Children's pragmatic inferences as a route for learning about the world

Alexandra C. Horowitz

Department of Psychology, Stanford University

Michael C. Frank

Department of Psychology, Stanford University

Address all correspondence to Alexandra C. Horowitz, Stanford University, Department of Psychology, Jordan Hall, 450 Serra Mall (Bldg. 420), Stanford, CA, 94305. Phone: 650-721-9270. E-mail: `ahorowit@stanford.edu`

Abstract

Information about the world is not always stated explicitly. We investigated whether children can infer generalizable information pragmatically from speakers' word choices. We introduced novel objects described with prenominal adjectives and asked what other category members look like. Preschoolers reliably inferred implied contrast with the support of framing cues (Experiment 1). Performance was reduced after removing the contrastive framing, but improved after a pre-exposure to the relevant adjective pairs (Experiment 2). In a free-response task, children spontaneously produced relevant property contrasts (Experiment 3). Our findings suggest that preschoolers generalize from a single exemplar— not just about what that category is, but also what it isn't. Sensitivity to the implications of speakers' production choices may help children learn about the world.

Keywords: pragmatics; language development; adjectives; knowledge transmission

Introduction

Children learn much important information through explicit instruction (e.g., "put the fork on the left of the plate") and generic statements ("forks go on the left"), but not all information is stated directly. Sometimes information is implicit in the particular production choices a speaker makes. For example, if a parent says, "that's a salad fork," she is implicitly conveying that forks vary in the foods they are intended for (and that most other forks are likely used for non-salad items). More generally, the way we describe the world can reveal to a perceptive observer all sorts of biases about what we find notable, interesting, or generally worthy of comment—and such biases can in turn act as signals about our knowledge of the world. Are children able to use these implicit signals for learning?

We address this question using a simple case study: learning to generalize novel words via minimal contrastive descriptions. We focus on contrastive word choices, as in the above "salad fork" example. Contrastive word choices—the way we use labels and their modifiers—can help identify the speaker's intended referent in the current context (selecting the desired fork) but can also jointly signal generalizable knowledge (forks are associated with meal courses). In the current study, we investigate the idea that adults and children may learn generalizable knowledge via inferences about why speakers choose a particular word to convey a message. To motivate this case study, we begin by discussing two bodies of research: first, work on children's ability to learn about the world from language, and second, work on their ability to reason about the knowledge and beliefs underlying other agents' actions (both non-linguistic and linguistic).

Learning from others' explicit statements

Although learning from the world directly is a very powerful method for acquiring knowledge (Gopnik, 2012), there is no way that even the most precocious child-scientist could reconstruct an adult's knowledge from direct experience alone (Shafto, Goodman, & Frank, 2012; Harris, 2012). Instead, children's knowledge comes from a mixture of direct experiences and

knowledge transmitted by others.

Language in particular is an extremely powerful source of information about the world. From the time children begin to speak, they understand that language is used to communicate information (Vouloumanos, Onishi, & Pogue, 2012; Martin, Onishi, & Vouloumanos, 2012). They expect speakers of the same language to use conventional names for conventional meanings (E. Clark, 1987; Markman & Wachtel, 1988; Diesendruck, 2005), but learn to recognize that individual knowledge such as facts about objects may not be shared (Diesendruck & Markson, 2001). They also show early knowledge that language can communicate about information that goes beyond the here-and-now (Saylor & Ganea, 2007; Ganea, Shutts, Spelke, & DeLoache, 2007). This early, foundational set of assumptions—that speakers will use language in consistent and communicative ways to convey (relatively) abstract knowledge—is critical in allowing children to use language to learn about the world.

While some language describes the current state of the world (e.g., "the salad fork is on the outside"), other statements provide more general information that applies across situations ("salad forks go on the outside"). Generic language—cued in a number of ways, including the use of a bare plural ("salad forks") in the previous example—is a particularly powerful method for conveying such information (Leslie, 2008). Children can use generic language to infer general properties quite early (Gelman & Raman, 2003). In addition, they draw different conclusions from generic statements than non-generic statements, and are more likely to believe that information stated generically is conceptually and functionally central and more widely-known (Cimpian & Markman, 2009; Cimpian & Cadena, 2010; Cimpian & Scott, 2012). And in some contexts, generic language is not even necessary: The simple use of a label or even the use of particular communicative cues—child-directed speech, direct gaze, or pointing—may signal that a speaker is presenting information that is relevant to a kind, category, or practice (Csibra & Gergely, 2009; Butler & Markman, 2012).

Language is such a powerful source of information that preschoolers find it very difficult *not*

to believe what they are told. Three-year-olds can discount inconsistent evidence conveyed through physical markers (e.g. they can learn that an agent purposely places a sticker on the wrong cup, and select the opposite), but they have a much harder time discounting verbal evidence from an unreliable speaker in the same scenario: Even after learning that the agent consistently states the *wrong* location, children still look for the hidden item where the speaker says (Jaswal, Croft, Setia, & Cole, 2010). When given the option to choose between two potential informants, however, preschoolers can recognize which speaker is more accurate and prefer to trust that speaker (Pasquini, Corriveau, Koenig, & Harris, 2007), retaining this preference even after a time delay (Corriveau & Harris, 2009). In sum, children favor more reliable speakers when a choice is available, but they display a general bias to trust verbal information.

Inferences about others' actions

In nearly all of the work reviewed above, a parent, teacher, or experimenter presents the relevant information explicitly, via a demonstration or explicit utterance. But a parallel line of work suggests that children and even infants are able to make inferences about the *implicit* sources of both linguistic and non-linguistic actions. This literature is critical for motivating our hypothesis—that such inferences might not just inform guesses about particular agents' knowledge, abilities, preferences, or desires, but that they might also be a source of information about the world more generally.

By their first birthday, babies appear to make inferences about the unseen goals that underlie actions, even in very stripped-down displays. For example, they look longer when a shape that previously jumped over a wall toward another shape continues on the same path when the wall is removed (an action goal) instead of moving directly toward the shape (an end-state goal; Gergely, Nádasdy, Csibra, & Biro, 1995). This result is part of a broader body of work suggesting that infants expect agents to act rationally to achieve (inferred) outcomes most efficiently (Csibra & Gergely, 1998; Gergely & Csibra, 2003). In other words, very young children appear sensitive not

only to agents' particular actions, but also to the presumed purpose for these actions.

Young children also seem to be able to integrate information about constraints on knowledge and action into their inferences about goals. For example, infants can distinguish between actions that are produced intentionally (e.g. the choice of a particular ball by an agent viewing the contents of a box) versus randomly (by an agent wearing a blindfold; Xu & Denison, 2009). They can also reevaluate the likelihood of particular evidence when physical constraints make it more difficult for certain items to be selected (e.g., balls that stick to velcro, Denison & Xu, 2010). They can even infer that an agent demonstrates a preference by observing a pattern of choices that would be unlikely to occur by random selection (Kushnir, Xu, & Wellman, 2010).

Critical for our hypothesis here, some evidence suggests that young children can also work backwards from an agent's actions to infer generalizable knowledge about objects. In an experiment by Gweon, Tenenbaum, and Schulz (2010), fifteen-month-olds watched an experimenter pull a series of blue balls from a box and squeeze each toy to produce a squeaking sound. Babies were then handed a slightly different, yellow ball, and their generalizations about whether the new ball should also squeak were measured by their attempts to squeeze the toy. Depending on the evidence they saw, babies made different generalizations: If the blue balls were sampled by the experimenter from a box of mostly blue balls (implying that they were randomly sampled by the agent), they were more likely to think that a yellow ball would also squeak. But if they saw the blue balls picked out from a box of mostly yellow balls (where presumably the blue balls were less likely to be picked randomly, and thus were more likely to be intentionally selected for the demonstration of squeaking), they thought the yellow balls were less likely to squeak. In other words, children in this second condition made a general inference about the world (yellow balls don't squeak) based on a surprising thing that someone *didn't* do (did not pick out more common, yellow balls). The experiments we present below test a related type of inference in the linguistic domain; we therefore describe some of the background for these kinds of pragmatic inference.

Similar to the patterns of reasoning described above, listeners make pragmatic inferences in language comprehension by reasoning about the generating causes of a speaker's (linguistic) action and about the constraints on that action (Shafto et al., 2012). Just as babies form expectations about sampling likelihoods and infer that violations are intentional and informative (e.g. indicating others' preference or pedagogical demonstrations), they may learn to do the same for language, and make inferences about implicit, intended meaning when speakers' production choices differ from their expectations. Grice's (1975) maxims of cooperative communication—be truthful, informative, relevant, and clear—provide a framework for inferring implied meaning from linguistic evidence. For instance, Grice gives as an example a recommender who declares that his student has good penmanship. This recommender is choosing an action that completes his goal—write a letter that is maximally informative about the student—while complying with the restrictions on his actions—be truthful, don't say anything negative. On the basis of the above letter of recommendation, a reader of the letter can make the *implicature* that the writer does not believe the applicant has any other positive qualities (or else he would have mentioned them). A number of other theories have also attempted to describe the interplay between intention and production (Horn, 1984; Sperber & Wilson, 1986; H. Clark, 1996; Levinson, 2000), all preserving the basic idea of pragmatic inference as action understanding.

Since we are interested here in whether children can learn based on pragmatic inferences, we briefly review work on children's general pragmatic abilities. Initial work in this domain suggested that children might show deficits in pragmatic language understanding through age five or even later (Noveck, 2000; Papafragou & Musolino, 2003). When presented with a situation where three of three horses jumped a fence, children judged the utterance "some of the horses jumped over the fence" to be a felicitous description. While adults reject the statement under the assumption that "some" implies *some but not all*, children accept that "some" can overlap in meaning with "all". A more recent body of work, however, suggests that these deficits are specific to particular linguistic phenomena, namely scalar implicatures using quantifiers like "some" and

"all" (Barner, Brooks, & Bale, 2011; Katsos & Bishop, 2011).

One interpretation of failures in scalar implicature tasks is that children have trouble holding in mind and contrasting the alternative statements "some" and "all." This *alternatives hypothesis* makes a number of predictions that have now been tested. First, children have difficulties with even non-pragmatic interpretations that require considering these alternatives (e.g., reasoning about what "only some" means; Barner et al., 2011). Second, exposure to alternatives increases their success in computing implicatures (Skordos & Papafragou, 2014). Third, they perform better in contexts where alternative meanings are more explicit (e.g. pictured as alternatives in a forced-choice) (Miller, Schmitt, Chang, & Munn, 2005; Stiller, Goodman, & Frank, 2014). Thus, children appear to be more pragmatically sophisticated than some accounts have given them credit for. We appeal to the alternatives hypothesis below in describing the developmental pattern we observe in our experiments.

Our current study

Given that children are able to make a number of sophisticated inferences about the basis for both actions and utterances, we ask whether pragmatic inferences can provide a method for the transmission of information. In particular, we investigate preschoolers' ability to infer information about a general class from the specific word choices that a speaker makes in a description. For example, labeling a novel item as a "tall blicket" gives information not only that this item is a tall blicket, but it also suggests that height is a relevant variable property to blickets. Therefore, other blickets may be short.

We focus on adjectives as a case study. Because adjectives are optional modifiers, they can selectively be included in an utterance to draw contrasts between an intended referent and other unintended alternatives. Children show evidence of recognizing implied contrasts from prenominal adjectives (e.g. "red car") in their real-time language comprehension by age 3 (Fernald, Thorpe, & Marchman, 2010), and in more complex visual contexts by kindergarten (Nadig & Sedivy, 2002).

Work to date has focused on how adjectives are used to identify targets in referential communication tasks, however. Here we examine a novel question, asking how adjective use can help listeners infer *what the context is* that would lead a speaker to produce a partiular modified description. In other words, we ask whether children can infer that a contrastive description conveys not only information about the current referent, but also information about the implied property variability of other category members.

In the three experiments below, we test this hypothesis. In Experiment 1, we found that with a supportive framing, preschoolers made robust contrast inferences. In Experiment 2, performance decreased in a more stripped-down version of the task with reduced cues to contrast but was somewhat increased by pre-exposure to the appropriate alternative set, supporting a pragmatic interpretation of this behavior. In Experiment 3, children succeeded in a more open-ended production task, suggesting that they were able to summon to mind the relevant contrast dimension and not just select between alternatives.

Experiment 1

To investigate preschoolers' inferences about adjective use and category membership, we used a simple triad task. We introduced children to a novel shape, followed by two similar shapes: one that differed from the first only by size (e.g. small versus big), and the other that differed from the first only by a different polar feature (e.g. wet versus dry).[1] We described the first shape with either a size or feature adjective, and asked children to generalize what they thought other category members looked like. Children could follow at least two plausible strategies in this scenario. First, they could generalize from the exact property they heard, reasoning as follows: *You said, 'This is a special kind of tibu. This is a small tibu,' so most tibus are probably also small*. Second, they could

---

[1]In discussions of adjective semantics, the size adjectives we used are referred to as *gradable* adjectives because their meaning is relative to the head noun (Kennedy, 2012)—a small sofa is nevertheless bigger than a large mouse. In contrast, our alternative features were *non-gradable*—a sofa or a mouse exposed to water is equally considered wet. For convenience here and below, we refer to this distinction as "size" vs. "feature."

generalize from the property *dimension* they heard, reasoning instead that: *You said, 'This is a special kind of tibu. This is a small tibu.' If most tibus were small, you probably wouldn't have pointed out this one's size as special. So other tibus probably vary by size and can be big.* If children are sensitive to the pragmatic implications of speakers' choices, then they should take the latter route and infer that opting to include an adjective conveys an implied contrast with a set of alternatives, in this case other category members.
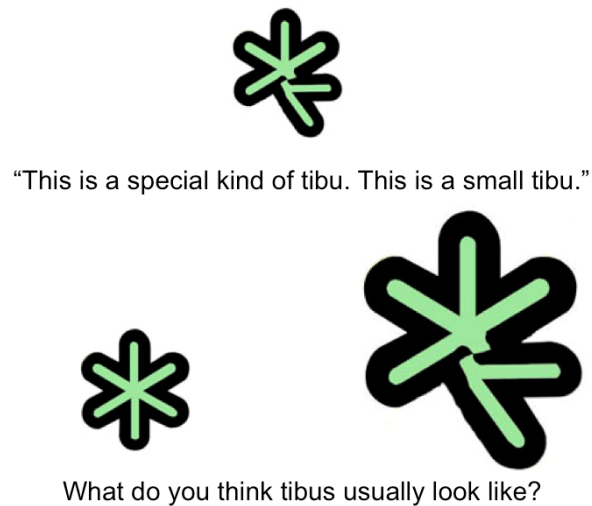
Methods

Participants. We recruited a planned sample of 96 children into four age groups: 3.0–3.5 years (n=24, mean age 3;3), 3.5–4.0 years (n=24, mean age 3;9), 4.0–4.5 years (n=24, mean age 4;3), and 4.5–5.0 years (n=24, mean age 4;8). Approximately half of the sample was recruited from the Bing Nursery School at Stanford University (n=52) and half was recruited from the Children's Discovery Museum (CDM) of San Jose (n=44); recruitment location was roughly even across age groups.

Children were tested individually in a quiet room at either the nursery school or the museum. At the CDM, parents accompanied their children and were asked to fill out a short demographic form about the child's language background. As a pre-specified selection criterion, only children who were reported to hear English at least 75% of the time were included in the final sample. Eight participants were excluded from analysis based on this criterion.[2] An additional two participants were excluded due to interruptions from family members during the testing session, and two were excluded for not completing all four experimental trials.

---

[2]In our partnership with the CDM, we invite any interested visitors to participate in our studies rather than prescreening children to meet our language requirements or to counterbalance all demographic factors (Callanan, 2012). Bing Nursery School is an English language preschool, and children included in the sample were fluent speakers of English. Children from Bing Nursery School and the CDM are demographically similar in terms of language exposure, ethnic backgrounds, and parental education, as reported by parents from each location. Both locations were mainly composed of educated middle class families. We tested for effect of location, and found no differences.

"This is a special kind of tibu. This is a small tibu."

What do you think tibus usually look like?

Figure 1. Example of a test trial for Experiment 1. Participants were introduced to a base shape (top) described with either a size or feature adjective. They were then shown two images, one that differed from the exemplar by a feature contrast (e.g. broken versus unbroken, left) and one that differed from the exemplar by size (e.g. small versus big, right) and were asked to point to which picture they thought was also a member of the same category.

We also recruited a comparison group of 128 adult participants through Amazon's Mechanical Turk online crowd-sourcing service. Participants all reported being native English speakers and residents of the United States. They were informed that the task was designed for children. Three participants were excluded for failing to complete the task.

Materials. We constructed the experiment as a storybook, illustrated with colorful clipart images. The book contained two training trials and four test trials. Each trial consisted of a novel shape (induction example) along with a pair of generalization stimuli: one that differed from the induction example only by size (e.g. small versus big), and one image that differed from the exemplar only by a feature contrast (e.g., broken versus unbroken; see example in Figure 1). Two

of the four trials used size adjectives and two of the trials used feature adjectives. Size terms were *small* (vs. big), *long* (vs. short), *tall* (vs. short), and *short* (vs. long); feature contrasts were *broken* (vs. unbroken), *pointy* (vs. smooth), *dirty* (vs. clean), and *wet* (vs. dry). To ensure that children were familiar with the words we used, we included a posttest of two-alternative displays. Children were able to recognize all of the contrasts used in our task, with 90% accuracy for 3–3.5 year-olds, 95% for 3.5–4.0 year-olds, 96% for 4.0–4.5 year-olds, and 98% for 4.5–5.0 year-olds.

Procedure. The experimenter read the storybook with children individually in a quiet room at either Bing Nursery School or the CDM. At the museum, parents accompanied children and sat either next to or behind the child. Siblings were sometimes also present, and were offered quiet activities such as coloring or reading.

To begin the book, children were introduced to a character named Allen the Alien who was visiting planet Earth. Children then participated in two training trials containing familiar items to teach Allen about some things on Earth and get children used to the study design. Training trials featured adjectives other than those used in critical trials, and training pictures displayed only one relevant contrast choice. For example, children were shown a picture of chocolate milk followed by two pictures, one of plain milk and one of orange juice. Children were told, "This is a special kind of milk. This is *chocolate* milk. What does milk usually look like? What does most milk look like?" and prompted to point to the picture. On the rare occasion that children answered incorrectly, the experimenter repeated the statements and encouraged children to point to the correct picture until they answered correctly.

After the training trials, children participated in four test trials. For each test trial, children were shown a picture of an induction example and told something about it, e.g. "This is a special kind of tibu. This is a small tibu." They were then shown two similar pictures, one that differed from the exemplar only by size (e.g., a big broken tibu) and one that differed from the exemplar only by a feature (e.g., a small unbroken tibu), and were asked "What do you think tibus usually look like? What do you think most tibus look like?" They were prompted to select one of the

non-exemplar images.

Participants were assigned to one of two lists, counterbalanced for adjective type and picture order. Adjectives were focused using contrastive stress. The experimenter averted her gaze while children pointed to their responses. Responses were coded online and double-coded offline using a video recording of the testing session. The task took about ten minutes to complete.

The task was adapted to an online format for adult participants. Adults viewed a single trial composed of one of the picture triads and read the same text that was spoken to children. We chose to use only a single trial for adults to avoid the possibility of task demands caused by repeating the same type of inference (Frank & Goodman, 2012). Picture type, side, and adjective were counterbalanced across participants. Adults indicated their response using a radio button below their image selection. Participants were paid 25 cents for completing the task, which took about two minutes to complete.

Results and discussion

Preschoolers' ability to make correct contrast inferences increased across the age range we tested (Figure 2).[3] We categorized a response as correct—representing what we will call a *contrast inference*—if children selected the item that differed from the exemplar along the referenced dimension (e.g., they chose the short item if the exemplar was referred to as "tall," and the clean item if it was described as "dirty"). The youngest children in our sample (age 3.0–3.5 years) were marginally above chance ($t(23) = 1.84$, $p = .08$) in their contrast inferences across adjective types, and all other age groups were above chance (all $ps < .01$). These results indicate that preschoolers are sensitive to adjective use as conveying contrast information about a relevant property dimension.

To measure differences across adjective types and age groups, we used a logistic mixed effect model, predicting correct responses as the interaction of age and contrast type, with random

---

[3]Data and analysis code can be found at `http://github.com/ahorowit/aliens`.
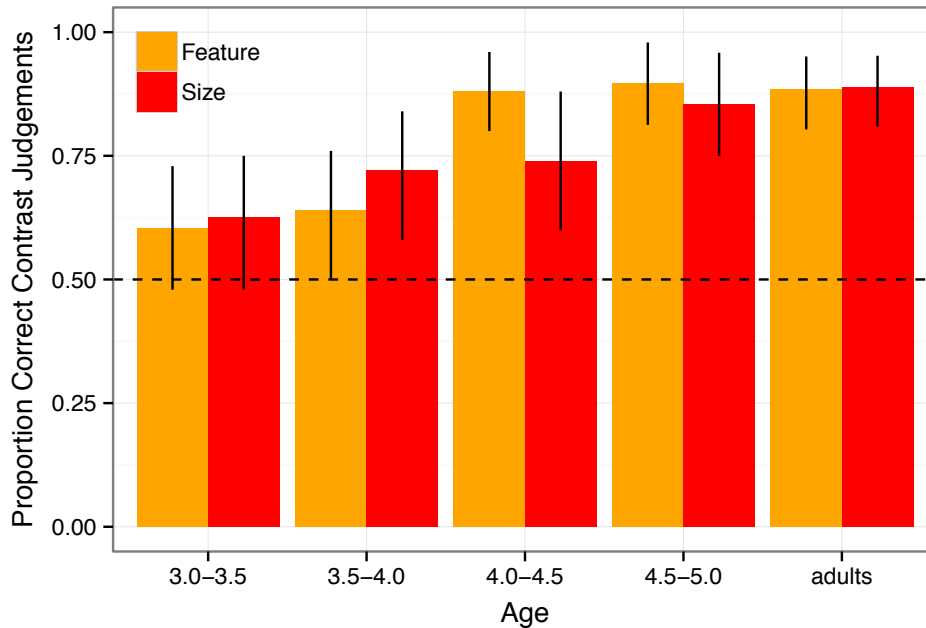
Figure 2. Mean proportion correct contrast judgments for preschoolers and adults in Experiment 1. Yellow bars depict feature adjective trials and red bars depict size adjective trials. Dashed line represents chance; error bars show 95% confidence intervals computed by non-parametric bootstrap.

intercept and slope (reflecting contrast type) for each participant. Children made increasingly more correct contrast judgments with age ($\beta = 1.52$, $p < .0001$). There was no significant effect of contrast type (feature vs. size adjectives), and there was no interaction between age and contrast type, suggesting that participants across ages did not differ in their responses to different property types. Overall, these analyses show that children demonstrate an increasing sensitivity to implicit contrast information from adjectives.

Experiment 2

In Experiment 1, we provided a supportive scaffolding to help children recognize that adjectives were being used contrastively: We told them explicitly that "This is a special kind of

tibu." In Experiment 2, we removed this contrastive framing to test whether the older children (who succeeded handily in Experiment 1) could still make contrast inferences based on the presence of a contrasting adjective alone, a far more subtle cue. Our revised framing was "This is a tibu. This is a small tibu."

We hypothesized that this *adjective only* condition would make the contrast inference substantially more difficult. Previous work on pragmatic inference has suggested that one major problem for preschool children in making inferences about contrasting terms is summoning to mind alternative terms that could have been used (e.g. that "some" is a weaker alternative to "all"; Barner et al., 2011). For this reason, we attempted to alleviate this burden in our task by providing children with pre-training on the relevant contrasts of interest. In this *alternatives pre-exposure* condition, we read children a book highlighting polar opposites prior to the experimental task to remind children that, for example, "dirty" is an alternative to "clean." We predicted that the increased experience comparing adjective alternatives in this condition would help support children's ability to make contrast inferences at test.

Methods

Participants. A new sample of 96 children was recruited from Bing Nursery School. Because of the presumed increased difficulty of this task, we recruited children from the older age groups: 4.0–4.5 years (n=48, mean age 4;4) and 4.5—5.0 years (n=48, mean age 4;9). Half of the children in each age group (24 younger 4s and 24 older 4s) were randomly assigned to each of the conditions (*Adjective Only* and *Alternatives Pre-Exposure*). Four children were excluded for not completing all four trials of the task.

We also ran a new group of 128 adult participants in the *Adjective Only* condition on Amazon's Mechanical Turk. All participants were reported to be US residents and native English speakers. They were informed that the task was designed for children. Seven were excluded for failing to complete the task.
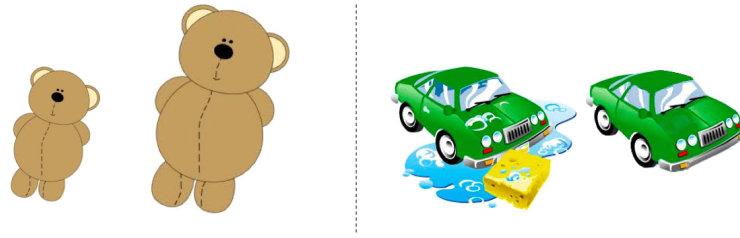
Figure 3. Sample images from the Alternatives Pre-Exposure book used in Experiment 2. Left: an example size contrast (small teddy bear versus big teddy bear). Right: an example feature contrast (wet car versus dry car).

Materials. Stimuli were identical to Experiment 1. In the Alternatives Pre-Exposure condition, participants read a book prior to the testing procedure. The book consisted of clip art images of familiar items depicting the size and feature contrasts portrayed in the test book. Opposites were paired so that scalar contrasts were viewed simultaneously and stated consecutively (e.g. "Here is a small teddy bear. Here is a big teddy bear."). Sample images are presented in Figure 3.

Procedure. Procedures for the experimental task were identical to Experiment 1 with the exception that the referential phrase was minimized by removing the phrase "special kind of" to reduce contrast cues other than the adjective. Instead, participants heard only "This is a [tibu]. This is a [broken tibu]." Children in the Alternatives Pre-Exposure condition were told that they would be reading two books in the experimental session. The experimenter read the adjectives book with children, labeling each picture in a neutral way on each page (e.g. "Here is a small teddy bear. Here is a big teddy bear.") before reading the test book. As in Experiment 1, adult participants were randomly assigned to a single test trial presented online.
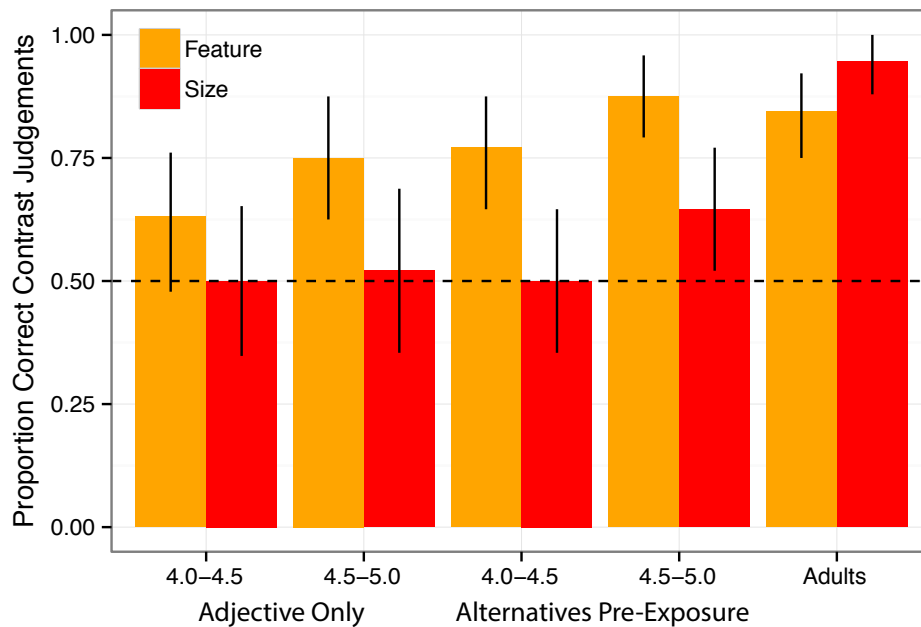
Figure 4. Preschoolers' and adults' proportion correct performance for the *Adjective Only* and *Alternatives Pre-Exposure* conditions in Experiment 2. Adults did not see the pre-exposure trials. Dashed line shows chance performance; error bars show 95% confidence intervals.

Results and discussion

With the removal of supportive language indicating contrast, performance in Experiment 2 was substantially lower than in Experiment 1. Contrast selections in size trials were especially low, while contrast judgements in feature trials remained higher. In the Adjective Only condition, averaging across adjective types, 4.0-year-olds were not above chance ($t(22) = 1.10$, $p = .28$), but 4.5-year-olds were ($t(23) = 2.18$, $p = .04$). Breaking performance down by adjective type, on feature trials the younger 4s were marginally above chance, while the older 4s were substantially above chance ($t(22) = 1.82$, $p = .08$ and $t(23) = 3.71$, $p = .001$). Both groups' performance did not differ from chance for size adjectives. Thus, without the contrastive language in Experiment 1,

children had a substantially harder time but the oldest children could still make contrast inferences at above-chance levels for feature adjectives.

One potential source of the asymmetry between feature and size adjectives could be due to the relatively greater contrast implied by our featural adjectives. Saying that something is "broken" almost always implies a changed state from having been unbroken at another point in time. In contrast, saying something is "small" can imply that there are bigger others—but it can also simply reflect some sort of general, stable comment on size. If this ambiguity about the contrastiveness of the size adjectives was the source of the lowered performance in the Adjective Only condition, the Alternatives Pre-Exposure condition might increase performance for size adjectives by virtue of highlighting the contrastive use of alternative size terms.

Congruent with this hypothesis, the alternatives pre-exposure appeared to boost performance somewhat. Children in the Alternatives Pre-Exposure condition were above chance at both ages ($t(23) = 2.33$, $p = .03$ and $t(23) = 6.33$, $p < .0001$), aggregating across adjective types. Breaking down by adjective types, the younger 4s were above chance for feature but not size adjectives ($t(23) = 4.51$, $p = .0001$ and $t(23) = 0$, $p = 1$). Older children were above chance on both ($t(23) = 8.31$, $p < .0001$ and $t(23) = 2.07$, $p = .05$). Nevertheless, no pairwise tests between the Adjective Only and Alternatives Pre-Exposure conditions reached significance.

We again analyzed our results using a logistic mixed model. We initially fit a model that included all interactions of age, adjective type, and book type but found that no effects reached significance; indeed, this model did not increase fit over a model that only included main effects ($\chi^2(4) = 2.47$, $p = .65$). This main-effects only model included a significant effect of adjective type such that they made fewer contrast selections on size trials than feature trials ($\beta = -1.08$, $p < .0001$). It also included marginal effects of age and book type ($\beta = 1.04$, $p = .07$ and $\beta = .50$, $p = .09$, respectively), indicating that older children performed better than younger children, and that performance was higher in the Alternatives Pre-Exposure condition than the Adjectives Only condition.

In sum, Experiment 2 provides support for the hypothesis that children's judgments in Experiment 1 were strengthened by the presence of the contrastive language "a special kind of." When we removed this language, performance dropped markedly, especially for size adjectives (which might not be as clearly contrastive as feature adjectives). We also saw some signs that pre-exposure to a storybook that used the target adjectives contrastively improved performance, providing additional support to the idea that it was specifically the recognition of pragmatic contrasts that led children to make the appropriate generalization.

One possible alternative explanation for our findings is that contrast inferences were in fact not warranted by the simpler adjective framing (rather than being warranted but not being identified by children). Ruling out this explanation, the performance of adults in Experiment 2 (without supportive language) was essentially identical to that of adults in Experiment 1. Thus, we believe that the contrastive framing in Experiment 1 helped children differentially in interpreting the contrast inference implied by the use of the adjective features.

Experiment 3

In Experiment 3, we tested the extent of children's inferences from contrastive adjective use by examining their performance in a free-response task. One possibility in Experiments 1 and 2 was that children in fact did not infer contrast immediately from the linguistic frame and adjective, but instead recognized that the two-alternative forced choice task required a contrastive interpretation of the adjective. A free response task circumvents this issue by testing children's interpretation of the concept without asking them to choose between alternatives. For the linguistic framing in this experiment, we chose an intermediate level of support for contrast; less extreme than Experiment 1 but more supportive than Experiment 2: we told children that "This is a plizzle. There are different kinds of plizzles. This one is a small plizzle."

Methods

Participants. We recruited a new planned sample of 24 4-year-old children (mean age 4;6) from the CDM. Only children whose parents reported that they heard English at least 75% of the time were included in the final sample. Two participants were excluded for using less than this amount of English, and one participant was excluded for not producing responses to the experimenter's questions.

Materials. We used a similar task design as Experiments 1 and 2, but showed children only a single picture rather than a triad. In addition, because some of the original items depicted contrasts in which one pole was visually salient but perhaps difficult for children to name (e.g. "broken" vs. "unbroken"), we modified the set of items we used. The named size contrasts used were *small* (vs. big), *tall* (vs. short), *long* (vs. short), and *skinny* (vs. fat). The named feature contrasts were *hot* (vs. cold), *dark* (vs. light), *wet* (vs. dry), and *open* (vs. closed). We also included a post-test to ensure that children were familiar with all of the properties used in the task. Children successfully identified pictures that corresponded with the meanings of the adjectives in 96% of trials.

Procedure. The experimenter read the storybook with children individually in a quiet room at the CDM. As before, children were introduced to Allen the Alien and then participated in two training trials with familiar items. Unlike in Experiments 1 and 2, children saw only a single image per trial. For example, in a training trial, children were shown a picture of a heart-shaped cookie and told, "This is a cookie. There are different kinds of cookies. This one is a heart-shaped cookie. What do other cookies look like?" Most children answered immediately that most cookies are *round* or *circle*-shaped. A few children were slower to respond, and were promoted to think again what most cookies look like. If they still did not respond, they were asked what shape most cookies are. If children provided an answer other than shape, they were encouraged to think about the description again (e.g. "You're right, cookies can be different flavors. Did you see this one is heart-shaped? What do most cookies look like?").

Following the two training trials, children participated in four test trials. For each test trial, children were shown a picture of a single exemplar and told something about it, e.g. "Wow, this is a plizzle. There are different kinds of plizzles. This one is a small plizzle. What do you think other plizzles look like?" Their verbal responses were recorded. Two of the test trials referred to size adjectives (e.g. *small*), and two of the trials referred to feature adjectives (e.g. *hot*). The order of trial items varied across two lists, each of which was counterbalanced for adjective type and picture order. Adjectives were focused using contrastive stress. Responses were coded online and double-coded offline using a video recording of the testing session. The task took about ten minutes to complete.

Results and discussion

Despite the open-ended nature of the task, children gave contrastive responses more than half of the time (57% and 64% overall for feature and size, Figure 5). We coded responses as either an *Exact contrast* (e.g. hearing "small" and saying "big") or as an *Other contrasting term* if they were an approximate contrast to the named property (e.g. hearing "small" and saying a size property other than "big", e.g. "tall"). Matching, non-contrastive descriptions (e.g. hearing "small" and repeating "small") were not included in the approximate contrast group. More than a third of productions were exact opposites (35% for feature terms and 39% for size terms), and another quarter were non-exact contrasts but related to the stated property information (22% for feature terms and 25% for size terms). There were no differences in the proportion of response scores across feature and size trials. Thus, Experiment 3 provides evidence that, even without seeing a contrastive alternative test item, children were able to spontaneously generate appropriate descriptions based on the adjective the speaker used.

General Discussion

If a speaker describes an "art museum," can children learn that there are other types of museums? And if children hear someone described as a "female scientist" or "male librarian," will
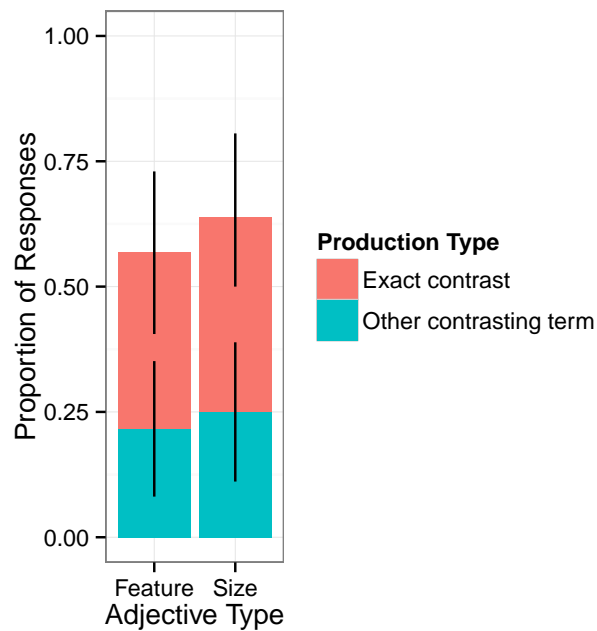
Figure 5. Four-year-olds' free response descriptions of other category members upon hearing a feature adjective (left) or size adjective (right). Productions were coded as being *Exact contrasts* if they were opposite the description and as being *Other contrasting terms* if they were related to the referenced property but neither a direct contrast nor an exact match. Error bars show 95% confidence intervals.

they make (potentially harmful) inferences about gender-typical roles? Our findings support the idea that children are indeed sensitive to contrasts of this type. In our experiments, they were able to learn from not just the literal content of a speaker's utterance, but from the choices she made in expressing that content; and they did so most effectively in cases where the particular contrast of interest was made salient.

In Experiment 1, preschoolers reliably made contrast inferences from size and feature adjectives with the support of a contrastive linguistic framing. In Experiment 2, we found that

unlike adults, 4-year-olds had difficulty making these inferences when the supportive framing was removed. Their performance increased somewhat after exposure to the relevant adjective contrasts, however, suggesting that a number of cues to the contrastiveness of a particular adjective could contribute to their ability to make inferences. Finally, in Experiment 3, which used a free-response task, four-year-olds were able to generate the relevant property contrasts (without seeing them pictured in the two-alternative-forced choice we used in the other experiments). Taken together, our results suggest that preschoolers can reason about speakers' adjective choices as providing information not only about the particular entity being described at the time, but also about other category members.

Although the design of our task was simple, it nevertheless required a counterintuitive response: hearing a description and inferring that its opposite was represented in the broader set. In other words, children needed to process the stated adjective and then select the picture that *differed* from that description instead of the one that shared that property, even though both picture options were available. The task thus involved both pragmatic reasoning about how people use language, and also the exercise of executive function in inhibiting the bias to match the description directly at test. We found that preschoolers nonetheless made contrast inferences, though performance was stronger in older children; the development of executive function in this age range provides a plausible source for the developmental effects we observed (Davidson, Amso, Anderson, & Diamond, 2006; Zelazo et al., 2003).

Our data speak to children's ability to learn one particular piece of information from pragmatic language use: the typical property for exemplars of a category. We selected this particular example because the generalization of category structure from individual exemplars is a key problem for children (Markman, 1991). But there are potentially many different inferences that could be made on the basis of the same sort of optional modification. As in the case of "art museum," sometimes a contrastive modifier does not license specific inferences about what is typical of a category (i.e. if "art" does not have an opposite, what can we infer about other

museums?) but only that there is some important variability in a particular dimension (there are other types of museums other than art museums). The pragmatic and discourse context of an utterance can also affect the kind of inference that is licensed by a contrastive modifier: Depending on context, labeling someone as a "good student" can imply that others in the comparison set are either better (where the student is implied to be "merely good") or worse (where the student is "very good"). Future work should investigate a broader range of cases of pragmatic learning, from other kinds of world knowledge to the idiosyncrasies of social judgement that are implied by descriptive language.

Our experiments contribute to the growing literature suggesting that children consider how and why evidence is generated to reason about the social world, in both non-linguistic and linguistic contexts. As reviewed above, even young children robustly infer probabilities from random sampling while making inferences about social preferences or generalizable knowledge from conspicuous non-random sampling (Xu & Denison, 2009; Kushnir et al., 2010; Butler & Markman, 2012). In the domain of language, preschoolers are beginning to make similar (pragmatic) inferences about the motivations for language use (e.g. Stiller et al., 2014; Katsos & Bishop, 2011), though many factors may constrain their ability to succeed in more complex situations (Papafragou & Musolino, 2003; Barner et al., 2011).

To explain failures of pragmatic reasoning, Barner et al. (2011) proposed the *alternatives hypothesis*—that children's ability to make inferences is often constrained by their ability to bring to mind and contrast the relevant inferential alternatives. In the case of scalar implicature, these alternatives can be quantifiers like "some" and "all"; in our case, they were pairs of polar adjectives. In both cases, the pragmatic inference is supported by an appreciation of what *wasn't* said. Our pre-exposure manipulation in Experiment 2 provides some preliminary evidence that the alternatives hypothesis explains children's difficulty in our task. An earlier study we conducted with color adjectives also supports this hypothesis (Horowitz & Frank, 2012). Four-year-olds in this study were considerably more proficient in making contrast inferences with size adjectives

than with colors. One explanation for the difficulty that children had with colors was the lack of an obvious alternative: While a "*small* tibu" might be exceptional relative to the larger, more typical tibus, a "*blue* tibu" less clearly implies what color other tibus are (other than not being blue). Although the stimuli in this earlier experiment were different than those used in the current study and hence direct numerical comparisons between levels of performance are not possible, it nevertheless provides additional circumstantial evidence about the role of pragmatic alternatives in allowing such inferences.

Most work investigating children's pragmatic abilities has focused on reasoning about speakers' intended meanings. In contrast, we examined children's inferences about the state of the world that would lead a speaker to make particular production choices. While preschoolers show evidence of learning generalizable knowledge from specific descriptions based on framing cues (Cimpian & Markman, 2009), our work suggests an additional pragmatic route to such general knowledge. In this way, we connect the mechanisms of pragmatic inference with processes of social learning and generalization. If children assume that speakers are communicating pragmatically, then they can take advantage of opportunities for learning wherever they recognize a speaker's choice to produce an utterance in one form over another.

References

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84–93.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416–1428.

Callanan, M. A. (2012). Conducting cognitive developmental research in museums: Theoretical issues and practical considerations. *Journal of Cognition and Development*, *13*(2), 137–151.

Cimpian, A., & Cadena, C. (2010). Why are dunkels sticky? Preschoolers infer functionality and intentional creation for artifact properties learned from generic language. *Cognition*, *117*(1), 62–68.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to childrens biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14–25.

Cimpian, A., & Scott, R. M. (2012). Children expect generic knowledge to be widely shared. *Cognition*, *123*(3), 419–433.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition*, *1*, 33.

Clark, H. (1996). *Using language* (Vol. 4). Cambridge University Press Cambridge.

Corriveau, K., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*(1), 188–193.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255–259.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of

memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

Denison, S., & Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science*, *34*(5), 885–908.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: an empirical examination. *Developmental Psychology*, *41*(3), 451.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: a pragmatic account. *Developmental Psychology*, *37*(5), 630.

Fernald, A., Thorpe, K., & Marchman, V. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*(3), 190–217.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Ganea, P. A., Shutts, K., Spelke, E. S., & DeLoache, J. S. (2007). Thinking of things unseen infants' use of language to update mental representations. *Psychological Science*, *18*(8), 734–739.

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, *74*(1), 308–325.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.

Gergely, G., Nádasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623–1627.

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard
    University Press.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based
    implicature. *Meaning, Form, and Use in Context*, *42*.

Horowitz, A., & Frank, M. C. (2012). Learning from speaker word choice by assuming adjectives
    are informative. In *Proceedings of the 34th Annual Conference of the Cognitive Science
    Society.*

Horowitz, A., & Frank, M. C. (2014). Preschoolers infer contrast from adjectives if they can
    access lexical alternatives. In *Proceedings of the 36th Annual Conference of the Cognitive
    Science Society.*

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific,
    highly robust bias to trust testimony. *Psychological Science*.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of
    informativeness and implicature. *Cognition*, *120*(1), 67–81.

Kennedy, C. (2012). Adjectives. In G. Russell & D. G. Fara (Eds.), *Routledge Companion to
    Philosophy of Language.* Routledge.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the
    preferences of other people. *Psychological Science*, *21*(8), 1134–1140.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, *117*(1), 1–47.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational
    implicature*. MIT Press.

Markman, E. M. (1991). *Categorization and naming in children: Problems of induction*. Mit
    Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the
    meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech

in communication at 12 months. *Cognition*, *123*(1), 50–60.

Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29th annual boston University Conference on Language Development* (pp. 389–400).

Nadig, A., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329.

Noveck, I. A. (2000). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165–188.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*, 253–282.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216.

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, *43*(3), 696.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others the consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351.

Skordos, D., & Papafragou, A. (2014). Scalar inferences in 5-year-olds: The role of alternatives. In *Proceedings of the 38th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Harvard University Press Cambridge, MA.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2014). Ad-hoc implicature in preschool children. *Language, Learning & Development*.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of*

Sciences, *109*(32), 12933–12937.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*(1), 97–104.

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., et al. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, i–151.