tapraid5/z2p-devpsy/z2p-devpsy/z2p00512/z2p3035d12z | xppws | S=1 | 7/5/12 | 14:39 | Art: 2010-0203

Developmental Psychology

0012-1649/12/$12.00 DOI: 10.1037/a0029477

The Use of Lexical and Referential Cues in Children's Online Interpretation of Adjectives

Yi Ting Huang
University of Maryland College Park

Jesse Snedeker
Harvard University

Recent research on moment-to-moment language comprehension has revealed striking differences between adults and preschool children. Adults rapidly use the *referential principle* to resolve syntactic ambiguity, assuming that modification is more likely when there are 2 possible referents for a definite noun phrase. Young children do not. We examine the scope of this phenomenon by exploring whether children use the referential principle to resolve another form of ambiguity. Scalar adjectives (*big*, *small*) are typically used to refer to an object when contrasting members of the same category are present in the scene (big and small coins). In the present experiment, 5-year-olds and adults heard instructions like "Point to the big (small) coin" while their eye-movements were measured to displays containing 1 or 2 coins. Both groups rapidly recruited the meaning of the adjective to distinguish between referents of different sizes. Critically, like adults, children were quicker to look to the correct item in trials containing 2 possible referents compared with 1. Nevertheless, children's sensitivity to the referential principle was substantially delayed compared to adults', suggesting possible differences in the recruitment of this top-down cue. The implications of current and previous findings are discussed with respect to the development of the architecture of language comprehension.

AQ: 1 *Keywords:* language processing, referential contrast, comprehension, scalar adjectives

Our moment-to-moment interpretation of language depends not only by the words that we hear but also on the situation in which they occur. Take for example the fragment in (1).

(1) I'll eat the donut with . . .

If you heard this snippet while waiting in line at a Dunkin' Donuts, you would probably expect the speaker to complete the sentence with a modifier like *the sprinkles* or *chocolate icing*. With so many flavors around, a more specific description is called for. In contrast, if the same comment was made by a friend who had just been served at a diner, you might expect it to end with an instrument like *a fork* or *my hands*. With only one donut in sight, a modifier would be redundant. Empirical work demonstrates that adults rapidly use the number of referents in a situation to guide their interpretation of ambiguous prepositional phrases (Altmann & Steedman, 1988; Crain & Steedman, 1985). This rapid sensitivity to the *referential principle*, and the ability to combine referential and lexical information, has been central in supporting theories that characterize the mature language comprehension system as interactive and opportunistic (MacDonald, Pearlmutter, & Seidenberg, 1994; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell & Tanenhaus, 1994).

In contrast, the evidence to date suggests that children may have substantial difficulties coordinating linguistic and contextual cues during real-time comprehension. When confronted with ambiguities like (1), 5-year-olds draw on lexical information such as the meaning of verbs to guide their interpretations, but they fail to use contextual information such as the number of referents in the scene (Choi & Trueswell, 2010; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999; Weighall, 2008). Nevertheless, while this disparity between adults and children is striking, the scope of the phenomenon is unclear. Are children's failures to use referential information limited to interpretations of syntactic structure? Or are they unable to use context to make any predictions about the use of descriptions and modifiers during real-time comprehension?

In the remainder of the introduction, we briefly review recent studies on children's use of referential context. Next we turn to another area of comprehension—the interpretation of scalar adjectives like *big* and *tall*—and examine reasons why developmental patterns in the processing of these terms may help us understand the acquisition of the referential principle. Finally, we describe how the present studies isolate children's use of lexical and referential cues during real-time comprehension.

Children's Failure To Use Referential Context During Comprehension

Children's failure to use referential information was highlighted in a study by Trueswell et al. (1999). Following Tanenhaus et al. (1995), the authors presented 5-year-olds with spoken instructions like (2) while their eye movements were recorded to a display featuring a napkin, a box, a frog sitting on a napkin, and a horse sitting on a table.

(2) Put the frog on the napkin in the box.

Yi Ting Huang, University of Maryland College Park; Jesse Snedeker, Harvard University.
AQ: 13
Correspondence concerning this article should be addressed to Yi Ting Huang, ADDRESS. E-mail: ythuang1@umd.edu
AQ: 14

These instructions contained a temporary ambiguity, in which the syntactic properties of the verb (*put*) strongly support an initial interpretation of "on the napkin" as a destination. In fact, in a context with just one frog, adults initially looked over to the napkin, suggesting that they had misanalyzed the phrase as a destination. However, after hearing the second prepositional phrase ("in the box"), adults correctly revised their interpretation of "on the napkin" to be a modifier of the noun. Thus, on the vast majority of the trials, they picked up the frog that was on the napkin and put it into the box. Critically, when the horse in this scene was replaced with a second frog (a frog sitting on the table), adults were able to immediately use this contrast to infer that "on the napkin" must be a modifier that distinguishes the two referents and no longer made spurious looks to the napkin.

Children, in contrast, were unaffected by the referential context. In the presence of either one or two frogs, they frequently looked at the napkin, suggesting that they incorrectly interpreted the modifier as a destination. Moreover, on over half of the trials, children performed an action that involved this incorrect destination (taking a frog and putting it on the napkin before putting it in the box), suggesting that they never revised their misanalysis. This failure to use information from the scene to facilitate interpretation of syntactically ambiguous sentences has been replicated under a variety of conditions (Choi & Trueswell, 2010; Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Weighall, 2008). It is also typically most robust in 5-year-olds and less pronounced in 8- and 11-year-olds (Trueswell et al., 1999; Weighall, 2008).

Subsequent studies have ruled out several potential explanations for this pattern. First, children are clearly capable of interpreting ambiguous phrases as postnominal modifiers (Snedeker & Trueswell, 2004). Four- and 5-year-olds prefer modifier interpretations when the properties of the verb support this interpretation: When asked to "Choose the frog with the feather," they select a frog in the display that's holding the feather. Second, children are able to use information such as the prosody of an utterance to access appropriate interpretations, suggesting that their failure to recruit referential context is not related to a global inability to use nonlexical cues (Snedeker & Yuan, 2008). Finally, while children are sensitive to the need for modification to distinguish between referents during *production*, they often fail to exploit this knowledge to inform interpretations of ambiguous phrases during *comprehension* (Trueswell & Gleitman, 2004). Five-year-olds reliably produce situationally appropriate modifiers: When asked to identify the appropriate character in a story about two frogs in the scene ("Which frog visited Mrs. Squid's house?"), they describe the intended referent with informative statements like "The frog on the napkin" (Hurewitz et al., 2000). Yet just moments later, these same children, when presented with two frogs in the eye-tracking task, continue to incorrectly interpret "on the napkin" as a destination.

How can we account for this failure? Here current theories in adult psycholinguistics may offer potential insights into children's behavior (MacDonald et al., 1994; Trueswell & Gleitman, 2004; Trueswell & Tanenhaus, 1994). These accounts suggest that the architecture of the language comprehension system has two critical characteristics: (a) it maps input onto multiple representations that are situated across distinct levels (phonological, prosodic, syntactic, semantic, discourse), and (b) these representations are systematically linked such that processing at one level constrains processing at other levels. Since the connections between levels of representation are primarily acquired through experience, children continue to learn about how each level constrains the others even after the representational systems themselves are place. The rapidity with which particular constraints develop will depend, in whole or in part, upon the strength of the correlation between two phenomena. This predicts that children's sensitivity to the referential principle in a given situation will depend on the likelihood that they have heard multiple referents in a scene appropriately disambiguated with modifiers.

Critically, evidence from adult production suggests the correlation between the number of referents and postnominal modification may be fairly weak. In an unscripted conversation task, Brown-Schmidt and Tanenhaus (2008) asked speakers to instruct listeners to move a target to different locations on a board (e.g., a square among an array of other squares). To do so, speakers produced utterances that disambiguated the target with a modifier 53% of the time (e.g., "Pick up the square on the left"). Yet nearly just as often, they used ambiguous bare definite noun phrases to describe the intended referent (e.g., "Pick up the square"). Adult listeners had little difficulty understanding these ambiguous phrases since referents could often be inferred based on the task goals or prior utterances. Nevertheless, this variability suggests a reason why children may fail to use scene information to predict postnominal modification. If their input does not support a direct mapping between the number of referents in a scene and the production of modifiers, then children are confronted with a much trickier problem of tracking the relative salience of possible referents in the discourse. However, these findings also predict that in situations where scene information might be a more reliable predictor of language use, children's sensitivity to the referential principle may emerge early in development.

Test Case: Interpretation of Scalar Adjectives

In our current study, we turn to scalar adjectives like *big* and *tall* as a critical case where language use is systematically linked to the number of referents in the scene. These adjectives, like postnominal modifiers, are typically used in situations where an unadorned noun would be insufficient to distinguish the referents (Brown-Schmidt & Tanenhaus, 2006; Ferreira, Slevc, & Rogers, 2005; Gregory, Joshi, Grodner, & Sedivy, 2003; Nadig & Sedivy, 2002; Sedivy, 2003). Thus, they are felicitous when there are at least two objects of the same kind in the context (two glasses, one tall and one short), and they are infelicitous when there is only one object that can be referred to by the noun alone (one glass). Critically, unlike postnominal modifiers, prior work has highlighted the tight correlation between referent number and production of scalar adjectives (Ferreira et al., 2005; Gregory et al., 2003; Nadig & Sedivy, 2002; Sedivy, 2003). Brown-Schmidt and Tanenhaus (2006) found that speakers described a target (e.g., large triangle) with a scalar adjective 25% of the time when it was the only one of its kind in the scene. However, they did so 98% of the time when the scene included a contrast of the same kind (e.g., small triangle).

This tendency to modify may reflect linguistic properties associated with scalar adjectives (Kennedy, 1999). At the level of lexical meaning, these terms specify (a) a scale along which entities are ordered and (b) a pole along that scale that generates

tapraid5/z2p-devpsy/z2p-devpsy/z2p00512/z2p3035d12z | xppws | S=1 | 7/5/12 | 14:39 | Art: 2010-0203

the comparisons. For example, an adjective like *big* compares objects along the positive polarity of the size scale, while a term like *short* compares objects along the negative polarity of the height scale. Nevertheless to interpret these terms, listeners must also set a standard of comparison to determine what values along that scale are sufficient to count as examples of the adjective (Bierwisch, 1987; Kamp & Partee, 1995). For example, to evaluate a statement like "Hummers are big cars," one must know information not only about the size of Hummers but also the size of cars in general (Hummers are *big* relative to Civics and Escorts). The appropriate scale can shift dramatically when the same referent is evaluated relative to a different category (Hummers are not *big vehicles* relative to trains, cruise ships, Boeing 747s). Even within a single modified category, the standard of comparison can change with respect to a contextually defined comparison class (*big cars* in Europe versus *big cars* in Texas).

Prior work examining off-line interpretations has found that scalar adjectives emerge early in development, with most children producing terms like *big*, *nice*, *pretty*, and *heavy* by their second birthday (Dale & Fenson, 1996). Furthermore, at the earliest stages of acquisition, these words are used to refer to known (Ebeling & Gelman, 1988; Sera, Troyer, & Smith, 1988; L. B. Smith, Cooney, & McCord, 1986; Syrett, Kennedy, & Lidz, 2009) and novel entities (Barner & Snedeker, 2008) in a contextually sensitive manner. However, far less is known about how children interpret scalar adjectives during online comprehension.[1] Fn1

What little work there is has focused on the interpretation of nonscalar adjectives like color words (Fernald, Thorpe, Marchman, 2010; Sedivy, Demuth, Chunyo, & Freeman, 2000). Fernald et al. (2010) found that when asked to find the *blue car*, 3-year-olds abandoned their looks to different colored competitors (red car or red house) shortly after adjective onset. While these findings suggest that young children can rapidly recruit the meanings of adjectives to restrict reference (distinguishing *blue* from *red*), they do not directly address whether they are also sensitive to the referential principle. Prior adult work has shown that unlike scalar adjectives, speakers often produce color words, even in situations where modifiers are unnecessary to pick out referents (e.g., when there is only one blue car in the scene; Sedivy, 2003). Consequently, even in the presence of two referents from the same category, listeners typically fail to interpret these terms contrastively. Furthermore, children's sensitivity to referential contrast may be difficult to isolate using a two-alterative force-choice task like the preferential-looking paradigm (Fernald et al., 2010). Since the presence of contrast also directly increases the number of potential referents for the noun, children may experience interference with reference resolution, particularly after the onset of the noun. In fact, Fernald et al. (2010) found that 2- and 3-year-olds were actually faster to look at the correct target when it was paired with a between-category item (red house) than a within-category one (red car).

The Present Study

Thus, in order to examine the role of referential contrast in children's comprehension, we borrowed a paradigm developed in adult psycholinguistics (Sedivy et al., 1999) and adapted it for use in children. Five-year-olds were given instructions like "Point to the big coin," and their eye-movements were measured to visual displays containing four items that varied in size and category membership (see Figure 1). These displays always featured a F1 *Target* object that matched the adjective/noun combination (big coin) and a *Contrast* object that differed in size. In the two-referent trials, this item came from the same category as the Target (small coin), while in the one-referent trials, it came from a different category (small button). The displays also featured a *Competitor* that matched the Target in size but not by category (big stamp), paired with an unrelated object that matched the Target in neither size nor category (small marshmallow).

The presence of the Competitor was critical since it allowed us to gauge the relative advantage of the Target when it was paired with a between-category item (when both the Target and Competitor can be distinguished without a modifier) versus a within-category item (when the Target but not the Competitor required a modifier). Unlike preferential-looking, these displays allowed us to examine children's sensitivity to the referential principle independently of their looks to the contrast items. We predicted that if prior failures to recruit referential context reflected the weak correlation between referent number and postnominal modification, then children who hear an ambiguous adjective ("Point to the big . . .") should now restrict reference to a member of the contrast set (big coin, not big stamp). This would lead to more looks to the Target and fewer looks to the Competitor in the two-referent trials compared with the one-referent trials. If, however, children are broadly insensitive to the referential principle, then we would expect the number of referents to have no effect on the interpretation of scalar adjectives as well. This would lead to no differences in Target and Competitor looks across the one- and two-referent trials.

We also compared children's use of referential contrast with their use of the lexical meaning of scalar adjectives. This was done by measuring when children rule out referents that were incompatible with the specified pole: Upon hearing *big*, we would expect looks to the big objects to increase (Target and Competitor) and looks to the small objects to decrease (Contrast and Unrelated item), while upon hearing *small*, we would expect looks to the small objects to increase and looks to the big objects to decrease. Prior work has highlighted the rapid use of lexical meaning to restrict reference for nonscalar adjectives (Fernald et al., 2010; Sedivy et al., 2000) and resolve ambiguity for syntactic constructions (Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008). This suggests that children would draw on the polarity of scalar adjectives to guide real-time interpretations. In fact, given the complexity of these terms and previous failures to find sensitivity to the referential principle, it is possible that children may simply rely on the polarity of the adjective plus the subsequent noun information to distinguish referents in our task.

[1] We know of just one study which has specifically examined children's online comprehension of scalar adjectives. Nadig et al. (2003) found that while adults were faster to interpret *big car* in the presence of a same category contrast (small car) compared with an unrelated object (baseball), 5-year-olds' preference for the target were not directly affected by the number of referents in the scene. This insensitivity to the referential principle led the authors to suggest that "children may not yet have the processing capacity to successfully incorporate referential context" (Nadig et al., 2003, p. 577).

tapraid5/z2p-devpsy/z2p-devpsy/z2p00512/z2p3035d12z | xppws | S=1 | 7/5/12 | 14:39 | Art: 2010-0203

Figure 1. Visual display for the “big coin” trial.

In Experiment 1, we first examine the use of lexical meaning and referential information in adult interpretation of scalar adjectives. Since our experimental design provides tight controls for several features of the display that are somewhat different from those previously used (Nadig et al., 2003; Sedivy et al., 1999), it was necessary to collect additional data from adult participants to establish the expected pattern of performance in this task. The goals of this experiment were twofold. First, we wanted to replicate the contrast effects seen in the previous adult studies by Sedivy et al. (1999; Sedivy, 2003). Second, we wanted to establish the time-course of these contrast effects with respect to the use of polarity information.

Experiment 1

Method

Participants. Thirty-two undergraduates at Harvard University participated in this study and received either course credit or $5 for their participation. All participants were native English speakers.

Procedure. Participants sat in front of an inclined podium divided into four quadrants (upper left, upper right, lower left, and lower right), each containing a shelf where objects could be placed. A camera at the center of the display focused on the participant’s face and recorded the direction of their gaze while they were performing the task. A second camera, located behind the participant, recorded both their actions and the location of the items in the display. For every trial, the experimenter took out four objects from a bag and placed them each on a shelf in a prespecified order. This was followed by a prerecorded utterance that instructed participants to point to one of the objects. Once the participant pointed to an object, the trial ended, the objects were removed from the display, and the next trial began.

Materials. Scalar adjectives were selected from the size (*big, small*) and height scale (*tall*, *short*). Each item was rotated through the four conditions of a 2 × 2 design. The first factor, polarity, indicated whether the Target item was from the negative pole (*small, short*) versus the positive pole (*big*, *tall*) of the scale. These terms were embedded in commands like (3).

(3) Point to the *big coin*.

These sentences were prerecorded by a female actor. The second factor, contrast, indicated whether the Contrast item belonged to the same basic-level category (two-referent trials) versus a different one (one-referent trials). Unlike prior studies (Sedivy et al., 1999; Nadig et al., 2003), we ensured that features of these items were tightly controlled in three ways. First, Contrast objects always differed in size from the Target in both one-referent and two-referent trials (small button and small coin). This made certain that any difference that emerged between these conditions can be specifically attributed to the category membership of the Contrast and not from other extraneous features. Second, we arranged the objects in a way that increased the likelihood of encoding the size difference between the Target and Contrast. Thus, the Contrast was always placed to the left/right of the Target, while the Competitor was always placed above/below it. Finally, we counterbalanced the Target and Competitor items so that all Targets were used as Competitors and vice versa. This was achieved by creating eight versions of each base item that were used to create eight presentation lists such that each list contained two items in each of the eight cells (four items in each of the critical condition), and each base item appeared just once in every list. Table A1 provides a full list of the materials.

The 16 critical trials were randomized and interspersed with eight filler trials featuring displays similar to the two-referent trials (big ball vs. small ball) but asking for a non-Contrast item (“Pick up the big tomato”). This was critical since the effect of contrast sets on interpretation is assessed by comparing trials in which the Target appears with a Contrast item from the same category with ones in which it does not. Thus, it was possible that if the remaining two distractor items were never members of a contrast set, then participants could learn that whenever they see two items of the same kind, one of those two items will always be the Target. This could facilitate Target identification on two-referent trials relative to one-referent trials, but it would not reveal whether participants were sensitive to the informational implications of modification or whether they were simply sensitive to a specific contingency in our stimuli. The presence of these filler trials ensured that eye-movements to the Target did not reflect this type of strategy.

Finally, the objects in the displays came from 24 sets of real household items, matched to the relative scale of other members of the set. These items were always consistent with their real-world size/height and were never miniaturized versions of otherwise large objects (e.g., a toy vehicle as a representation of a car). This was critical to ensure that the interpretations of scalar adjectives were not complicated by questions of the appropriate scale to use (i.e., the real-world scale or one that was specific to the display). To ensure that our items were good exemplars of the adjective/noun combination, we conducted a separate rating task. An additional group of 36 adults were asked to rate how an object (big coin) compared with typical members of its category (coins) along a particular dimension (size). Participants were asked to make their judgments on a 1 to 7 scale, where 1 = *much smaller/shorter than usual* and 7 = *much bigger/taller than usual*. In order to avoid any direct comparisons across objects from the same category, participants only saw one member of each kind (either a big coin or small coin). We found that Targets for the positive polarity trials were rated significantly higher ($M = 5.3$, $SD = 1.3$) than those for the negative polarity trials ($M = 1.6$, $SD = 0.7$), $F(1, 126) = 220.30$, $p < .001$, $\eta^2 = .64$.

Coding. Trained research assistants watched videotapes of the participants' actions and coded the object that was selected on each trial. Across both experiments, analyses of eye-movements only included trials where participants correctly selected the Target. However, in Experiment 1, no trials were excluded on this basis. Approximately 0.5% of test trials were excluded from further analysis because of experimenter error.

Eye-movements were coded by a research assistant, who was blind to the location of each object, using frame-by-frame viewing of the participant's face on a Sony digital VCR (Snedeker & Trueswell, 2004). Each recorded trial began from the onset of the instruction and ended with completion of the corresponding action. Each change in direction of gaze was coded as toward one of the quadrants, at the center, or missing due to looks away from the display or blinking. These missing frames accounted for approximately 2% of all coded frames and were excluded from analysis. Twenty-five percent of the trials were checked by second coder, who confirmed the direction of fixation for 94.6% of the coded frames. Any disagreements between the two coders were resolved by a third coder.

Results

For each analysis, we first identified differences across conditions by conducting analyses of variance (ANOVAs) over four broad time windows: (a) *Baseline region*: (a 667-ms period from the onset of the instruction to the onset of the adjective; "*Point to the*"), (b) *Adjective region* (a 433-ms period from the onset of the adjective to the onset of the noun; "*big*"), (c) *Noun region* (a 667-ms period from the onset of the final noun to the offset of the command; "*coin*"), (d) *End region* (a 700-ms period following the offset of the command). Each period was shifted 200 ms after the relevant marker in the speech stream to account for the time it takes to program saccadic eye-movements (Matin, Shao, & Boff, 1993). Critical differences in fixations during these large time windows were followed up by fine-grained analyses using 100-ms intervals, starting from the onset of the adjective until 1,900 ms later. The onsets for these analyses were not offset in time and matched their real-time occurrence in the speech stream.

Use of polarity information. Our first analysis examined when adults first used the polarity of the scalar adjective to rule out objects from the opposite end of the height or size scale. Our dependent measure was the sum of looking time to the Target and Competitor items divided by the sum of looking time to all four objects. These scores ranged from zero (exclusive looks to objects that did not match the specified pole) to one (exclusive looks to objects that matched the specified pole). For each polarity, we determined when scores exceeded what would be predicted by chance ($M = 50\%$).

F2 Figure 2 illustrates that during the Baseline region, looks to the matching referents in the positive polarity trials were no different from chance ($M = 52\%$, $SE = 2\%$; ts < 1.00, ps $> .40$) and in the negative polarity trials were marginally below chance ($M = 45\%$, $SE = 3\%$), $t(31) = 1.75$, $p < .10$, $d = 0.40$. We believe that this latter effect reflects an early perceptual bias against looking at less salient smaller/shorter items. Critically, during the Adjective region, looks to the Target and Competitor increased across both polarities, leading to significantly higher looks to the matching referents in both the positive ($M = 68\%$, $SE = 3\%$), $t(31) = 6.89$, $p < .001$, $d = 0.99$, and negative trials ($M = 56\%$, $SE = 3\%$), $t(31) = 2.25$, $p < .05$, $d = 0.65$. These looks continued to increase during the Noun region across positive ($M = 84\%$, $SE = 2\%$), $t(31) = 18.86$, $p < .001$, $d = 0.99$, and negative trials ($M = 81\%$, $SE = 2\%$), $t(31) = 14.55$, $p < .001$, $d = 0.99$. This pattern was sustained in the End region where looks to matching referents dominated both the positive ($M = 93\%$, $SE = 2\%$), $t(31) = 26.10$, $p < .001$, $d = 0.99$, and negative trials ($M = 95\%$, $SE = 2\%$), $t(31) = 29.39$, $p < .001$, $d = 0.99$.

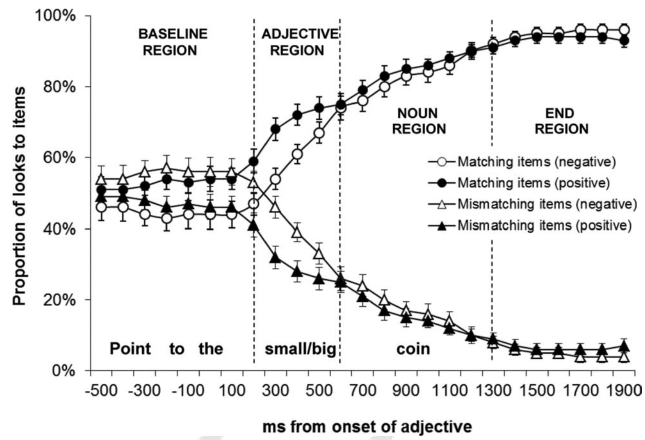

Figure 2. In Experiment 1, adult looks to matching referents (Target and Competitor items) and mismatching referents (Contrast and Unrelated items) in negative and positive polarity trials. AQ: 15

To explore the timing of these polarity effects in greater detail, we examined Target and Competitor looks during 100-ms intervals after the onset of the adjective. For the positive polarity trials, we found that looks to the matching referents exceeded chance approximately 200 ms after adjective onset ($M = 59\%$, $SE = 3\%$), $t(31) = 2.75$, $p < .01$, $d = 0.74$. For the negative polarity trials, this moment occurred shortly after, in the 400-ms time window ($M = 61\%$, $SE = 3\%$), $t(31) = 4.23$, $p < .001$, $d = 0.98$. Comparisons across the two polarities revealed that looks to the matching referents in the positive trials exceeded those in the negative trials from the 200-ms window, $t(31) = 2.49$, $p < .05$, $d = 0.93$, through the 400-ms window, $t(31) = 3.43$, $p < .01$, $d = 0.94$. Altogether this suggests that adult processing of adjective polarity occurred rapidly but was slightly delayed following a negative expression compared to a positive one.

Use of referential contrast. Our second set of analyses examined whether referential contrast facilitated adults' interpretation of scalar adjectives. Here our dependent measure was the proportion of looking time to the Target divided by the sum of looking time to the Target and the Competitor. These scores ranged from zero (exclusive looks to the Competitor) to one (exclusive looks to the Target) and allowed us to focus on the relative advantage of the Target when it was paired with a within- versus between-category contrast. These scores were analyzed in a 2×2 ANOVA with contrast (one- vs. two-referent) and polarity (negative vs. positive) as within-subject variables.

tapraid5/z2p-devpsy/z2p-devpsy/z2p00512/z2p3035d12z | xppws | S=1 | 7/5/12 | 14:39 | Art: 2010-0203

F3 Figure 3 illustrates that that during the Baseline region, the proportion of looks to the Target initially remained around chance across all trials, leading no significant effects of contrast and polarity and no interaction between the two (all *F*s $<$ 1.20, all *p*s $>$ .20). Similarly, following the onset of the adjective in the subsequent region, looks to the Target across conditions did not substantially diverge (all *F*s $<$ 1.20, all *p*s $>$ .20). During the Noun region, however, adults made more looks to the Target in the two-referent trials (M = 73%, SE = 2%) compared with the one-referent trials (M = 69%, SE = 2%), leading to a significant effect of contrast type, $F(1, 31) = 4.48$, $p < .05$, $\eta^2 = .13$. There was no additional effect of polarity or interaction between the two (both *F*s $<$ 2.50, both *p*s $>$ .10). Finally, during the End region, adults across all conditions closed in on the Target, leading to no significant effects of contrast and polarity and no interaction between the two (all *F*s $<$ 1.00, all *p*s $>$ .30).

We explored the timing of these differences in greater detail using 100-ms intervals after the onset of the adjective. Approximately 600 ms after adjective onset (or around at the onset of the noun), there emerged an advantage for Target looks in two-referent trials. During this window, looks in the two-referent trials exceeded those in the one-referent trials (M = 59%, SE = 2% vs. M = 53%, SE = 3%), $F(1, 31) = 3.96$, $p < .05$, $\eta^2 = .10$, and continued to do so through the 900-ms time window (M = 76%, SE = 3% vs. M = 69%, SE = 3%), $F(1, 31) = 4.82$, $p < .05$, $\eta^2 = .13$. This indicates that the presence of a within-category contrast facilitated real-time interpretation of scalar adjectives in adults. There was no additional effect of polarity or interaction between the two in any of the fine-grained time windows (all *F*s $<$ 2.00, all *p*s $>$ .15).

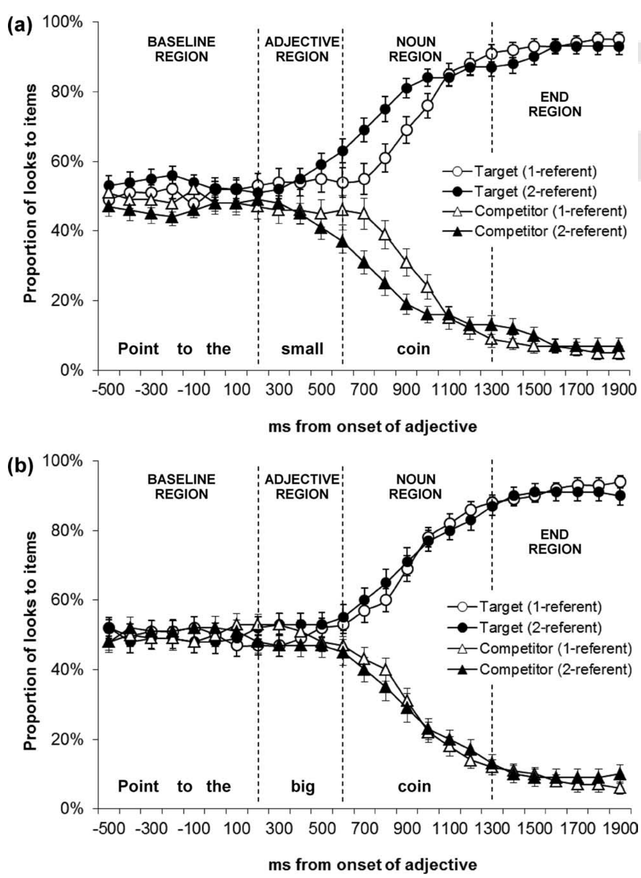

Figure 3. In Experiment 1, adult looks to the Target and the Competitor in (a) negative polarity trials and (b) positive polarity trials.

Discussion

In Experiment 1, we found that adults used both the meanings of scalar adjectives and their referential implications to constrain the referent of noun phrases. Approximately 300 ms after the onset of the adjective, adults retrieved polarity information and used it to rule out incompatible referents of the opposite size/height. Curiously, we found evidence that adult processing of the negative pole was slightly delayed relative to the positive pole. This weaker use of the negative pole is consistent with either an account where these terms are more semantically complex and marked (H. Clark, 1969) or simply less frequent than their positive counterparts (Carey & Considine, 1973; Wepman & Hess, 1969). We return to this point in the next section. Critically, approximately 600 ms after adjective onset, adults were faster to comprehend a modified noun in the presence of another contrasting member of the same category. This replicates prior findings (Grodner & Sedivy, 2011; Sedivy, 2003; Sedivy et al., 1999) and demonstrates that adult listeners prefer to interpret scalar adjectives as modifying nonunique referents in the context.

In Experiment 2, we tested 5-year-olds with the same task to examine whether their on-line interpretations of scalar adjectives are influenced by the same linguistic and contextual cues. If children use polarity information to restrict reference, then we would expect their looks to the Target and Competitor to increase after hearing adjectives. Critically, if children are also sensitive to referential contrast, we would expect their looks to the Target to increase more quickly in the two-referent trials compared with the one-referent trials.

Experiment 2

Method

Participants. Forty 5-year-olds (ranging from 4;6 to 5;8, mean age 5;0) were recruited from the Arlington Children's Center in Arlington, Massachusetts and the McGlynn Elementary School in Medford, Massachusetts. This age group was examined because prior studies have found robust failures to use the referential principle during this period of development (Choi & Trueswell, 2010; Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008). Information on participants' ethnicity, parental education, income, and occupation was not recorded, but the data from the 2000 Census for these communities suggest that participants predominantly came from middle socioeconomic homes and were primarily Caucasian. All children were native English speakers.

Procedure and materials. The procedure and materials were identical to those used in Experiment 1.

Coding. The data were coded in the manner described for Experiment 1. Approximately 2.2% of trials were excluded from further analysis due to experimenter error, while approximately 3.1% of trials were excluded because of a participant's incorrect action. Missing frames due to blinks or looks away accounted for 5.4% of all coded frames and were also excluded from analysis. First and second coding had 93.8% intercoder reliability.

Results

We examined children's use of lexical meaning and referential contrast using the same coarse- and fine-grained analyses employed in Experiment 1.

F4 **Use of polarity information.** Figure 4 illustrates that during the Baseline region, looks to the Target and Competitor in the positive (M = 47%, SE = 2%) and negative polarity trials (M = 51%, SE = 3%) were no different from chance (ts $<$ 1.00, ps $>$ .30). However during the Adjective region, children in the positive trials quickly shifted their fixations to the matching referents, exceeding what would be predicted by chance (M = 59%, SE = 2%), $t(39) = 3.61$, $p < .001$, $d = 0.96$. In contrast, looks in the negative trials were marginally below chance, demonstrating delayed evidence of the same perceptual bias seen in adults (M = 46%, SE = 2%), $t(39) = 1.84$, $p < .10$, $d = 0.38$. Critically during the Noun region, looks to the matching referents increased across both polarities and were greater than chance in the positive (M = 80%, SE = 2%), $t(39) = 16.70$, $p < .001$, $d = 0.99$, and negative trials (M = 65%, SE = 2%), $t(39) = 6.21$, $p < .001$, $d = 0.99$. This pattern continued into the End region, where looks to matching referents dominated in the positive (M = 88%, SE = 2%), $t(39) = 23.32$, $p < .001$, $d = 0.99$, and negative trials (M = 80%, SE = 2%), $t(39) = 13.51$, $p < .001$, $d = 0.99$.

Fine-grained analyses confirmed a temporal difference in children's use of lexical information across the two polarities. While looks to the matching referents exceeded chance approximately 300 ms after the onset of positive polarity adjectives (M = 56%, SE = 3%), $t(39) = 2.22$, $p < .05$, $d = 0.58$, they failed to do so until the 600-ms window for negative polarity terms (M = 58%, SE = 3%), $t(39) = 2.73$, $p < .01$, $d = 0.74$. Matching referent looks in the positive trials also exceeded those in the negative trials for an extended period from the 200-ms window, $t(39) = 2.91$, $p < .01$, $d = 0.99$, through the 1,600-ms window, $t(39) = 2.96$, $p < .01$, $d = 0.93$.

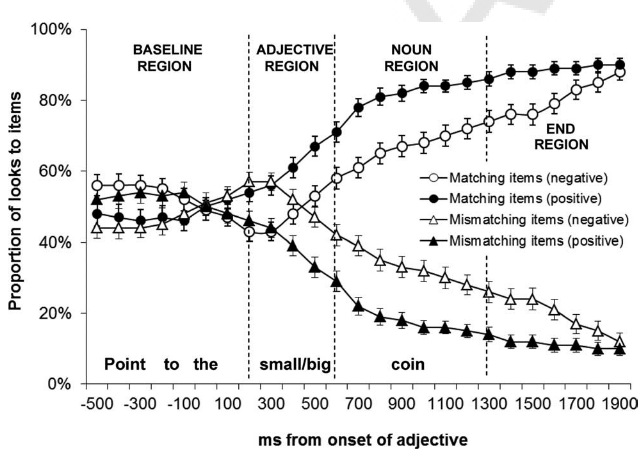

Figure 4. In Experiment 2, child looks to matching referents (Target and Competitor items) and mismatching referents (Contrast and Unrelated items) in negative and positive polarity trials.

We explored this delay in greater detail by examining whether the processing of polarity differed across the two age groups. We focused on an extended region when fixations in the negative trials lagged behind those in the positive trials for either adults or children. The relative difficulty of the negative trials was calculated as a difference score of matching referent looks in the negative trials minus those in the positive trials. Negative scores indicated fewer looks in the negative trials compared to the positive trials while positive scores indicated fewer looks in the positive trials compared to the negative trials. These scores were analyzed in a 2 × 2 ANOVA with age (adult vs. children) as a between-subjects variable and time window (100-ms intervals from 200ms to 1,600 ms after adjective onset) as a within-subject variable. This analysis revealed that children (M = –13%, SE = 2%) experienced greater difficulties with the negative pole compared with adults (M = –4%, SE = 2%), leading to a main effect of age, $F(1, 70) = 9.92$, $p < .01$, $\eta^2 = .12$. This disadvantage in the negative trials lasted longer in children than adults, leading to an interaction between age and time, $F(14, 980) = 1.87$, $p < .05$, $\eta^2 = .03$. Follow-up comparisons confirmed that children exhibited disproportionate difficulties with the negative pole from the 600-ms window, $F(1, 70) = 4.58$, $p < .05$, $\eta^2 = .06$, through the 1,600-ms window, $F(1, 70) = 7.12$, $p < .01$, $\eta^2 = .09$.

Use of referential contrast. Figure 5 illustrates that during the Baseline region, the proportion of looks to the Target remained around chance across all trial types, leading no significant effects of contrast and polarity and no interaction between the two (all Fs $<$ 1.00, all ps $>$ .40). In the Adjective region, there was a preference to look at the Target in the two-referent trials (M = 53%, SE = 1%) relative to the one-referent trials (M = 48%, SE = 2%). This led to a significant main effect of contrast, $F(1, 39) = 4.53$, $p < .05$, $\eta^2 = .10$, but no additional effect of polarity or interaction between the two (both Fs $<$ 1.00, both ps $>$ .70). Curiously, this preference disappeared following the onset of the noun (M = 62%, SE = 2% vs. M = 60%, SE = 2%), leading to no significant effects of contrast or polarity and no interaction between the two (all Fs $<$ 1.00, all ps $>$ .30). However, in the final region of analysis, Target looks in the two-referent trials (M = 82%, SE = 2%) were again higher than those in the one-referent trials (M = 76%, SE = 2%), leading to a main effect of contrast, $F(1, 39) = 8.09$, $p < .01$, $\eta^2 = .17$. During this region, Target looks were also higher in the positive trials (M = 83%, SE = 2%) than in the negative trials (M = 75%, SE = 2%), leading to an additional main effect of polarity, $F(1, 39) = 10.56$, $p < .01$, $\eta^2 = .21$, but no interaction between the two (Fs $<$ 1.00, ps $>$ .40). F5

Fine-grained analyses of the contrast effect confirmed that Target looks in the two-referent trials exceeded those in the one-referent trials during two distinct periods. The first coincided with the Adjective region, beginning approximately 300 ms after adjective onset (M = 53%, SE = 2% vs. M = 48%, SE = 2%), $F(1, 39) = 4.12$, $p < .05$, $\eta^2 = .10$, and continuing through the 400-ms window (M = 53%, SE = 2% vs. M = 47%, SE = 2%), $F(1, 39) = 5.59$, $p < .05$, $\eta^2 = .10$. The second occurred after the offset of the command, beginning approximately 1,300 ms after adjective onset (M = 75%, SE = 2% vs. M = 69%, SE = 2%), $F(1, 39) = 4.16$, $p < .05$, $\eta^2 = .09$, and continuing through the 1,700-ms window (M = 41%, SE = 3% vs. M = 34%, SE = 3%), $F(1, 39) = 4.09$, $p < .05$, $\eta^2 = .09$.

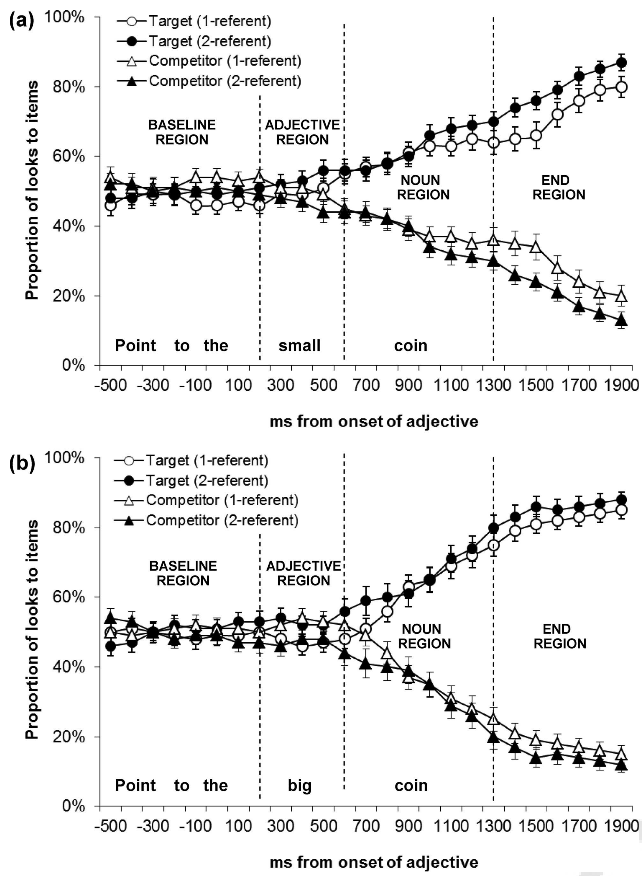

Figure 5. In Experiment 2, child looks to the Target and the Competitor in (a) negative polarity trials and (b) positive polarity trials.

Understanding the timing of contrast effects. While the results above suggest that within-category contrast items facilitated children's real-time interpretation of scalar adjectives, we were puzzled by why this pattern emerged during two distinct periods. In particular, the earliness of the first period (about 300 ms before contrast effects in adults) led us to consider whether children's fixations reflected perceptual biases to look at matching items in a contrast set (matching effect). In the current design, the presence of filler trials eliminated the correlation between the presence of the contrast set and the likely Target. Nevertheless, children may have developed an early preference to attend to matching sets that was independent of their language comprehension.

In order to distinguish between the causes of Target preferences in the two-referent trials, we conducted additional analyses that focused on correlations during two time windows: early (around adjective onset: –100 ms to 500 ms) and late (around sentence offset: 1,300 ms to 1,900 ms). First, we asked whether increases in Target looks associated with the presence of a within-category contrast item were correlated across these two time regions. We reasoned that if early and late preferences were driven by a common procedure, then we should expect the two to be correlated. If, however, increases in early Target looks reflected one procedure (e.g., a perceptual bias to look at matching items) and later Target looks reflected another (e.g., sensitivity to the referential principle), then we would not expect these effects to be correlated.

For each subject, we calculated the facilitation of the within-category contrast item on Target looks as a difference score of Target proportions (Target looks divided by the sum of Target and Competitor looks) in the two-referent trials minus the one-referent trials. Negative scores indicated fewer Target looks in the two-referent trials compared to the one-referent trials, while positive scores indicated greater looks in the two-referent trials compared to the one-referent trials. We found that scores in early region were not correlated with those in the late region, $r(40) = -0.06$, $p > .70$. This suggests that increases in Target looks during the two time regions were caused by distinct procedures.

Next we addressed the possibility that Target facilitation in the late region was caused by a different kind of matching effect. Recall that Fernald et al. (2010) found that after the onset of the noun, 2- and 3-year-olds generated both correct looks to the Target (*blue car*) and incorrect looks to a within-category item (*red car*). Similarly, it was possible that children who heard "big coin" increased their Target fixations based on their comprehension of the noun (*coin*) and not the adjective (*big*). This tendency would *not* explain why children generated fewer Target looks in the one-referent trials (where the Target was the only possible match for the noun) compared with the two-referent trials (where both the Target and Contrast matched the noun). Nevertheless, we reasoned that if the facilitation of Target looks reflected a matching bias triggered by the noun, then we might expect this to affect Contrast looks in the same way as well.

To examine this possibility, we correlated the facilitation of Target looks with the facilitation of Contrast looks during the late region. Target scores were calculated in the same way described above. For each subject, Contrast scores were calculated as a difference of Contrast proportions (Contrast looks divided by the sum of all looks) in the two-referent trials minus the one-referent trials. Negative scores indicated fewer Contrast looks in the two-referent trials compared with the one-referent trials, while positive scores indicated greater looks in the two-referent trials compared with the one-referent trials. We found that Target facilitation was not significantly correlated with Contrast facilitation, $r(40) = -.21$, $p > .15$. In fact, the negative relationship between the two was opposite from what would be predicted by a matching bias. This suggests that Target preferences during the final region were not triggered by the noun alone.

Finally, we examined whether Target looks during the late region reflected a linguistically driven contrast effect. We did so by focusing on children's fixations during the early region, specifically looks to items that were semantically consistent with the adjective polarity. We reasoned that if later facilitation of Target looks reflected sensitivity to the referential principle, then it should be correlated with prior processing of the adjective meaning. If, however, these effects reflected perceptual biases, then we would not expect the two to be correlated.

For each subject, we calculated the preference for semantically consistent items in the early window as the sum of Target and Competitor looks divided by the sum of all looks. For the late window, we used the same Target facilitation score described above. Unlike the prior analysis, we now found that preferences in the early region were significantly correlated with those in the late region, $r(40) = 0.34$, $p < .05$. Greater looks to the semantically consistent items predicted greater Target looks in the presence of within-category contrast items. This suggests that contrast effects during the final region of the two-referent trials reflected a late-emerging sensitivity to the referential principle during children's comprehension of adjectives.

Comparison between adults and children. We also explored whether the processing of referential context differed across the two age groups. We focused on an extended region when either adults or children exhibited significant facilitation in the two-referent trials and calculated the same Target facilitation score described above. These scores were analyzed in a 2 × 2 ANOVA with age (adult vs. children) as a between-subjects variable and time window (100-ms intervals from 600 ms to 1,700 ms after adjective onset) as a within-subject variable.

This analysis revealed that the timing of contrast effects varied across adults and children, leading to a significant interaction between time and age, $F(11, 70) = 4.23$, $p < .001$, $\eta^2 = .05$. Follow-up comparisons confirmed that contrast effects were initially greater in adults compared with children during the 800-ms, $F(1, 70) = 3.94$, $p < .05$, $\eta^2 = .05$, and 900-ms, $F(1, 70) = 3.98$, $p < .05$, $\eta^2 = .05$, time windows. However, this pattern reversed during a later period, from the 1,300-ms window, $F(1, 70) = 4.36$, $p < .05$, $\eta^2 = .06$, through the 1,700-ms window, $F(1, 70) = 5.93$, $p < .05$, $\eta^2 = .05$. This suggests that children were sensitive to referential contrast but that their ability to recruit this information was substantially delayed. Finally, we found no additional effect of age ($F < 1.00$, $p > .40$), suggesting no developmental differences in the overall strength of the contrast effects.

Discussion

In Experiment 2, we found that 5-year-olds interpreted scalar adjectives in ways that were similar to adults. Like adults, children exploited the polarity of these terms to distinguish between referents of different sizes about 400 ms after adjective onset. Similarly, like adults, children were better at interpreting positive polarity adjectives compared to negative ones. Critically, we found that about 1,300 ms after adjective onset, children became sensitive to the presence of multiple members of the same category in the visual scene and used this information to facilitate resolution of the correct referent. Like adults, they were quicker to restrict reference in the two-referent context compared to the one-referent context. These results differ from prior findings in the syntactic ambiguity literature (Choi & Trueswell, 2010; Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008) and clearly demonstrate that 5-year-olds can make use of the referential principle during real-time comprehension.

However, children's patterns of comprehension also exhibited notable differences from adults. First, while adults were able to rapidly overcome their difficulties with processing the negative polarity, children demonstrated prolonged delays with these terms. This pattern is consistent with prior off-line studies showing that 3- and 4-year-olds are more accurate at understanding positive adjectives relative to their negative counterparts (Donaldson & Wales, 1970; Farnham-Diggory & Bermon, 1968; Klatzky, Clark, & Macken, 1973; Wales & Campbell, 1970). This asymmetry also suggests that despite acquiring early knowledge of contrasting poles along a dimension (E. V. Clark, 1972), the interpretation of negative terms continues to present challenges to children, perhaps due to their semantic complexity (H. Clark, 1969) or relative infrequency (Carey & Considine, 1973; Wepman & Hess, 1969). Thus adult-like levels of processing may critically depend on additional experiences with these expressions.

A second striking developmental difference comes from the timing of the contrast effects in these two groups. While adults were sensitive to the presence of within-category contrast after the onset of the noun, children failed to show these effects until noun offset. This is consistent with prior studies highlighting an asymmetry between children's use of various cues. While children demonstrate adult-like efficiency in using lexical information to isolate referents (Fernald et al., 2010; Huang & Snedeker, 2011; Swingley, Pinto, & Fernald, 1999), their abilities to recruit higher level linguistic information like prosody (Snedeker & Yuan, 2008), discourse status (Pyykkönen, Mathews, & Järvikivi, 2007; Song & Fisher, 2005), and pragmatic inference (Huang & Snedeker, 2009) are typically delayed or altogether absent. These patterns suggest that children have greater difficulty generating inferences about how higher level linguistic representations constrain the interpretation of lower level ones. The complexity of these top-down procedures may also interact with children's slower processing speed (Kail, 1991; Kail & Salthouse, 1994). Developmental differences in this domain-general capacity may lead processes that occur with measurable delays in adults to become magnified in children.

General Discussion

This study explores the use of linguistic meaning and referential contrast in the real-time interpretation of scalar adjectives. We found that language comprehension in adults and children was influenced by both sources of information. Critically, we demonstrated that 5-year-olds understand that modification is warranted in the presence of multiple referents of the same kind and are able to apply this referential principle to constrain online comprehension. These findings add to a growing literature demonstrating that children use multiple sources of information to interpret language in real time (Arnold, 2008; Pyykkönen et al., 2007; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Song & Fisher, 2005; Trueswell et al., 1999).

Yet these findings are also somewhat surprising since prior studies have largely found that children fail to use the number of referents as a cue for resolving syntactic ambiguity (Choi & Trueswell, 2010; Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008). We have suggested that this asymmetry reflects the robustness of the contrast cue for predicting the production of modification through scalar adjectives but not through postnominal prepositional phrases. However, the results of the present experiment may also be potentially compatible with two other alternative explanations. In the following section, we lay out these possibilities and evaluate how they account for the full range of data on children's language comprehension.

Alternative Explanations

One possibility is that differences between current and prior findings reflect the relative position of the modifier and the noun. In our task, the modifier occurred *before* the upcoming noun ("*big* coin"), while in the syntactic ambiguity tasks (Trueswell et al., 1999; Hurewitz et al., 2000; Snedeker & Trueswell, 2004; Weighall, 2008), the modifier often occurred *after* the noun ("frog *on the napkin*"). This difference in ordering may have critical implications for real-time comprehension. Children may have a strong bias to establish reference immediately after identifying a noun, regardless of whether they have sufficient evidence to do so. Support for this proposal comes from children's behavior in the two-referent trials in Trueswell et al. (1999). In the presence of two frogs, children typically looked at one of the referents shortly after hearing the direct-object noun ("the frog") and whichever frog they happened to look at was often the one that they used to carry out the action. Thus, by committing to an interpretation immediately after hearing the noun, children may have resolved the referential ambiguity for themselves, even before encountering the ambiguous prepositional phrase. In contrast, by moving the modifier to a position prior to the critical noun in our study, we created a context in which the presence of referential contrast can be used to facilitate the prediction of an upcoming referent, rather than to revise a previous referential commitment.

However, recent cross-linguistic evidence suggests that differences in ordering may not fully account for prior failures to interpret modifiers. Choi and Trueswell (2010) presented Korean-speaking 5-year-olds with displays featuring two frogs and sentences like (4).

(4) naypkhin-ey kaykwuli-lul cipu-sey-yo

napkin-on frog pick up

"Pick up the frog on the napkin"

As in English, the prepositional phrase *naypkhin-ey* is temporarily ambiguous between the modifier and the destination interpretation. Furthermore, while the presence of the case marker *-ey* initially biases interpretation in favor of the destination, the final verb disambiguates the phrase as a modifier. Critically, since Korean is a head-final language, the modifier in these sentences precedes the noun. Thus, if prior failures to recruit contrast reflect difficulties with postnominal modifiers, then the use of prenominal modifiers should allow Korean-speaking children to avoid misinterpreting *naypkhin-ey*. Following this prepositional phrase, Korean-speaking adults generated few looks to the destination (the empty napkin), suggesting that they were sensitive to the referential requirements of the situation. Nevertheless, children continued to interpret the modifier as a destination, producing erroneous actions 54% of the time.

A second alternative explanation for the differences between previous and current studies is that they examine fundamentally different processes. While children fail to use contextual cues for syntactic parsing, they may be able to do so for the purpose of lexical prediction. This could reflect an asymmetry in how referential cues are integrated into different subsystems of language. Perhaps referential information is more systematically implicated in lexical processing, or perhaps the coordination of multiple cues is more easily accomplished here than it is during syntactic processing. This would predict that use of referential contrast should always emerge earlier in lexical processing than it does in syntactic processing.

While it is difficult to completely rule this possibility (to do so would simply require more data than is currently available on developmental sentence processing), such an account would be somewhat puzzling in light of children's early ability to match the number of referents in the scene with syntactic interpretations. This capacity supports the acquisition of quantifiers and plural morphology (Kouider, Halberda, Wood, & Carey, 2006; Wood, Kouider, & Carey, 2009) and features prominently in verb learning (Gleitman, 1990; Naigles, 1990; Yuan & Fisher, 2009; Arunachalam & Waxman, 2010). For example, Yuan and Fisher (2009) found that 2-year-olds use the syntax of an utterance to generate expectations about the appropriate number of referents in the scene. When introduced to a novel verb in a series of transitive sentences ("Jane blicked the baby"), children subsequently preferred to look at a scene with two referents (a girl pulling another girl's leg). Similarly, following a series of intransitive sentence ("Jane blicked"), children preferred to look at a scene with just one (a girl raising her hand). These findings suggest that prior failures to use context for syntactic parsing is unlikely to be caused by a uniform deficit in incorporating reference information in this domain.

Scalar Adjectives and Gricean Inferences

Finally, in addition to helping us understand how language comprehension develops, studying children's interpretation of scalar adjectives may clarify the processes that give rise to these contrast effects in the first place (Gregory et al., 2003; Grodner & Sedivy, 2011; Sedivy, 2003; Sedivy et al., 1999). Sedivy (2003) suggested that the presence of a contrast item facilitates reference restriction by causing listeners to generate a rapid Gricean inference (Grice, 1975): Hearing *tall* leads to the inferences that the adjective likely modifies a member of a contrastive set since modification would be overinformative for items that could be uniquely identified from the noun alone. Recently, Grodner and Sedivy (2011) found that these context effects are also sensitive to a listener's knowledge of the speaker. When listeners were told that the speaker had "an impairment that causes social and language problems," they no longer showed facilitation in the two-referent context. This is consistent with the idea that Gricean inferences follow from the assumption that speakers are rational and cooperative.

On this proposal, evidence of facilitation in the two-referent context is unexpected since children are notoriously poor at making other kinds of Gricean inferences (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Huang & Snedeker, 2009; Noveck, 2001; Papafragou & Musolino, 2003; C. L. Smith, 1980). Studies of scalar implicatures have shown that children are often insensitive to underinformative statements. For example, they find "Minnie started the puzzle" to be a perfectly acceptable way of describing a scene where she actually finished it (Papafragou & Musolino, 2003). In contrast, our results indicate that children have implicit knowledge of the informativity requirements that guide the use of scalar adjectives. Why would they succeed with these inferences but fail for other terms? One possibility is that Gricean inferences are a heterogeneous category with some expressions emerging earlier than others during development. However, this hypothesis would still have to explain why a particular class of inferences would be easier to acquire than another when both rely on the same logical mechanism.

A second possibility is that contrast effects for scalar adjectives do not depend on Gricean inferences but instead reflect the role that comparison sets in informing the meanings of these terms. To use an adjective like *tall*, most semantic theories suggest that we must establish a standard of comparison to determine what counts as *tall* in a given context (Bierwisch, 1987; Kamp & Partee, 1995; Kennedy, 1999). In some cases, we may use our stored knowledge of other tokens of a given kind (*tall* for a glass), while in other cases, we may use a salient contrast item in the scene (*tall* for a glass on this table). On this account, the contrast effects observed during children's processing are part and parcel of their interpretation of these terms. This provides an explanation for why children show contrast effects for scalar adjectives but fail to calculate other, parallel, Gricean inferences. Yet while removing one theoretical obstacle, this explanation creates another: It fails to account for evidence that adults show robust contrast effects for nonscalar adjectives like material terms (*plastic*, *leather*) and highly associated color-noun combinations (*yellow banana*, *pink eraser*; Grodner & Sedivy, 2011; Sedivy, 2003). These terms can be semantically interpreted without constructing a contrast set. However, the presence of a within-category item has been shown to generate comparable contrast effects in adults. Thus to resolve these issues, additional research is required to explore how these various pragmatic effects emerge during development.

Conclusion

The findings of this study provide evidence that children's interpretations of scalar adjectives are influenced by multiple sources of information including the meanings of these terms and their referential implications. This is consistent with prior evidence demonstrating that the many of the fundamental features that characterize adult language comprehension are also present and operational in the child listener. Critically, we found that 5-year-olds understand that modification is warranted in the presence of multiple referents of the same kind and are able to apply this referential principle to constrain interpretations during real-time comprehension. This contrasts sharply with prior studies that document that children fail to apply the referential principle during syntactic parsing. This divergence in findings suggests that further research on a wider range of phenomena is necessary to understand when and how children are able to use context to inform moment-to-moment language comprehension.

References

Altmann, G. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238. doi:10.1016/0010-0277(88)90020-0

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition, 108,* 69–99. doi:10.1016/j.cognition.2008.01.001

Arunachalam, S., & Waxman, S. (2010). Meaning from syntax: Evidence from two-year-olds. *Cognition, 114,* 442–446. doi:10.1016/j.cognition.2009.10.015

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child Development, 79,* 594–608. doi:10.1111/j.1467-8624.2008.01145.x

AQ: 2 Bierwisch, M. (1987). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives.* Berlin, Germany: SpringerVerlag.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54,* 592–609. doi:10.1016/j.jml.2005.12.008

Brown-Schmidt, S., & Tanenhaus, M. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32,* 643–684. doi:10.1080/03640210802066816

Carey, S., & Considine, T. (1973). *The domain of comparative adjectives.* Unpublished working paper, Massachusetts Institute of Technology.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domingues, & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (pp. 157–168). Somerville, MA: Cascadilla Press.

Choi, Y., & Trueswell, J. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology, 106,* 41–61. doi:10.1016/j.jecp.2010.01.003

Clark, E. V. (1972). On the child's acquisition of antonyms in two semantic fields. *Journal of Verbal Learning & Verbal Behavior, 11,* 750–758. doi:10.1016/S0022-5371(72)80009-4

Clark, H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76,* 387–404. doi:10.1037/h0027578

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunnen, & A. Zwicky (Eds.), *Natural language parsing.* Cambridge, England: Cambridge University Press. AQ: 3

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28,* 125–127. doi:10.3758/BF03203646

Donaldson, M., & Wales, R. J. (1970). On the acquisition of some relational terms. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York, NY: Wiley. AQ: 4

Ebeling, K., & Gelman, S. (1988). Coordination of size and standard by young children. *Child Development, 59,* 888–896. doi:10.2307/1130256

Farnham-Diggory, S., & Bermon, M. (1968). Verbal compensation, cognitive synthesis, and conservation. *Merrill-Palmer Quarterly, 14,* 215–227.

Fernald, A., Thorpe, K., & Marchman, V. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology, 60,* 190–217. doi:10.1016/j.cogpsych.2009.12.002

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96,* 263–284. doi:10.1016/j.cognition.2004.09.002

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1,* 3–55. doi:10.1207/s15327817la0101_2

Gregory, M. L., Joshi, A., Grodner, D., & Sedivy, J. C. (2003). *Adjectives and processing effort: So, uh, what are we doing during disfluencies?* Paper presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA. AQ: 5

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3). New York, NY: Academic Press. AQ: 6

Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference.* MIT Press: Cambridge, MA. AQ: 7

Huang, Y., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in five-year-olds: Evidence from real time spoken language comprehension. *Developmental Psychology, 45,* 1723–1739. doi:10.1037/a0016704

Huang, Y., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language, 38,* 644–661. doi:10.1017/S0305000910000206

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L., & Trueswell, J. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research, 29,* 597–626. doi:10.1023/A:1026468209238

AQ: 8 Kail, R. (1991). Development of processing speed in childhood and adolescence. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 25). New York, NY: Academic Press.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86,* 199–225. doi:10.1016/0001-6918(94)90003-5

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57,* 129–191. doi:10.1016/0010-0277(94)00659-9

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison.* New York, NY: Garland.

Klatzky, R., Clark, E., & Macken, M. (1973). Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual. *Journal of Experimental Child Psychology, 16,* 32–46. doi:10.1016/0022-0965(73)90060-X

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular–plural distinction. *Language Learning and Development, 2,* 1–25. doi:10.1207/s15473341lld0201_1

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703. doi:10.1037/0033-295X.101.4.676

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics, 53,* 372–380. doi:10.3758/BF03206780

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

Nadig, A., Sedivy, J., Joshi, A., & Bortfeld, H. (2003). The development of discourse constraints on the interpretation of adjectives. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 568–579). Somerville, MA: Cascadilla Press.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17,* 357–374. doi:10.1017/S0305000900013817

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

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

AQ: 9 Pyykkönen, P., Matthews, D., & Järvikivi, J. (2007). *Children's online comprehension of pronouns in spoken language: The role of verb semantics.* Poster presented at the 20th annual CUNY Conference on Human Sentence Processing, San Diego, CA.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effect of informativity expectations. *Journal of Psycholinguistic Research, 32,* 3–23. doi:10.1023/A:1021928914454

Sedivy, J., Demuth, K., Chunyo, G., & Freeman, S. (2000). Incremental referentiality-based language processing in young children: Evidence from eye movement monitoring. In S. C. Howell, S. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 684–695). Somerville, MA: Cascadilla Press.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147. doi:10.1016/S0010-0277(99)00025-6

Sera, M., Troyer, D., & Smith, L. (1988). What do two-year-olds know about the sizes of things? *Child Development, 59,* 1489–1496. doi:10.2307/1130663

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

Smith, L. B., Cooney, N., & McCord, C. (1986). What is 'high'? The development of reference points for 'high' and 'low'. *Child Development, 57,* 583–602. doi:10.2307/1130338

Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49,* 238–299. doi:10.1016/j.cogpsych.2004.03.001

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58,* 574–608. doi:10.1016/j.jml.2007.08.001

Song, H., & Fisher, C. (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language, 52,* 29–57. doi:10.1016/j.jml.2004.06.012

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71,* 73–108. doi:10.1016/S0010-0277(99)00021-9

Syrett, K., Kennedy, C., & Lidz, J. (2009). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics, 27,* 1–35. doi:10.1093/jos/ffp011

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. doi:10.1126/science.7777863

Trueswell, J., & Gleitman, L. (2004). Children's eye movements during listening: Evidence for a constraint based theory of parsing and word learning. In J. Henderson & F. Ferreira (Eds.), *Interface of language, vision, and action: Eye movements and the visual world.* New York, NY: Psychology Press. AQ: 10

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134. doi:10.1016/S0010-0277(99)00032-3

Trueswell, J., & Tanenhaus, M. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton & L. Frazier (Eds.), *Perspectives on sentence processing.* Hillsdale, NJ: Erlbaum. AQ: 11

Wales, R., & Campbell, R. (1970). On the development of comparison and the comparison of development. In G. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics.* Amsterdam, the Netherlands: North-Holland. AQ: 12

Weighall, A. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology, 99,* 75–95. doi:10.1016/j.jecp.2007.10.004

Wepman, J. M., & Hess, W. (1969). *A spoken word count (children ages five, six, and seven).* Chicago, IL: Language Research Associates.

Wood, J., Kouider, S., & Carey, S. (2009). Acquisition of singular–plural morphology. *Developmental Psychology, 45,* 202–206. doi:10.1037/a0014432

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20,* 619–626. doi:10.1111/j.1467-9280.2009.02341.x

(*Appendix follows*)

Table A1
Stimuli Used in Experiments 1 and 2

Item	Adjective	Target	Contrast: Two-referent trials	Contrast: One-referent trials	Competitor	Unrelated
			Critical trials			
1	Big - Small	Candy bar	Candy bar	Glove	Envelope	Sock
2	Big - Small	Card	Card	Sponge	Calculator	Videotape
3	Big - Small	Crayon	Crayon	Thread	Glue stick	Clothespin
4	Big - Small	Tape	Tape	Egg	Post-it	Lock
5	Big - Small	Teabag	Teabag	Lipstick	Band-aid	Sugar packet
6	Big - Small	Lego	Lego	Paper clip	Battery	Eraser (pencil)
7	Big - Small	Coin	Coin	Button	Stamp	Marshmallow
8	Big - Small	Spoon	Spoon	Toothpaste	Scissor	Fork
9	Tall - Short	Roll	Roll	Bookmark	Eraser (chalk)	Comb
10	Tall - Short	Ruler	Ruler	Candle	Vase	Doll
11	Tall - Short	Flashlight	Flashlight	Juice box	Bottle	Salt shaker
12	Tall - Short	Candy cane	Candy cane	Feather	Toothbrush	Flag
13	Tall - Short	Frame	Frame	Glass	Book	Can
14	Tall - Short	Toothpick	Toothpick	Nail	Cotton swab	Chalk
15	Tall - Short	Straw	Straw	Paint brush	Pencil	Chopsticks
16	Tall - Short	Brush	Brush	Flower	Remote	Lollipop
			Filler trials			
1	Big	Tomato	—	Apple	Ball	Ball
2	Big	Candy	—	Ring	Clip	Clip
3	Small	Stapler	—	Whisk	Knife	Knife
4	Small	Tape measure	—	Soap	Puzzle piece	Puzzle piece
5	Tall	Match	—	Chalk	Chess piece	Chess piece
6	Tall	Party hat	—	Carrot	Ice cream cone	Ice cream cone
7	Short	Jar	—	Postcard	Kleenex	Kleenex
8	Short	Pen	—	Highlighter	Stick	Stick

Received April 30, 2010
Revision received March 9, 2012
Accepted June 11, 2012 ■