Knowledge transmission through pragmatic inference in preschool children

Alexandra C. Horowitz

Department of Psychology, Stanford University

Michael C. Frank

Department of Psychology, Stanford University

Special thanks to the staff and families at the Bing Nursery School and the Children's Discovery Museum of San Jose and to Octavia Zahrt for assistance with Experiment 3. This work supported by a John Merck Scholars Fellowship and ONR grant N00014-13-1-0287. Earlier versions of this work were presented to the Cognitive Science Society in Horowitz and Frank (2012) and Horowitz and Frank (2014).

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

Children learn many instances of cultural norms through direct instruction (e.g. “red lights mean stop”), but not all conventional knowledge is stated explicitly. We investigate the hypothesis that children infer generalizable information about the world pragmatically, by making inferences from speakers’ descriptive word choices. We introduced preschoolers to picture triads: an exemplar of a novel category as the base, and then two options that varied from it either by size or by a binary feature. We described the exemplar contrastively, using either a size term or another polar adjective (e.g., “this is a [small/broken] tibu”), and asked children which of the two pictures they thought other category members looked like. In Experiment 1, we used a supportive framing that provided an additional signal that the adjective was contrastive and found that children reliably inferred the typical category member. In Experiment 2, children’s performance was reduced when the contrastive framing was removed, but partially restored via a pre-exposure to the relevant adjective pairs. In Experiment 3, a free-response task, children spontaneously produced relevant property contrasts for both size and feature terms. Our findings suggest that preschoolers can learn about a novel category from a single exemplar—and not just about what that category is, but also what it isn’t. More generally, sensitivity to why speakers choose to describe the world the way they do may allow children to learn efficiently about the world around them.

Keywords: pragmatics; language development; adjectives; cultural learning

Introduction

Children learn some important cultural 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 norms are stated directly. Sometimes information is implicit in *how* a statement is made. For example, if a parent says, "that's a salad fork," he 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 test this hypothesis using a simple case study: learning to generalize novel words via minimal contrastive descriptions. We focus on contrastive word choices, as in the above 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.

While some language describes the particulars 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: They 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 a broader set of 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).

In fact, 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 telling children the *wrong* location consistently, 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).

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 underly 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 flexibly into their inferences about goals. For example, infants can distinguish between actions that are taken 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 the agent has a preference after 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 saw a series of blue balls squeezed to make a squeaking sound and then were presented with the opportunity to generalize by squeezing a slightly different yellow ball. Depending on the evidence they saw, the children made different generalizations. If the blue balls were sampled by the experimenter from a box of blue balls (implying that they were randomly sampled by the agent), the children more likely to think that a yellow ball would also squeak. But if the children 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 (pick out yellow balls). The experiments we present below test a very 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, in making pragmatic inferences in language comprehension, listeners reason about the generating causes of a speaker's (linguistic) action and about the constraints on that action (Shafto et al., 2012). In Grice's (1975) classic example, the recommender who declares that his student has good penmanship 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. Grice famously formalized these tradeoffs in reasoning about linguistic actions as a series of maxims—be truthful, informative, relevant, and clear—and derived a number of important linguistic phenomena from their intersection. For example, on the basis of the above letter of recommendation, a reader of the letter can make the *implicature* that the speaker does not believe the applicant has any other positive qualities (or else the letter writer would have mentioned them). Since Grice's initial formulation, a number of other theories have described the exact tradeoff that governs speakers' actions differently (Horn, 1984; Sperber, Wilson, He, & Ran, 1986; H. Clark, 1996; Levinson, 2000), even as they have preserved the basic idea of pragmatic inference as action understanding.

Are children able to make such inferences as well? Initial work suggested that they might show deficits in this kind of 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 would judge the utterance "some of the horses jumped over the fence" to be a felicitous description. The necessary pattern of reasoning to make the implicature in this case would be that if the speaker said "some," she probably couldn't have truthfully uttered the stronger term "all." A more recent body of work suggests that these deficits are specific to particular linguistic phenomena—in particular, 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 levels of implicature (Skordos & Papafragou, 2014). Third, in contexts where alternative meanings are more explicit (e.g. pictured as alternatives in a forced-choice), children perform better (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, 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 choice 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, and that other blickets may be short.

We focus on adjectives as a case study. Because adjectives are optional modifiers, they can be included in an utterance to draw contrasts between an intended referent and other unintended alternatives. Children show evidence of making contrastive use of prenominal adjectives (e.g. "red car") in their real-time language comprehension by age 3 (Fernald, Thorpe, & Marchman, 2010), and in more complex referential communication task by kindergarten (Nadig & Sedivy, 2002). Work to date has focused on how adjectives are used in context, however. In contrast, here we examine a novel question, asking how adjective use can help listeners infer *what the context is*. In other words, we ask whether children can infer that a contrastive description conveys not only information about the referent, but also information about the properties 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 category feature 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, and the other that differed from the first only by a different polar feature (e.g. wet/dry).[1] We described the first shape contrastively using either a size or feature adjective, and asked children to generalize what they thought other category members looked like. If children generalize the category without taking into account contrast, then they should expect other category members to *match* on that property (e.g., hearing "small tibu" and selecting that tibus are usually small). If they infer the category structure based on a contrastive interpretation, they should expect other category members to *mismatch* (e.g., hearing "small tibu" and selecting that tibus are usually larger).

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.

[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 tree is nevertheless bigger than a large mouse. In contrast, our alternative features were non-gradable—a wet sock is likely to be as wet as a wet dog. For convenience here and below, we refer to this distinction as "size" vs. "feature."

Children were tested individually in a quiet room at either the nursery school or the museum. At the CDM, parents accompanied children and sat either next to or behind the child, and children received a sticker and a certificate for their participation. If siblings also attended, they were provided quiet toys or activities such as coloring or reading during the testing prodecure. Parents were asked to fill out a short demographic form about their children'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. [FILL IN NUMBER] were excluded from analysis based on this criterion.[2] An additional [FILL IN NUMBER] were excluded for not completing all four experimental trials, and [FILL IN NUMBER] were dropped from analysis due to interference from parents or siblings interrupting the study procedure.

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. Each 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, 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.

[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 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.

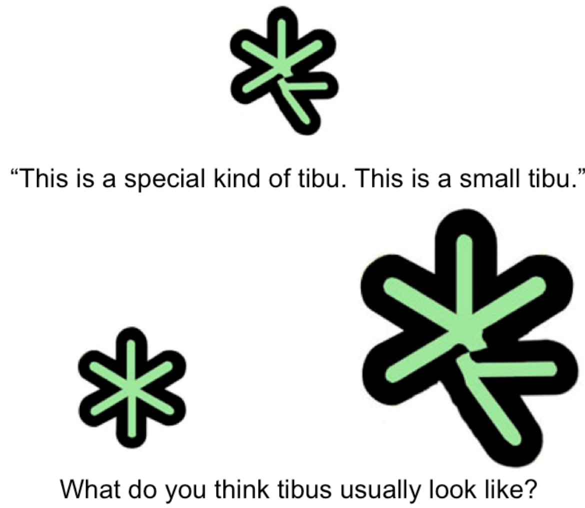

Figure 1. Example of a test trial for Experiment 1a. Participants were introduced to a base exemplar shape (top) described with either a size or feature adjective. They were then shown two images, one that differed from the exemplar by a feature (left) and one that differed from the exemplar by size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

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% 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 of the San Jose Children's Discovery Museum. 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 fixed 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 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

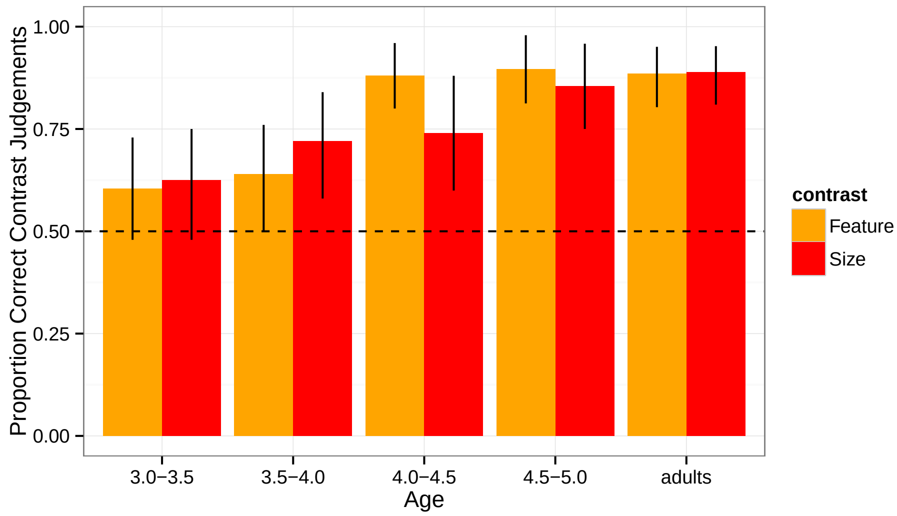

Figure 2. Mean proportion correct for preschoolers and adults in Experiment 1. Yellow bars depict feature adjective trials and red bars depict size trials. Dashed line represents chance; error bars show 95% confidence intervals computed by non-parametric bootstrap.

their image selection. Participants were paid 25 cents for completing the task, which took about two minutes.

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 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," but the clean item if it was described as "dirty"). The youngest children in our sample were numerically above chance in their responding across categories ($M = .61$) but were only marginally above chance, consolidating

[3]Data and analysis code can be found at `http://github.com/ahorowit/aliens`.

across adjective types ($t(23) = 1.84$, $p = .08$). All other groups were above chance (all ps $< .01$).

To measure differences across adjective types and age groups, we used logistic mixed effect model, predicting correct responses as the interaction of age and contrast type, with random effects of participant and item. 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 (Barner et al., 2011). For this reason, we attempted to alleviate this burden by providing children with pre-training on the relevant contrasts of interest. In the *alternatives pre-exposure* condition, we read children a book highlighting featural opposites prior to the experimental task.

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

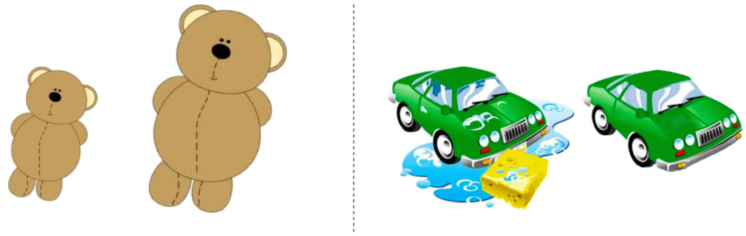

Figure 3. Sample images from the adjective naming book used in Experiment 3.

groups: 4.0–4.5 years (n=48, mean age FIXME) and 4.5—5.0 years (n=48, mean age FIXME). 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*).

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.

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 teddybear. Here is a big teddybear.") 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

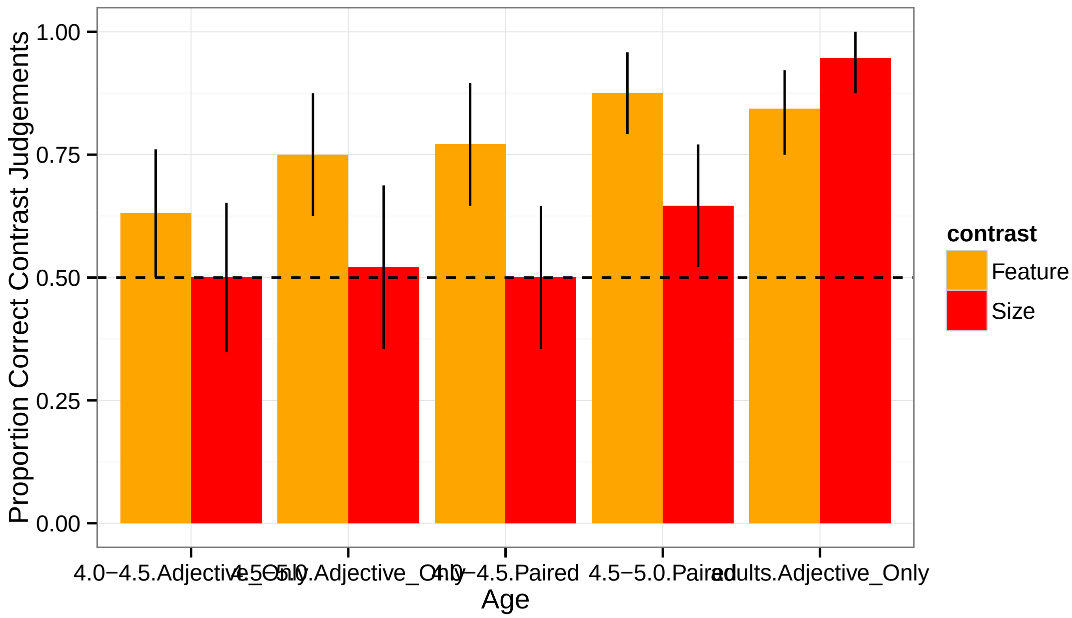

Figure 4. Preschoolers' and adults' proportion correct performance for the *adjective only* and *alternatives pre-exposure* conditions in Experiment 2. Dashed line shows chance performance; error bars show 95% confidence intervals.

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 teddybear. Here is a big teddybear."). As in Experiment 1, adult participants were randomly assigned to a single test trial presented online.

Results and discussion

Although preschoolers showed increasing contrast selections from adjective use with age in Experiment 1b, they were essentially at chance when the contrastive language framing was removed. We analyzed our results using a logistic mixed model, predicting correct responses as an interaction between age and contrast type with random effects of participant and shape, and we found no significant effects and no significant interaction. In post-hoc followup tests, older 4s showed a significant feature contrast bias ($p = .001$, exact binomial test), but this contrast was not reliable in the full model when controlling for participant and item effects, and may have been driven primarily by the “broken” and “clean” items. Although adults remained attentive to implicit contrast information in both the contrastive language and adjective only framings, children performed substantially worse without the additional linguistic cues to guide their contrast judgements.

UPDATE THIS INFO WITH LAST FEW KIDS

As above, we measured the proportion of correct contrast judgments for which participants selected the test picture that differed along the referenced property dimension. Adults performance was significantly about chance ($p < .001$ in exact binomial tests for feature and size terms) and did not differ by adjective type. They showed only a slight decrease in performance in this adjective only framing from the contrastive language framing in Experiment 1c (see Figure 4). These results indicate that adjective use in our task is a strong indicator of relevant property information of novel category members for adults. Their nearly equal performance across Experiment 1a and 2a suggests that adjectives provided salient cues to implicit contrast dimensions on their own without the necessity of additional semantic support.

Increasing preschoolers’ access to relevant linguistic alternatives helped older 4s select property contrasts for both feature and size adjectives. These results suggest that supporting children’s abilities to bring relevant alternatives to mind plays a strong role in their pragmatic inferences. Beyond relying on rich semantic framing cues to intended meaning, which are not always available in natural speech, reminding children of different types of modifiers increases their likelihood of forming contrast inferences from adjectives alone.

Older children selected the contrast property for both feature and size terms more often than chance ($p < .001$ and $p < 0.01$ respectively in exact binomial tests) and younger children for feature terms ($p = .01$ in exact binomial test), though younger children’s performance did not differ across feature and size trials. A logistic mixed model predicting correct responses as an interaction between age and contrast type with random effects of participant and shape revealed no significant effects or interaction, however.

Our results from Experiment 2 suggest that exposing children to a book of unrelated pictures with the scalar alternatives used in our test trials helped older children to select opposites more consistently, without any framing cues. An alternative hypothesis, that the initial book served to train children to always select named opposites, is not supported because we did not see a change in performance for the youngest children. In addition, anecdotally none of the children remarked on any relationship between the books, even though they conveyed the same adjective properties. Instead, we believe that the opposites book served to make the lexical scales more accessible to children so that, at least for the oldest children in our task, they could spontaneously infer implicit contrast information from an adjective produced alone.

Experiment 3

In Experiments 1 and 2, we found that preschoolers can make contrast inferences from adjective use, but that they may rely on the support of contrastive linguistic framing or a reminder of the contrastive nature of modifiers before recognizing the informativeness of these word choices on their own as adults do. Although we were able to measure developmental differences in our first experiments, our results are limited to childrens selections in a forced-choice procedure. In Experiment 3, we wanted to test the extent of childrens more natural responses to adjective use. We modified our design to examine childrens productions in a free response version of the task. We told children that, e.g. “This is a plizzle. There are different kinds of plizzles. This one is a small plizzle. What do you think other plizzles look like?” Even in this open-ended context, 4-year-olds produced relevant alternatives about half the time. Our findings suggest that preschoolers truly are sensitive to the informativeness of speakers word choices, and can recognize opportunity for inference spontaneously. Even without forced-choice alternatives available, preschoolers show strong proficiency at inferring cues and implications from language use.

Methods

Participants. A new planned sample of 24 children (mean age 4;6) were recruited from the Childrens Discovery Museum of San Jose.

Materials. We used a similar task design as Experiments 1 and 2, but showed children only a single picture rather than a triad. We also modified the items to convey stronger alternatives to children because some of the original items had difficult contrasts to articulate (e.g. “broken” to “unbroken” or “fixed”). 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 ran a posttest to ensure that children were familiar with all of the properties used in the task. Children successfully identified the contrasts 96% of the time.

Procedure. The experimenter read the storybook with children individually in a quiet room at the CDM. As before, children were introduced to a pictured character: Allen the Alien who was visiting planet Earth. Children participated in two training trials with familiar items before proceeding to the test trials. The adjectives used in the training trials were different than those in experimental trials, and portrayed properties very familiar to children. Unlike Experiments 1 and 2, children saw only a single image per trial rather than an exemplar and two pictured alternatives. For example, 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 cookies usually look like? What do most 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, and, if they still did not respond, were asked what shape most cookies are. If children provided an answer other 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 plizzles usually look like? What do most 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

Childrens free response for each trial was coded along a 3-point scale: a score of 2 indicated that they produced an exact opposite (e.g. hearing "small" and saying "big"); a score of 1corresponded to a production that was related to the referential property, but not a precise contrast (e.g. hearing "small" and saying a size property other than "big"); a score of 0 was given when the child produced an unrelated description to the referential property (e.g. hearing "small" and producing a response unrelated to size).

Despite the open-ended nature of the task, children demonstrated sensitivity to the speakers word choices as indicators of cultural knowledge. Children were under no constraints in their productions, and could presumably invent any description that appealed to them or repeat the information stated by the experimenter. We found, however, that 4-year-olds spontaneously produced relevant contrasts (a score of 1 or 2) more than half the time for both feature and size modifiers (54% and 60% respectively). More than a third of their 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. In all, preschoolers produced remarkably on-task descriptions based on the perceived informativeness of a single word choice: sensitivity to the adjective used.

Although we provided some subtle cues about variability in this task (e.g. stating that "There are different kinds of plizzles"), the referential expression did not particularly convey a contrastive linguistic framing the way Experiment 1 did (stating that "This is a special kind of..."). Instead, the actual reference resembled the *adjective only* framing of Experiment 2. The finding that 4-year-olds were sensitive to the informativeness of adjective use to bring to mind relevant contrast information gives further support that preschoolers are learning to use the pragmatics of word choice to infer cultural information. Even without the confines of a forced-choice procedure, children still infer that adjective choice informs a speakers knowledge or perspective to produce the particular utterance.

General Discussion

Our three experiments give support that children are sensitive to not just *what* speakers say, but *why* they say it in a particular way. We use the case study to investigate whether children make inferences about a speaker's knowledge or perspective that would lead to a modified description. In Experiment 1, we showed that by age 3.5, preschoolers reliable made contrast inferences from size and feature adjectives with the supportive cue of a contrastive linguistic framing ("This is a special kind of tibu"). Unlike adults, 4-year-olds had difficulty making contrast inferences from the bare cue of adjective use alone, although their performance increased after exposure to adjective contrasts before undergoing the experimental task. These findings suggest that preschoolers can infer relevant alternatives implied by word choice information, but they may need additional support from linguistic framing for adjective contrasts before spontaneously inferring contrast between competitors in our forced-choice task (where one item represented the *same* property as the one stated, and one represented the *contrast*. In Experiment 3, we used a free response

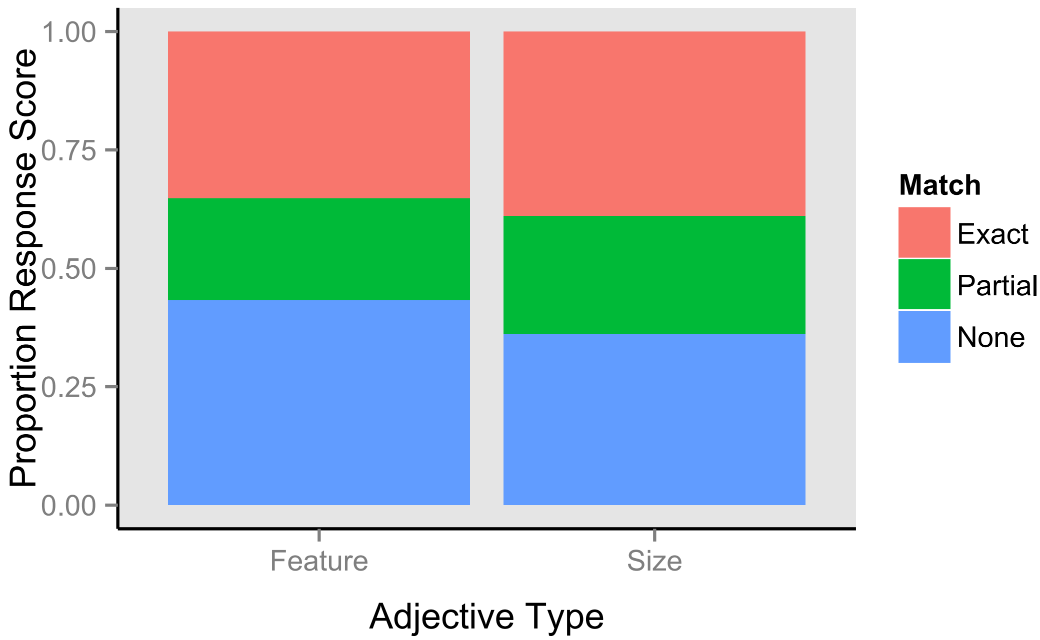

Figure 5. 4-year-olds free responses about properties of other category members upon hearing a feature adjective (left) or size adjective (right). Productions were coded as *exact matches* if they were opposite to the referenced description, *partial matches* if they were related to the referenced property but not a direct contrast, or *no match* if they were unrelated to the referenced modifier.

paradigm to examine childrens inferences from adjective use without the constraints of the forced-choice task. We found that, although 4-year-olds could produce any response of their liking, they spontaneously stated relevant contrast information more than half of the time. Altogether, our results suggest preschoolers appreciate adjective choice as conveying contrast information about other category members.

Although simple in design, our task was required a fairly counterintuitive response: selection of a property contrast rather than a property match. 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. Our findings that preschoolers nonetheless made these contrast inferences and robustness increased age suggests that they are reasoning backwards to infer contrast from speakers' word choices. Their sensitivity to the pragmatics of adjective use gives evidence that they are indeed processing information not just about what is stated explicitly, but also about what is culturally relevant to lead to such a description.

Extending their understanding of rational agents and sampling (e.g. Xu & Garcia, 2008; Xu & Denison, 2009; Gweon et al., 2010), preschoolers may use the informativeness of word choices to consider the implied state of the world that might lead to those choices. Their sensitivity to statistical regularities in the world may help them expect that modified referents contrast with other category members along the stated property dimension (i.e. that "big dog" implies that dogs can vary by size). Just as they can reason about likely sample to population of pingpong balls, language cues such as adjective use may implicitly convey cultural knowledge about the world. In this way, word choices refer not just to references in the present (e.g. the big dog currently in sight), but also to contrasts with other possible contexts and category members (that this dog happens to be big, and others are smaller).

If children assume that speakers are being both rational and communicative, then they can take advantage of opportunities for inference wherever they recognize a speaker's choice to produce an utterance in one form over another. Building on the hypothesis that children's abilities to compute implicatures relates to their ability to recognize linguistic alternatives (Barner et al., 2011), children may need to accumulate experience with how cultural information about the world is conveyed before recognizing relevant opportunities for pragmatic inference. Our case study of adjectives provides evidence that children are learning to process both explicit and implicitly conveyed information. Their increasing familiarity with opposite pairs leads them to robustly infer contrast with age given a supportive linguistic framing in Experiment 1 and with recency of exposure to contrasts in Experiment 2. Without any constraints, they generated relevant alternatives on their own more than half of the time. The ability to infer culturally-implied knowledge may require both the recognition of production choices that imply alternatives and also what those particular alternatives are.

Our work connects mechanisms of pragmatic inference, which have been well-studied in language development, with processes of cultural learning and generalization. We show that children can figure out that naming a “tall glorp” implies that most glorps are shorter. The same mechanism might also allow them to infer (usefully) that the term ”cognitive scientist” implies the existence of other types of scientists, or to infer (negatively) that “female scientist” implies that most scientists—at least in the mind of the speaker—are not female. Thus, pragmatic inferences of the type we study here may constitute an important mechanism of transmission for cultural knowledge.

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.

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 naıve 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.* 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., & 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 12months. *Cognition*, *123*(1), 50–60.

Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29 th 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., He, Z. . ·., & Ran, Y. . ·. (1986). Relevance: Communication and cognition.

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.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.