pink shirt is going to do something, and her friend is going to watch her. In the other video, the twin in the blue shirt is going to do something, and her friend is going to watch her. Sometimes the friends fall asleep, and sometimes the friends stay awake. At the end of the videos, we are going to ask you what the friends saw. Even if they fell asleep or stayed awake, the friends saw something, and you will have to tell us what they saw.

The experimenter then played the videos.

Children completed both a quantity and a transformation task. Each task had four trials. In a typical quantity trial, in one of the videos, the twin completed an action partway (e.g., colored half of a star) and paused briefly, at which point the "friend"/observer yawned, stretched, and fell asleep. The agent went on to complete the action and conceal the result (e.g., colored all of the star and covered the drawing) and then woke the sleeping observer. In the other video, the twin used an identical star and completed the whole action (e.g., colored all of the star, with a short pause midway) while the observer stayed awake; at the end, the twin covered the drawing (see Figure 1, for a sample). After each video was done, its final frame remained on the screen. The experimenter next said, "Now one of the friends is going to tell us what she saw. *(Turning to the "friends.")* What did you see?" Children then heard one of two prerecorded responses containing either a weak (*some*) or a strong (*all*) quantifier in the partitive form (e.g., "The girl colored some/all of the star") and were

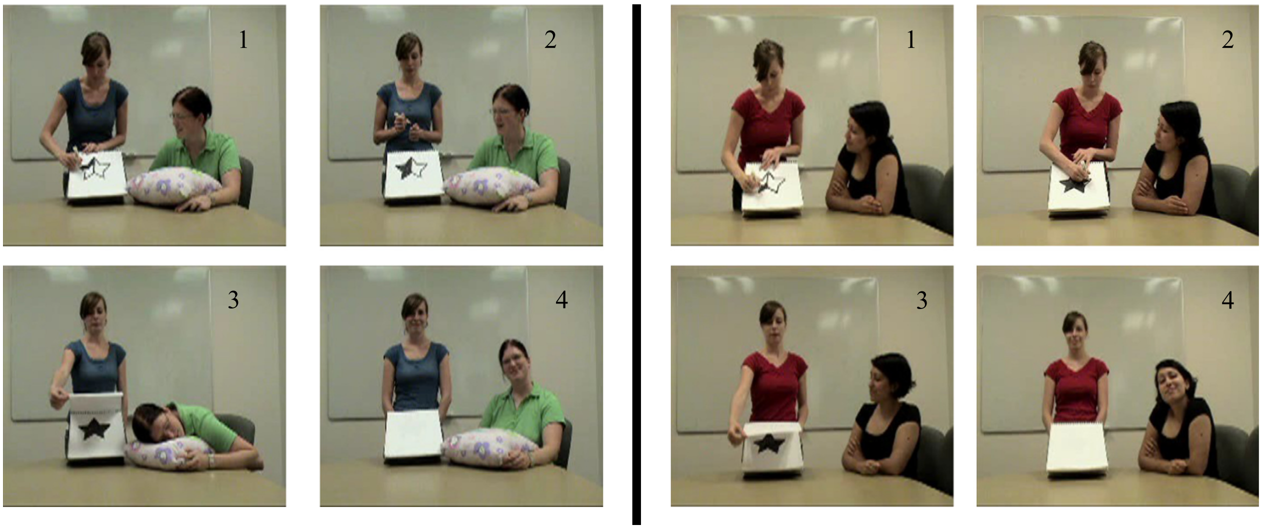

Figure 1. Panels 1–4 show critical points in the two videos for a sample quantity trial in Experiment 1. In the video on the left, one of the twins colors a star partway and pauses briefly; the "friend" stretches, yawns, and falls asleep; then the twin goes on to color all of the star and covers the result of her action before she wakes the friend. In the video on the right, the other twin performs the same action, but the "friend" remains awake throughout. [Color figure can be viewed at wileyonlinelibrary.com]

asked, "Which friend said that? Point to the friend who said that." Of interest was whether children would match the weak statement with the character having witnessed the partially complete action and the strong statement with the character having witnessed the completed action. To help children choose between the two observers, after the second video played and before the test questions were asked, a circle appeared around each of the friends' faces. Furthermore, to simplify the task of tracking what each observer knew, we used single objects (*some/all of the star*) as opposed to sets (*some/all of the stars*)—and the observer always fell asleep halfway through the action.

In the transformation task, the procedure was the same as in the quantity task, but the agent drew pictures on a sketchpad rather than act on objects. Both of the "friends"/observers initially saw each twin make the same drawing (e.g., a cloud). In one video, the observer fell asleep after the initial drawing was completed and did not see the twin add to it to transform it into something else (e.g., an ice cream; see Figure 2). In the other video, the observer watched the initial drawing turn into something else. Then the experimenter said, "Now one of the friends is going to tell us what she saw. *(Turning to the "friends.")* What did you see?" Children next heard one of two prerecorded responses ("The girl drew a cloud/an ice cream") and were asked, "Which friend said that? Point to the friend who said that."

Across participants, order of the quantity and transformation tasks was counterbalanced. Within each task, each participant heard two weak and two strong statements (blocked), with block order counterbalanced across participants. Block order for each participant was consistent across the quantity and transformation tasks. The pairing of each video with a weak/strong statement was reversed for half of the participants. The same set of "twins" appeared throughout, but each video used a different pair of "friends." If a friend was repeated, she appeared in a video playing on the opposite side from before and had a different role (knowledgeable vs. ignorant observer). All friends were female, and all prerecorded statements used different female voices. Friends were always seated to the right of each twin. The position of the video that contained the sleeping character (left/right) was counterbalanced. The experimenter recorded whether children chose the friend in the left or right video on an answer sheet. In the few cases where children failed to make a clear response, the test question was repeated.

Two pretest trials were shown in the beginning of each session. These trials were meant to alert participants to the importance of tracking not only events in the videos but also observers' visual access to the events. Additionally, they were meant to counter the possibility that children might avoid attributing any knowledge to a partially informed observer. The pretests were identical to the main phase except that the observer who fell asleep saw an event that the always-awake observer did not (and vice versa). For example, in one of the videos of a pretest, one twin waved, and then the observer stretched, yawned,

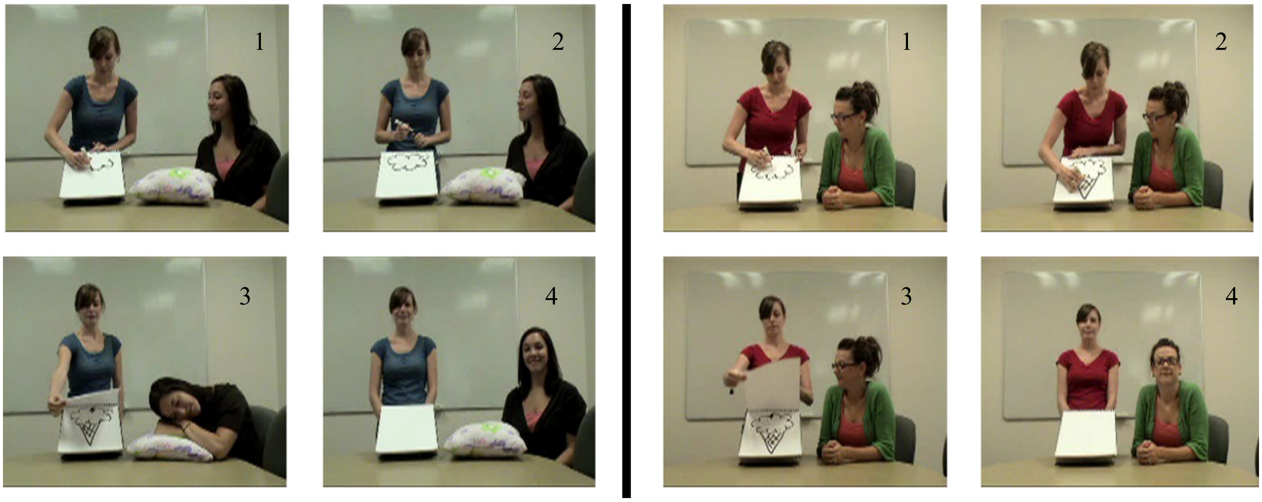

Figure 2. Panels 1–4 show critical points in the two videos for a sample transformation trial in Experiment 1. In the video on the left, one of the twins draws a cloud and pauses briefly; the "friend" stretches, yawns, and falls asleep; then the twin goes on to turn the cloud into an ice cream cone and covers the result of her action before she wakes the friend. In the video on the right, the other twin performs the same action, but the "friend" remains awake throughout. [Color figure can be viewed at wileyonlinelibrary.com]

and fell asleep; afterward, the twin went on to draw a picture. In the second video, the other twin drew a picture while the observer watched. At the end of both videos, the twins covered their drawings. Then the experimenter said, "Now one of the friends is going to tell us what she saw. *(Turning to the 'friends'.)* What did you see?" Children next heard a statement ("The girl waved/drew a picture") and had to match it to the right person. The statement "The girl waved" could only be attributed to the observer who fell asleep, and the statement "The girl drew a picture" could only be attributed to the observer who stayed awake. Order of the characters that were the target of each pretest was counterbalanced. Assignment of statement version to videos was also counterbalanced across participants. Feedback was given after each pretest (feedback for the weak trials included the phrase: "She fell asleep, but she knew the answer!"). Less than one-fourth of the children made an error on one of the pretests and had to be corrected.

Two control questions ("Who fell asleep? Who didn't fall asleep?") followed presentation of the videos in all test and pretest trials. The character who would fall asleep always had a pillow that served as a visual reminder to the children. Children always gave correct answers to the control questions.

Results

Results from test trials are presented in Figure 3. Correct responses for the strong trials were defined

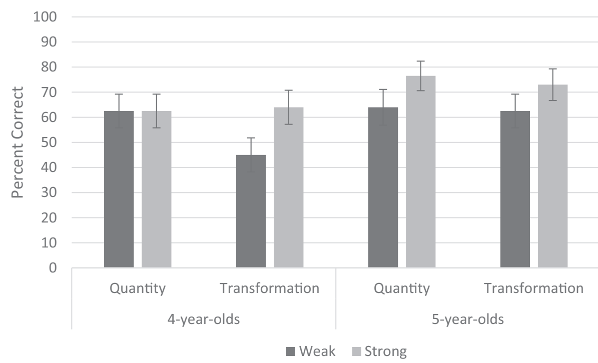

Figure 3. Four- and 5-year-olds' performance in Experiment 1. Error bars depict standard error.

as selections of the fully informed observer (i.e., the one who had witnessed all of the event in the quantity task or the complete drawing in the transformation task); correct responses for the weak trials were defined as selections of the partly informed observer (i.e., the one who fell asleep and thus witnessed only some of the event in the quantity task or part of the drawing in the transformation task). The proportion of children's correct responses on test trials was submitted to a mixed analysis of variance (ANOVA) with *trial* (strong, weak) and *task* (quantity, transformation) as within-subjects variables and *age* (4, 5) as a between-subjects variable. The ANOVA revealed only an effect of age, $F(1, 62) = 5.65$, $p = .02$, with 5-year-olds being overall better than 4-year-olds. Further analyses showed that 5-year-olds performed at levels different from chance in both strong, *t*

(31) = 4.45, $p = .0001$, two-tailed, $M = 74.5\%$, and weak statements, $t(31) = 2.46$, $p = .01$, two-tailed, $M = 63\%$. Four-year-olds, however, performed at levels different from chance in strong statements, $t(31) = 2.63$, $p = .03$, two-tailed, $M = 63\%$, but not in weak ones, $t(31) = .68$, $p = .50$, two-tailed, $M = 53.5\%$.

Discussion

Experiment 1 yielded two major findings. First, it demonstrated that 5-year-old children could link the content of an utterance to an observer's state of mind: Children at this age were able to attribute a strong scalar statement to a character who was fully informed and a weak scalar statement to a character with partial knowledge. Even though very error prone, 5-year-olds' performance in the present data is much higher than in previous studies that tested understanding of the pragmatics of scalars at this age group (e.g., Barner et al., 2011; Guasti et al., 2005; Noveck, 2001; Papafragou & Musolino, 2003; but cf. Katsos & Bishop, 2011; Papafragou & Tantalou, 2004). Importantly, the present study is the first to show that 5-year-olds go beyond computing the relative informativeness of weak and strong statements but to some degree make assumptions about the kinds of speakers for whom these statements would make reasonable (i.e., knowledge-appropriate) conversational contributions. In this sense, 5-year-olds in this study appear sensitive not simply to *potential*, but to *actual* scalar implicatures communicated by weak scalar statements.

A second major finding from the present study is that there are developmental differences in the way children link utterance content to a speaker's epistemic state, with 4-year-olds overall performing worse than 5-year-olds when assessing scalar statements. Furthermore, even though 4-year-olds were able to attribute a strong scalar statement to a character who was fully informed, they could not reliably associate a weak scalar statement with a character with partial knowledge.

What can explain 4-year-olds' difficulty with attributing scalar statements to speakers? There are at least two possibilities. One possibility is that children's difficulty stemmed from the demands of reasoning about other people's mental states, an ability known to develop during the preschool years (Wellman et al., 2001). On this hypothesis, 4-year-olds' difficulty with both strong and weak scalars might be related to the demands of tracking two observers, one of whom holds a false belief or is at least ignorant of an outcome to which the children themselves are privy. Alternatively, or additionally, younger children's failure might be related to the ability to access and use members of a linguistic scale, which has also been argued to develop with age (see Barner et al., 2011, among others). On this hypothesis, for instance, children's difficulties with scalar statements in the present task might be related, at least in part, to the difficulty of spontaneously accessing a scalar alternative (e.g., *all* when *some* is uttered) and contrasting it with the newly heard scalar term.

The present data cannot adjudicate between these two possibilities. However, there are two straightforward empirical tests that can do so. First, the epistemic (but not the lexical) hypothesis predicts that 4-year-old children should be more likely to succeed in a version of Experiment 1 that does not involve complex mental-state tracking. Second, the lexical (but not the epistemic) hypothesis leaves open the possibility that children might succeed in a nonlinguistic version of Experiment 1 that preserves complex mental-state tracking. We tested these two predictions in Experiments 2a and 2b, respectively.

Experiment 2a

Experiment 2a used versions of the stimuli from Experiment 1. Children saw two videos in which two identical agents ("twins") performed an action while observed by another character (a "friend"). Unlike Experiment 1, each agent performed a slightly different action, but the actions still formed a set–subset relationship (e.g., one girl colored part of a star and the other girl colored all of the star). Also unlike Experiment 1, each observer watched the corresponding action in its entirety (neither one fell asleep). As a result, each observer ended up with a different (but true) belief about the witnessed event. In the case of the observer who watched part of the star being colored, that observer knew for a fact that not all of the star was colored. Finally, there was no discrepancy between what the child knew and what each of the observers knew about the corresponding event, hence there was no need to track the observers' mental states separately from the child's own knowledge.

As in our previous experiment, children had to attribute either a weak or a strong scalar statement of the form "The girl colored some/all of the star" to one of the observers. Of interest was whether children would succeed in these new contexts that did not require complex mental perspective taking (as predicted by the epistemic hypothesis) or continue to fail (as expected by the position that children's difficulties are due to immature linguistic scalar knowledge).

Method

Participants

Twenty-five 4-year-old children (M_{age} = 4;6, range = 4;0–4;11; 12 female) participated. They were recruited from Newark, Delaware preschools. None of the children had participated in the earlier experiment.

Materials and Procedure

Materials and procedure were similar to those in Experiment 1 with the following modifications. In a typical quantity trial, one video showed one twin color part of a star while a "friend" watched. The other video showed the other twin perform the same action but end up coloring the entire star as a different "friend" watched. At the end of the second video, the experimenter said, "Now one of the friends is going to tell us what happened. *(Turning to the "friends.")* What did the girl color?" Then children heard a prerecorded weak or strong statement ("The girl colored some/all of the star") and were asked, "Who said that? Point to the friend who said that."

In the transformation task, the procedure was the same, but the twins drew pictures rather than act on objects. In a typical trial, one video showed one twin draw a cloud as an observer watched. The other video showed the other twin perform the same action but then add to it (e.g., transforming the cloud into an ice cream cone) as a different observer watched. At the end of the second video, the experimenter said, "Now one of the friends is going to tell us what happened. What did the girl draw?" Children heard a prerecorded statement ("The girl drew a cloud/ice cream/") and were asked, "Who said that? Point to the friend who said that."

Two pretest trials were shown in the beginning to familiarize children with the act of attributing an utterance to the appropriate speaker. In one pretest trial, one video showed a twin drawing a flower while a friend watched. The second video showed the other twin draw a butterfly as a different friend watched. At the end of the second video, the experimenter said, "Now one of the friends is going to tell us what happened. What did the girl draw?" Children then heard two statements (e.g., "The girl drew a flower" and "The girl drew a butterfly") and had to say after each statement "who said that." Only one child had to be corrected on one of the four (two per trial) pretest questions.

Results

Results from test trials are presented in Figure 4. Correct responses for the strong trials were defined as selections of the speaker whose knowledge corresponded to a complete event (i.e., the observer who had witnessed the completed event in the quantity task or the complete drawing in the transformation task); correct responses for the weak trials were defined as selections of the speaker whose knowledge corresponded to a narrower state of affairs (i.e., the observer who had witnessed a partly completed event in the quantity task or part of the drawing in the transformation task). The proportion of children's correct responses on test trials was submitted to a repeated measures ANOVA with *trial* (strong, weak) and *task* (quantity, transformation) as within-subjects variables. The ANOVA revealed only a marginal effect of trial, $F(1, 24) = 3.42$, $p = .07$: strong trials led to correct performance 84% of the time and weak trials did so 74% of the time. Performance was consistently different from chance levels for both strong, $t(24) = 6.83$, $p < .0001$, two-tailed, and weak statements, $t(24) = 3.29$, $p = .0031$, two-tailed.

Comparison to Experiment 1

We compared 4-year-old children's performance in Experiments 1 and 2a through a mixed ANOVA with trial (strong, weak), task (quantity, transformation), and experiment (1, 2a) as factors. The ANOVA revealed only an effect of experiment, $F(1,$

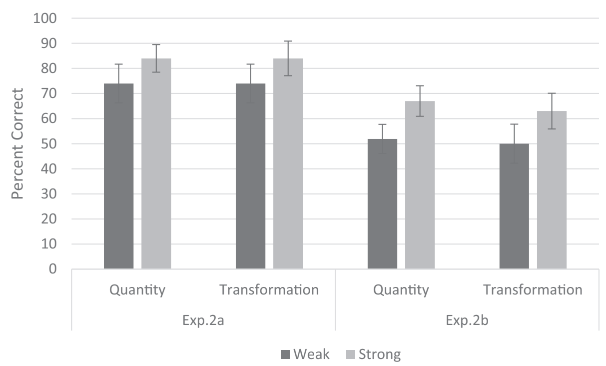

Figure 4. Four-year-olds' performance in Experiments 2a and 2b. Error bars depict standard error.

55) = 10.37, $p = .002$, with performance being much better in Experiment 2a ($M = 79\%$) than in Experiment 1 ($M = 58\%$).

Discussion

Experiment 2a tested the extent to which 4-year-old children's errors in Experiment 1 were due to epistemic factors (e.g., reasoning about others' mental states) as opposed to purely linguistic factors (e.g., accessing linguistic scalemates). Unlike Experiment 1, in Experiment 2a, statements had to be matched to speakers with full knowledge whose epistemic state was the same as that of the child's: these speakers had witnessed one of two events that happened to form a set–subset relationship (e.g., a girl colored some/all of the star). Under these conditions, 4-year-olds were able to match *some* and *all* statements to appropriate observers (with only a marginal advantage of strong over weak trials). We conclude that 4-year-old children are capable of distinguishing between more and less informative statements, but the ability to link these statements to agents' mental states develops with age. The fact that 4-year-olds' performance improved in both weak and strong trials compared to Experiment 1 lends further support to the conclusion that children of this age have difficulties with taking the epistemic stance in our task (as opposed to interpreting a specific type of quantifier).

Even more than the results from Experiment 1, the 4-year-olds' success in Experiment 2a is striking in the context of prior reports of older preschoolers' failure to derive scalar implicatures from weak statements (e.g., Barner et al., 2011; Guasti et al., 2005; Noveck, 2001; Papafragou & Musolino, 2003). Unlike Experiment 1, the present paradigm provides two distinct states of affairs—each accessed by a different observer—for children to choose from in interpreting a linguistic utterance. Crucially, unlike Experiment 1, each observer holds a true belief that is also shared by the child.

Experiment 2b

Experiment 2b investigated whether the difficulties faced by 4-year-olds in Experiment 1 would persist if the linguistic stimuli were replaced by nonlinguistic, but still representational, stimuli that could be perceived as clues to the observers' mental states. The epistemic hypothesis expects that, to the extent that the task continues to require children to "take the epistemic step," 4-year-olds should still have difficulties in matching stimuli to observers. The lexical hypothesis, however, leaves it open for children to perform better, because the main source of children's difficulty (incompletely acquired linguistic scalar items) is now removed.

Experiment 2b was identical to Experiment 1 with the exception of the test prompts. Recall that the prompts in Experiment 1 were weak and strong scalar statements (e.g., "The girl colored some/all of the star," "The girl drew a cloud/an ice cream") that had to be attributed to a speaker in response to the question "Who said that?" In Experiment 2b, these statements were replaced with the corresponding pictures taken from the videos (e.g., a picture of a partly/completely colored star, a picture of a cloud/an ice cream) that had to be attributed to an observer in response to the question "Who saw that?"

As in Experiment 1, to answer correctly, children had to take into account what each observer saw (and hence knew) rather than what they themselves had seen. For the "strong" pictures, there was a straightforward mapping to the fully informed observer. But for the "weak" pictures, children had to determine on pragmatic grounds that such pictures would be underinformative representations of the knowledge state of the observer with full access to the event (and had to therefore assign them to the observer with partial access). Just like utterances, pictures are representations and may be more or less informative with respect to the state of affairs they encode. Notice that the weak pictures, similarly to the weak statements in Experiment 1, do not exclude the complete event in a definitive way (e.g., saying that the picture where some of the star has been colored represents what someone saw does not exclude the possibility that the person actually saw that *all* of the star has been colored). Just like a weak scalar statement, the picture of a partly colored star *does* represent what the fully informed participant saw—but it does not represent *everything* that the fully informed participant saw.

Method

Participants

Twenty-six 4-year-olds (M_{age} = 4;7, range = 4;2–4;11; 13 female) participated. Children were recruited from Newark, Delaware preschools. None of the children had participated in the earlier experiments. We also tested a control group of 12 adults (7 females) that were undergraduate students at the

University of Delaware attending Introductory Psychology classes. None of them had participated in Experiment 1. Adults answered correctly 100% of the time, and their results will not be discussed further.

Materials and Procedure

Materials and procedure were as in Experiment 1 with modifications only in the final test prompts. Recall that, in the earlier study, children were presented with four quantity and four transformation trials. In each trial, they watched two videos, each of which showed one of two twins performing a certain action while an observer ("friend") watched. In a typical quantity trial, in one of the videos, one twin completed an action partway (e.g., colored some of the star), and then the "friend"/observer fell asleep. The twin then completed the action and concealed the result (e.g., colored all of the star and covered the drawing) and then woke the sleeping observer. In the other video, the other twin used an identical star and completed the whole action (e.g., colored all of the star) and then covered the drawing while the observer stayed awake. After each video was done, its final frame remained on the screen. Unlike Experiment 1, where children heard a statement and had to say who uttered it, in Experiment 2b the experimenter said, "One of the friends saw the girl do this to the star"—and presented a card with either a half-colored or a fully colored star. The picture on the card was taken from the actual videos. Then children were asked, "Who saw this? Point to the friend who saw this."

In the transformation task, both of the "friends"/observers initially saw each twin make the same drawing (e.g., a cloud). In one video, the observer fell asleep after the initial drawing was completed and did not see the twin add to it to transform it into something else (e.g., an ice cream; see Figure 2). In the other video, the observer watched the initial drawing turn into something else. Unlike Experiment 1, the experimenter said, "One of the friends saw the girl draw this"—and presented a card with still photographs of either a cloud or an ice cream taken from the videos (see Figure 2). Then children were asked, "Who saw this? Point to the friend who saw this."

As in Experiment 1, two pretest trials were shown in the beginning of each session. Recall that the pretests were designed in such a way that the observer who fell asleep saw an event that the always-awake observer did not (and vice versa). For example, in one of the videos of a pretest, one twin waved, then drew a picture; the observer only witnessed the waving event and then fell asleep. In the second video, the other twin drew a picture while the observer watched. Unlike Experiment 1, where participants heard a statement during the pretests, participants in Experiment 2b saw a card corresponding to one of the events (here, waiving or drawing) and had to match it to the right person. The card where the girl waived could only be attributed to the observer who fell asleep and the card where the girl drew a picture could only be attributed to the observer who stayed awake. Order of the characters that were the target of each pretest was counterbalanced. Assignment of card version to videos was also counterbalanced across participants. Feedback was given after each pretest (feedback for the weak trials included the phrase, "She fell asleep, but she knew the answer!"). Less than one-fourth of the children made an error on one of the pretests and had to be corrected.

Also as in Experiment 1, two control questions ("Who fell asleep? Who didn't fall asleep?") followed presentation of the videos in all test and pretest trials. The character who would fall asleep always had a pillow that served as a visual reminder to the children. Children always gave correct answers to the control questions.

Results

Results from test trials are presented in Figure 4. Correct responses for the strong trials were defined as selections of the fully informed observer (i.e., the one who had witnessed all of the event in the quantity task or the complete drawing in the transformation task); correct responses for the weak trials were defined as selections of the partly informed observer (i.e., the one who fell asleep and thus witnessed only some of the event in the quantity task or part of the drawing in the transformation task). The proportion of children's correct responses on test trials was submitted to an ANOVA with *trial* (strong, weak) and *task* (quantity, transformation) as within-subjects variables. The ANOVA revealed no significant effects or interactions. Performance on strong trials differed significantly from chance, $t(25) = 2.60$, $p = .01$, two-tailed, $M = 65\%$, but performance on weak trials did not, $t(25) = 0.16$, $p = .87$, two-tailed, $M = 50\%$.

Comparison to Experiment 1

We compared 4-year-olds' performance across Experiments 1 and 2b using a mixed ANOVA with trial (strong, weak) and task (quantity, transformation) as within-subjects factors and experiment (1, 2b) as a between-subjects factor. This analysis revealed only a marginal effect of trial, $F(1, 56) = 3.16$, $p = .08$. There was no main effect of experiment or interactions with that factor.

Discussion

Experiment 2b investigated whether 4-year-olds' ability to evaluate the informativeness of a representation extended beyond situations where children had to match a *sentence* to what an observer saw to situations where children had to match a *picture* to what an observer saw. We found that the pattern of results did not change from Experiment 1; overall, children's performance was fairly weak. The present data complement the results from Experiment 2a and jointly point to the conclusion that the difficulty 4-year-olds face in Experiment 1 is not linked to the linguistic scalar expressions per se but to the epistemic requirements of the task.

General Discussion

In the present experiments, we investigated the development of children's pragmatic abilities. More concretely, we asked whether young children can generate conversational (scalar) implicatures from weak scalar statements (e.g., *The girl colored some of the stars*) and match those statements to speakers in accordance with the speakers' epistemic state. Recall that, in order to compute a scalar implicature, children need to implicitly understand that an utterance contains as much true information as the speaker is epistemically capable of conveying (given what is relevant within the conversational purposes). This ability requires sensitivity to a pragmatic mechanism that estimates expected levels of relevant information in a conversation, processes what the speaker says against these expectations, and assumes that, if what is said falls short of a stronger (and relevant) statement the speaker might have used, the reason must be that the speaker is not in an epistemic position to offer the stronger statement (see Horn, 1972, for details).

Our basic experimental design (Experiment 1) involved two observers who had differential access to information: one observer had full access to an event (e.g., a person watched while a girl colored all of the star), whereas a second observer only had access to an event subpart (e.g., another person watched while another girl colored only some of the star; later—unbeknownst to the observer—all of the star was colored). Notice that the second observer's access to the event was different from the child's and had to be tracked separately. Children heard either a weak or a strong scalar statement describing what happened (e.g., *The girl colored some/all of the star*) and had to match it to the corresponding observer. According to the rationale of this paradigm, strong statements unambiguously belonged to the fully informed observer, whereas weak statements were logically attributable to either observer but pragmatically pointed to the observer who knew less (because these statements would have been underinformative/infelicitous if uttered by the fully informed observer).

Our results suggest that 5-year-old children were capable of inferring how much the speaker knows from the strength of a spoken utterance: Children at this age understood that a strong statement needs to be attributed to the fully informed observer and also successfully attributed a weak statement to the observer with partial knowledge. Four-year-olds, however, succeeded with strong trials but were not consistently accurate in linking weak statements with less than fully informed speakers. There was no difference between logical and quantificational as opposed to ad hoc, contextual scales for either age group.

We next asked whether 4-year-olds' fragile performance was due to the epistemic complexities of the task or the linguistic knowledge of scalars such as *some* and *all*. In Experiment 2a, we removed the requirement of tracking other cognizers' perspectives as distinct from one's own perspective and found that 4-year-olds' performance improved considerably compared to Experiment 1. In Experiment 2b, we replaced the linguistic stimuli of Experiment 1 with nonlinguistic (pictorial) stimuli and found that 4-year-olds performed similarly to Experiment 1. These two further manipulations show that 4-year-olds' difficulties in our first study were linked to the complexity of mental state calculations and not to the linguistic stimuli themselves.

Together, our findings make several novel contributions to the literature on pragmatic development and early perspective taking. To begin with, the 5-year-olds' behavior in Experiment 1 offers the first evidence in the literature that preschoolers are able to some degree to consult a speaker's epistemic state when computing a scalar implicature and support rich, Gricean models of scalar implicature computation (cf. Carston, 1998; Geurts, 2011; Grice, 1989; Horn, 1972; Sauerland, 2004, 2012; Sperber & Wilson, 1986/1995, among others). Furthermore, the 5-year-olds' performance in Experiment 1 (as well as the 4-year-olds' success in the simpler epistemic context of Experiment 2a) differs from several prior studies showing that children of this age have difficulty with pragmatic inferences, especially the computation of scalar implicatures. In many of these studies, children's sensitivity to pragmatic inference was measured by the ability to reject underinformative utterances offered either by the experimenter (Guasti et al., 2005; Noveck, 2001) or by a speaker known to be incompetent (e.g., a "silly puppet"; Katsos & Bishop, 2011; Papafragou & Musolino, 2003). Our paradigm differed from this earlier work in two key respects. First, our task did not involve explicit judgment of linguistic utterances: Rather than ask children to point out the infelicity of weak scalar statements when uttered by fully informed speakers, we simply asked children to match scalar statements to speakers depending on the speakers' knowledge state. Second, and most importantly, the use of a weaker statement was motivated by the fact that the speaker's knowledge state did not warrant the stronger statement. We believe that these novel features of our study enabled 5-year-olds to perform systematically above chance in calculating scalar implicatures from weak scalar statements in Experiment 1. The success of the 5-year-olds in our studies is consistent with and extends prior results that show early sensitivity to the pragmatics of scalar terms to be context and task dependent (Guasti et al., 2005; Katsos & Bishop, 2011; Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Pouscoulous et al., 2007).

The same findings are also significant in the context of broader debates about children's early communication skills by showing that young children are sensitive to social pragmatic reasoning in calculating speaker meaning. In this sense, they are consistent with evidence demonstrating, among other things, that preschoolers consult a speaker's knowledge when learning a novel word (Papafragou, Fairchild, Cohen, & Friedberg, 2016; Sabbagh & Baldwin, 2001) and take into account their interlocutors' perspective during online reference assignment (Nadig & Sedivy, 2002).

Despite the above-chance performance of 5-year-olds in Experiment 1, our findings strongly suggest that the ability to link utterance content and speaker knowledge undergoes considerable development. This is highlighted by the fact that even 5-year-olds make many errors in the scalar implicature task of Experiment 1. Furthermore, 4-year-olds fail at consulting speaker knowledge when computing scalar inferences in Experiment 1 and also have difficulty evaluating the appropriateness of a nonlinguistic stimulus with respect to what someone knows in Experiment 2b.

Could there be alternative interpretations for the observed patterns in our data? One might hypothesize that children's difficulty is caused by confusion over surface features of our task, such as the fact that one character falls asleep in Experiments 1 and 2b. We do not believe that having a character fall asleep posed any difficulty by itself. Children had no problem answering questions about who fell asleep and who did not; furthermore, children made appropriate selections of the character who fell asleep versus the fully attentive observer in pretest trials.

Another possibility is that children might have a slight overall bias to pick one video over the other for reasons unrelated to the task itself. Inspection of Figures 3 and 4 shows a numerical difference between strong and weak trials that could suggest a selection bias. However, statistical analysis shows no reliable effect of trial in our experiments. Furthermore, given that both videos in Experiments 1 and 2b involve identical, completed events, and differ only in terms of the observers' access to the events, it is difficult to avoid the conclusion that children's difficulties in our task are mostly related to epistemic factors that govern the use of scalars.

The epistemic limitations we have documented in Experiments 1 and 2b are reminiscent of other domains in which young children have been shown to have difficulty tracking mental-state content. For instance, 4-year-olds but not 5-year-olds are known to face difficulties in reasoning about others' false beliefs, especially in tasks that involve overt verbal responses and complex perspective tracking (Wellman et al., 2001). Four-year-olds, unlike older children, also tend to make errors when evaluating whether and how others have gained knowledge after a certain experience (e.g., Pillow, Hill, Boyce, & Stein, 2000; Sodian & Wimmer, 1987). Even 5-year-olds have difficulty recognizing epistemic uncertainty: For instance, faced with an uncertain outcome, they tend to commit to a possible conclusion before decisive evidence is available (Acredolo & Horobin, 1987; Klahr & Chen, 2003; Robinson, Rowley, Beck, Carroll, & Apperly, 2006). In these other domains, as in our own data, the question arises as to whether early failures with perspective taking are due to conceptual immaturity or to the specific perspective-taking requirements of sometimes complex tasks. At least in the domain of false belief reasoning, recent evidence seems to favor the second possibility: For instance, several studies have uncovered early sensitivity to false beliefs using implicit measures with infants or toddlers (Buttelmann, Carpenter, & Tomasello, 2009; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; see Baillargeon, Scott, & He, 2010, for a review) or less taxing perspective-taking tasks with young preschoolers (Rubio-Fernández & Geurts, 2013; see also Robinson & Whitcombe, 2003, for related evidence in the area of monitoring knowledge sources).

Our findings and those sketched above raise the possibility that other, potentially more streamlined ways of investigating sensitivity to speaker knowledge might lead 4-year-olds to succeed in both the linguistic (Experiment 1) and the nonlinguistic (Experiment 2b) version of our basic paradigm. Recall that this paradigm involved watching two stories, each of which involved two characters (the agent and the observer). Because of the demands of following the unfolding action performed by the agent, it is possible that children's tracking of the observer's perspective within each story was disrupted (see also Rubio-Fernández & Geurts, 2013). Furthermore, our paradigm relied on false belief contexts to investigate young children's awareness of the knowledge states of others. Such contexts are only a subset of the situations in which what a character knows might diverge from what the child knows and different ways of implementing a mismatch between what the child knows and what the observer knows might yield different results. Further research is required to throw light on the mechanisms whereby young learners ultimately fine tune their calculations of others' mental states and connect these calculations to speakers' word choices when drawing pragmatic inferences in communication.

References

Acredolo, C., & Horobin, K. (1987). Development of relational reasoning and premature closure. *Developmental Psychology*, *23*, 13–21. https://doi.org/10.1177/014272379201203506

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *188*, 87–96. https://doi.org/10.1016/j.cognition.2010.10.010

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition*, *21*, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1450–1460. https://doi.org/10.1037/a0027850

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Braine, M., & Rumain, B. (1981). Children's comprehension of 'or': Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*, 46–70. https://doi.org/10.1016/0022-0965(81)90003-5

Breheny, R., Ferguson, H., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*, 423–440. https://doi.org/10.1016/j.cognition.2012.11.012

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*, 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Carston, R. (1998). Informativeness, relevance, and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam, The Netherlands: Benjamins.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belleti (Ed.), *Structures and beyond* (pp. 39–103). Oxford, UK: Oxford University Press.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J., Do L. Domínguez & A. Johansen (Eds.), *BUCLD 25 proceedings* (pp. 157–168). Somerville, MA: Cascadilla Press.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In C. Maienborn, K. von Heusinger & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2297–2331). Mouton: de Gruyter.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417–431. https://doi.org/10.1017/S0305000900013842

Conti, D., & Camras, L. (1984). Children's understanding of conversational principles. *Journal of Experimental Child Psychology*, *38*, 456–463. https://doi.org/10.1016/0022-0965(84)90088-2

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. https://doi.org/10.1016/j.cognition.2010.12.011

Degen, J., & Tanenhaus, M. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*, 667–710. https://doi.org/10.1111/cogs.12171

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*, 630–641. https://doi.org/10.1037//0012-1649.37.5.630

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*, 365–394. https://doi.org/10.1080/15475441.2011.626386

Fox, D., & Spector, B. (2018). Economy and embedded exhaustification. *Natural Language Semantics*, *26*, 1–50.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *1*, 1–82. https://doi.org/10.3765/sp.4.1

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. New York, NY: Academic Press.

Geurts, B. (2011). *Quantity implicatures*. Cambridge, UK: Cambridge University Press.

Glucksberg, S., Krauss, R., & Higgins, T. (1975). The development of referential communication skills. In F. D. Horowitz (Ed.), *Review of child development research* (Vol. 4, pp. 305–345). Chicago, IL: University of Chicago Press.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grodner, D., Klein, N., Carbary, K., & Tanenhaus, M. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In R. Hastings, B. Jackson & Z. Zvolenszky (Eds.), *Proceedings of SALT XI* (pp. 231–247). Ithaca, NY: Cornell University.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*, 667–696. https://doi.org/10.1177/00238309080510010901

Hirschberg, J. (1985). *A theory of scalar implicature*. Doctoral dissertation, University of Pennsylvania.

Horn, L. (1972). *On the semantic properties of the logical operators in English*. Doctoral dissertation, UCLA Distributed by IULC, Indiana University.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

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*, 1723–1739. https://doi.org/10.1037/a0016704

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67–81. https://doi.org/10.1016/j.cognition.2011.02.015

Klahr, D., & Chen, Z. (2003). Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development*, *74*, 1275–1296. https://doi.org/10.1016/S0065-2407(08)00010-4

Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*, 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*, 403–422. https://doi.org/10.1017/S0142716406060334

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of "scalar inferences." In N. Burton-Roberts (Ed.), *Advances in pragmatics* (pp. 184–212). Basingstoke, UK: Palgrave.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–677. https://doi.org/10.1111/j.1467-8624.1996.tb01758.x

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258. https://doi.org/10.1126/science.1107621

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*, 191–214. https://doi.org/10.1080/15475441.2015.1024834

Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language*, *33*, 721–757. https://doi.org/10.1017/S0305000906007550

Papafragou, A., Fairchild, S., Cohen, M., & Friedberg, C. (2016). Learning words from speakers with false beliefs. *Journal of Child Language*. Advance online publication. https://doi.org/10.1017/s0305000916000301

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8

Papafragou, A., & Skordos, D. (2016). Scalar implicature. In J. Lidz, W. Snyder, & J. Pater (Eds.), *Oxford handbook* *of developmental linguistics* (pp. 611–629). Oxford, UK: Oxford University Press.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*, 71–82. https://doi.org/10.1207/s15327817la1201_3

Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology*, *36*, 169–179. https://doi.org/10.1037/0012-1649.36.2.169

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition*, *14*, 347–375. https://doi.org/10.1080/10489220701600457

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*, B13–B27. https://doi.org/10.1016/j.cognition.2005.01.008

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819–865. https://doi.org/10.1207/s15516709cog0000_31

Robinson, E., Rowley, M., Beck, S., Carroll, D., & Apperly, I. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development*, *77*, 1642–1655. https://doi.org/10.1111/j.1467-8624.2006.00964.x

Robinson, E., & Whitcombe, E. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74*, 48–62. https://doi.org/10.1111/1467-8624.t01-1-00520

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*, 27–33. https://doi.org/10.1177/0956797612447819

Sabbagh, M., & Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070. https://doi.org/10.1111/1467-8624.00334

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367–391. https://doi.org/10.1023/B:LING.0000023378.71748.db

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass*, *6*, 36–49. https://doi.org/10.1002/lnc3.321

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, *38*, 1–38. https://doi.org/10.2307/1165783

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18. https://doi.org/10.1016/j.cognition.2016.04.006

Smith, C. (1980). Quantifiers and question-answering in young children. *Journal of Experimental Child Psychology*, *30*, 191–205. https://doi.org/10.1016/0022-0965(80)90057-0

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, *58*, 424–433. https://doi.org/10.2307/1130519

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *16*, 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Spector, B. (2006). Scalar implicatures: Exhaustivity and Gricean reasoning. In M. Aloni, A. Butler, & P. Dekker (Eds.), *Questions in dynamic semantics* (pp. 229–254). Amsterdam, The Netherlands; Singapore City, Singapore: Elsevier.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Cambridge, MA: Harvard University Press. (Original work published 1986)

Stiller, A., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc scalar implicature in preschool children. *Language Learning and Development*, *11*, 176–190. https://doi.org/10.1080/15475441.2014.927328

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science*, *18*, 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information*, *13*, 491–519. https://doi.org/10.1007/s10849-004-2118-6

Wellman, H. M., Cross, D., & Watson, J. (2001). A meta-analysis of theory of mind development: The truth about false belief. *Child Development*, *72*, 655–684. https://doi.org/10.1111/1467-8624.00304

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. https://doi.org/10.1016/0010-0277(83)90004-5

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Video S1. Sample quantity trial, Experiment 1. The video corresponds to the left-most panel of Figure 1.

Video S2. Sample quantity trial, Experiment 1. The video corresponds to the right-most panel of Figure 1.

Video S3. Sample transformation trial, Experiment 1. The video corresponds to the left-most panel of Figure 2.

Video S4. Sample transformation trial, Experiment 1. The video corresponds to the right-most panel of Figure 2.