

Children's Online Processing of Ad-Hoc Implicatures

Erica J. Yoon

ejyoon@stanford.edu
Department of Psychology
Stanford University

Yunan Charles Wu

ywu15@wabash.edu
Department of Psychology
Wabash College

Michael C. Frank

mcf Frank@stanford.edu
Department of Psychology
Stanford University

Abstract

Language comprehenders routinely make pragmatic inferences that go beyond the literal meanings of utterances. If A said "I ate some of the cookies," B should infer that A ate some *but not all*. Children perform poorly on experimental tests of scalar implicatures like this, despite their early-emerging sensitivity to pragmatic cues. Our current work explores potential factors responsible for children's successes and failures in computing pragmatic inferences. In two experiments, we used an eye-tracking paradigm to test children's ability to compute implicatures when they have access to contextual alternatives to the target word (Experiment 1), and when they hear prosodic cues that emphasize the contrast between the target and alternative (Experiment 2). We found that by the time children are four years old, they successfully identify the inferential target referent in this paradigm; with supportive prosodic cues, we saw evidence of success in three-year-olds as well. With sufficient contextual support, young children are capable of making online pragmatic inferences.

Keywords: Pragmatic cues; implicatures; cognitive development

Introduction

Language comprehension involves not only interpreting the literal meanings of words in utterances, but also understanding the communicative intentions behind what is said. Listeners make *pragmatic implicatures*, inferences about speakers' intended meanings that goes beyond the semantics of their utterances (Grice, 1975). One common type of implicatures, called *scalar implicatures*, involves scales built based on the knowledge of *lexical* alternatives (Horn, 1972). For example, if A says to B, "Some of the students failed the test," B may infer that A intended to say "Some, *but not all*, of the students failed the test." That is, A's use of the term "some" implicates that the stronger scalar alternative "all" is negated.

Whereas adults readily compute scalar implicatures (*SI*'s), children tend to perform poorly on *SI* tasks (e.g., Noveck, 2001; Papafragou & Musolino, 2003; Huang & Snedeker, 2009). For example, given a context in which three out of three horses jumped over a fence, adults reject a statement such as "some of the horses jumped over the fence" as infelicitous, whereas children typically judge it to be acceptable (Papafragou & Musolino, 2003).

Children's failures on *SI* computation are surprising, given their early-emerging sensitivity to the informativeness of utterances. For example, by around approx. five years, children adjust informativeness of their own expressions depending on the listeners' knowledge (Matthews, Lieven, Theakston, & Tomasello, 2006); reward speakers based on their informativeness (Katsos & Bishop, 2011); and provide more information when disambiguation between potential referents is difficult (Matthews, Butcher, Lieven, & Tomasello, 2012).

Given this body of research, it seems unlikely that children's lack of pragmatic ability per se causes their failures on *SI* tasks. What then causes children's failures, and what factors can help them succeed on implicature tasks? The current work investigates two potential factors: availability of alternatives to the current term, and cues that highlight the contrast between current term and its alternatives.

Implicature computation involves generating and negating alternatives to a given term. Upon hearing "some," the listener needs to generate a stronger alternative ("all") based on the lexical knowledge, and negate that alternative. One potential cause of children's difficulty with previous *SI* tasks is their lack of access to lexical alternatives to the term offered (Barner, Brooks, & Bale, 2011). Upon hearing a term such as "some", children may not be able to generate the relevant scalar alternative ("all") to negate, even if they know that there are alternatives to be negated. If this hypothesis is true, children might succeed on implicature computation if given access to alternatives in the context.

Indeed, there is evidence that children can compute *ad-hoc* implicatures, which depend on contextually-derived scales rather than lexically-derived ones (Stiller, Goodman, & Frank, 2014)¹. Children saw three faces, one wearing glasses and a top-hat, one wearing glasses only, and one with no item. When children heard: "My friend has glasses," 3.5-year-old children and older chose the face with glasses only as the referent above chance, successfully computing the implicature "My friend has glasses, *but not a top-hat*," given the contextual access to the stronger alternative (face with glasses and top-hat). In our current work we adopt a similar *ad-hoc* implicature paradigm for eye-tracking, to ask both about factors underlying the previously-observed developmental trajectory and about the decision-making processes underlying children's implicature computation.

Eye-tracking offers several advantages over purely behavioral measures for examining pragmatic inference. First, it is possible to track participants' gaze as an utterance is being produced, providing moment-by-moment data about responses to spoken language. Second, eye gaze reflects a more implicit measure of comprehension and hence allows for more direct developmental comparisons compared with behavioral choices that may reflect conscious deliberation.

A previous eye-tracking paradigm looking at *SI* computation in children (Huang & Snedeker, 2009) suggested that

¹These inferences are sometimes known in the pragmatics literature as "particularized implicatures," in contrast to "generalized" implicatures; here we use the term *ad-hoc* to remain agnostic with respect to this distinction.

children do not calculate SI during online language processing. For example, when they saw a girl who has two out of four (some but not all) of the socks and another girl who has three out of three (all) of the soccer balls, and heard "... the girl who has *some* of the soc...", unlike adults, children did not look more toward the girl with socks until they heard the disambiguating word "socks." Children might have struggled with SI computation from the lack of access to lexical scales (some-all), and the time constraint to process implicatures (in less than one second). Our current work uses a similar but simpler paradigm that tests children's inference of implicatures given scales that are set up contextually.

Thus, in addition to replicating previous research on ad-hoc implicatures in the online processing context, we are able to pursue two goals: measure the time-course of ad-hoc pragmatic inference; and identify potential factors that contribute to the developmental differences in implicature computation performance. In Experiment 1, we measure implicature performance across a wide developmental range; in Experiment 2, we examine the contribution of contrastive intonation on performance for a subset of age groups. Our findings suggest that young children are able to spontaneously generate implicature inferences when contextual support is present, even though these inferences are slower and harder to make than interpretations of semantically unambiguous utterances.

Experiment 1

Method

Participants Parents and their 2- to 5-year-old children visiting Children's Discovery Museum in San Jose, CA, were invited to participate in a short video study. The current sample comprised of children who were exposed to English at least 75% of the time as indicated by their parents. In addition, individual trials with more than 50% missing gaze data were excluded from analysis, and only participants who completed at least half of the trials according to this criterion were included in the analysis. These exclusion criteria led to a final sample of 108 (out of 113 participants): 24 2-year-olds ($M = 2.6$, range 2;1–2;11, 10 girls), 28 3-year-olds ($M = 3.5$, range 3;1–3;11, 19 girls), 24 4-year-olds ($M = 4.6$, range 4;1–4;11, 13 girls), 32 5-year-olds ($M = 5.4$, range 5;1–5;9, 9 girls). Children were given a sticker for participating in the study. We also tested fourteen adult participants, undergraduate students recruited through Stanford Psychology credit pool.

Stimuli and Design On each trial, participants saw two images: a target and distractor, which could either be an item with a single feature (e.g., a plate with only a carrot or only a banana), or an item with double features (e.g., a plate with a carrot and a banana). Each trial contained three phases: in the initial phase (8.5 seconds), two images were presented in silence for two seconds, then a pre-recorded voice said a sentence (e.g., "Look at these plates. Elmo's plate has a carrot."). Then, in the anticipatory phase (1.5 seconds), a chime sound played to induce participants' anticipatory gaze. In the following feedback phase (1.5 seconds), a character appeared

next to the target with an amusing sound effect. This outcome served to keep the task engaging for participants.

There were three types of test trials (pictured in Figure 1, bottom). In *inference* trials, the target item had a single feature (e.g., a carrot), and the distractor item had two features, one that was common with the target (e.g., a carrot) and the other feature that was unique (e.g., a banana). The test sentence named the feature that was common to the target and distractor. Thus, if participants understood that "Elmo's plate has a carrot" implicates "Elmo's plate has a carrot *but not a banana*," given the context, they should look more toward the target than the distractor, but otherwise look equally to both. There were two additional trial types, with semantically unambiguous targets: *Control-double* trials looked identical to inference trials, but the target and distractor were switched, such that the double-feature item was the target and the single-feature item was the distractor, and the test sentence named the unique feature on the target. *Control-single* trials presented two items that each had a unique single feature, and either could be the target. Children saw 4 inference, 4 control-double, and 4 control-single trials; adults saw 6 inference, 6 control-double, and 12 control-single trials.

There were six sets of item and feature types, and the features were named with nouns found on the MacArthur-Bates Communicative Development Inventory word list (Fenson et al., 1994). Two orders of the test trials were created, such that trial types and item types were counterbalanced and trial order was pseudo-randomized across the two orders.

Procedure Participants sat in a booster seat, approx. 60 cm away from the monitor of an SMI RED 120 Hz binocular remote eye-tracker. Participants were introduced to the task as watching a short video. The video began with a short Elmo video clip that lasted for 1-2 minutes, during which any necessary adjustments to the eye-tracker and participants' chair positions were made. The eye-tracker was then calibrated using a 2-point calibration and validation of the calibration points. Then participants were introduced to Sesame Street characters and told 'Today, [they] will show us lots of fun things. Are you ready? Let's go!' Following the introduction, participants saw two gaze-contingent practice trials, with unambiguous targets that differed from the test items. Then children watched 16 test trials and adults watched 24 test trials, as well as 4 filler photos of children playing and 2 Elmo video clips, presented at a pseudo-random points between test trials. The video lasted approximately 8 minutes.

Results and Discussion

Participants of all ages looked to the targets in both control-double and control-single trials reliably above chance (50%; Figure 1). There were age differences in the speed of looking at the target and the proportion of correct looking across both control trial types, congruent with the findings of Fernald, Pinto, Swingley, Weinberg, and McRoberts (1998), who found that children's efficiency of familiar word recognition increases over the second year; the current data show this de-

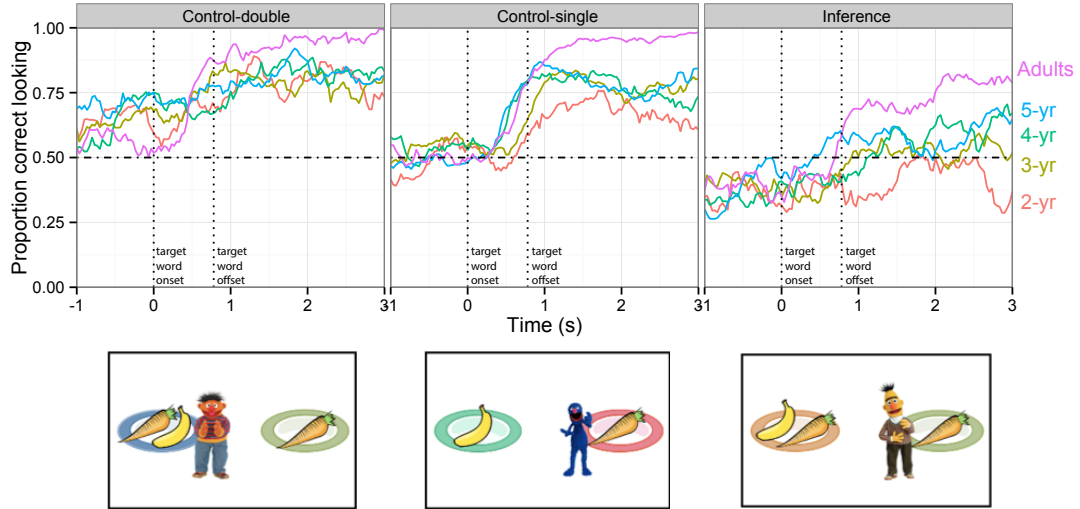


Figure 1: Proportion of 2- to 5-year-old children and adults looking to the target image as the utterance unfolds. Time 0 represents the target word onset. Proportion correct looking is defined by looks to the target divided by the total looks to both the target and the distractor. Bottom panels show example stimuli from each condition; the named character emerged at the end of the trial to mark the correct target.

developmental pattern continuing throughout childhood.

For inference trials, children of 4 years and above robustly looked to inferential targets (for 4-year-olds: $t(23) = 2.74$, $p = .01$). For example, upon hearing “Bert’s plate has a carrot,” older children identified the plate with only a carrot as the referent rather than the plate with a carrot and a banana, replicating Stiller et al. (2014)’s findings of ad-hoc implicature (though that study found successes in 3.5–4-year-old children as well). Although previous datasets are not directly comparable due to low-level differences in the task and materials, this finding is consistent with the hypothesis that children’s inferential ability might have been obscured in previous SI tasks due to the unavailability of lexical alternatives (e.g. “all” given “some”; Huang & Snedeker, 2009).

One additional finding emerged: two-year-olds’ looking at the target was below chance, rather than at chance: they tended to look more at distractors than targets ($t(23) = 1.93$, $p = .066$). Hence, two-year-olds did not even seem to consider inferential targets and distractors as equally likely referents. Rather, they did not disengage from distractors relative to their baseline bias prior to hearing the target word. We return to this pattern and speculate about the sources for the developmental changes observed in the General Discussion.

We fit a linear mixed-effects model² to measure the effects of trial type and age on the proportion of children looking to the target between 1 and 4s after noun onset (Table 1). We selected this time window because participants would have to wait until the end of target noun (0.8 seconds on average) to know they should switch to the inferential target, given

²All mixed-effects models were run using the lme4 package in R Studio Version 0.98.932. The random effects structure for this model was as follows: (trial type | subid) + (age | item)

the absence of a disambiguating continuation (e.g., “Elmo’s plate has a carrot *and banana*.”). Results of the mixed-effects model indicate significant main effects of trial type and age: participants looked to the target significantly less in inference trials compared to control-single trials, and across all trial types, participants’ looking to target increased with age.

To look more closely at the effect of age within inference trials, we fit a linear mixed-effects model³ to measure the effect of age on the proportion of children looking to the inferential target between the same time window (Table 2). Results indicate a significant main effect of age: participants looked toward the target at greater rates with increasing age.

We next analyzed participants’ reaction times (Fernald, Zangl, Portillo, & Marchman, 2008): We isolated trials on which participants were looking at the distractor at the point of disambiguation, and measured the average length of time prior to a shift to the target. Looks to the target were slower and overall lower in proportion in inference trials compared

³The random effects structure for this model was as follows: (1 | subid) + (age | item)

Table 1: Coefficient estimates from mixed-effects models predicting proportion of looks to target in Experiment 1.

Predictor	Value (SE)	<i>t</i> -value
Intercept (Control-single)	.60 (.05)	12.30
Age	.04 (.01)	3.24
Control-double	.10 (.06)	1.81
Inference	-.24 (.06)	-3.69
Age × Control-double	-.02 (.01)	-1.14
Age × Inference	.01 (.02)	.86

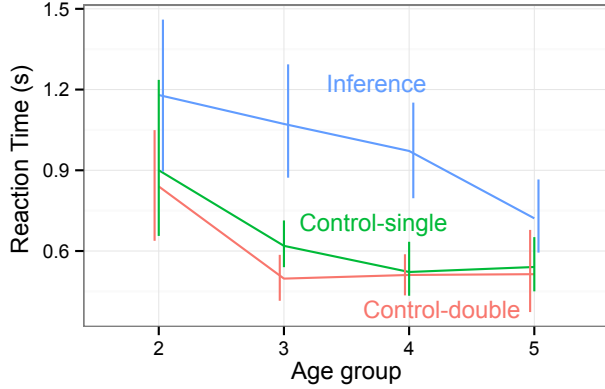


Figure 2: Average reaction times for first switches to target in trials in which participants were looking at the distractor (and not the target) at the target word onset.

to control trials across all ages (Figure 2). A linear mixed-effects model⁴ indicated significant main effects of trial type and age on the average time before switch to targets: switch to the target was slower in inferential trials compared to control trials, and the amount of time before switch to the target decreased as age increased. Hence, implicatures were generally slower and harder to process compared to the unambiguous semantic meanings, regardless of participants’ age.

Experiment 2

In Experiment 1, we found that children of 4 and above could compute implicatures given access to contextual alternatives. But children younger than 4 still struggled with implicature computation. In Experiment 2, we explored another factor that may improve children’s performance: prosody. Contrastive stress—a change in pitch, characterized by an initial drop followed by a rise—assists inference processing for both adults (Ito & Speer, 2008) and preschool children (Kurumada, 2014). For example, when preschoolers saw a picture of a zebra and a picture of an okapi that resembles a zebra, and heard “It LOOKS like a zebra” with a stress on the word “look,” they chose the picture of okapi if the context provided support. In the current experiment, we investigate whether a contrastive stress on the final noun (e.g., “Elmo’s plate has a CARROT”) in the inference trials would assist children in identifying the pragmatically-correct referent.

⁴The random effects structure for this model was as follows: (1 | subid) + (age | item)

Table 2: Coefficient estimates from mixed-effects models predicting proportion of looks to target in inference trials in Experiment 1.

Predictor	Value (SE)	<i>t</i> -value
Intercept	.36 (.05)	6.65
Age	.06 (.01)	3.96

Method

Participants Participants were recruited as in Experiment 1. For Experiment 2, we focused on 3- and 4-year-olds. Out of 57 initial participants, the final sample was chosen based on the same criteria as Experiment 1, and consisted of 17 3-year olds (8 girls), and 31 4-year olds (18 girls).

Stimuli, Design, and Procedure The stimuli, design and procedure were identical to Experiment 1, except for one change: Target nouns in inference trials were produced with contrastive stress (low-high-low pitch accent and longer duration, 1.2 seconds on average). Based on previous findings that children identify contrastive prosody based on the norms set within an experiment (Kurumada, 2014), we decided to include prosodic cues only in inference trials.

Results and Discussion

To determine the effect of prosodic cues on children’s inferential processing, we compared looking at targets across both Experiment 1 and 2 for inference trials (Figure 3). Children’s looking toward inferential targets increased slightly Experiment 2, especially towards the end of trials. In a post-hoc analysis, we split trials into an early and late period; Figure 4 shows this analysis directly. A linear mixed-effects model (Table 3) looking at the effects of age, trial type and window in Experiment 2 indicated a significant main effect of trial type, such that looking at target was lower in inference trials than in control trials. There was no significant interaction.

A closer look at inference trials in the two Experiments suggested that 3-year-olds identified inferential targets when there were supportive prosodic cues. In Experiment 2, both 3- and 4-year-olds looked at the correct inferential target above chance (for 3-year-olds: $t(16) = 2.47$, $p < .03$; Figure 4). This was in contrast to Experiment 1, in which 3-year-olds looked at the inferential target at chance level ($t(27) = 1.49$, $p = .15$). However, a linear mixed-effects model (Table 4) looking at the effects of experiment, age, and window in in-

Table 3: Coefficient estimates from mixed-effects models predicting proportion of looks to target in Experiment 2.

Predictor	Value (SE)	<i>t</i> -value
Intercept (Control-single and Early)	.80 (.04)	21.04
Age (4-year-old)	-.01 (.05)	-.19
Control-double	.03 (.05)	.59
Inference	-.26 (.05)	-5.07
Window (Late)	-.02 (.04)	-.46
Age × Control-double	.04 (.06)	.65
Age × Inference	.07 (.06)	1.07
Age × Window	.03 (.04)	.72
Control-double × Window	.05 (.06)	.82
Inference × Window	.10 (.06)	1.73
Age × Control-double × Window	-.07 (.07)	-.93
Age × Inference × Window	-.03 (.07)	-.51

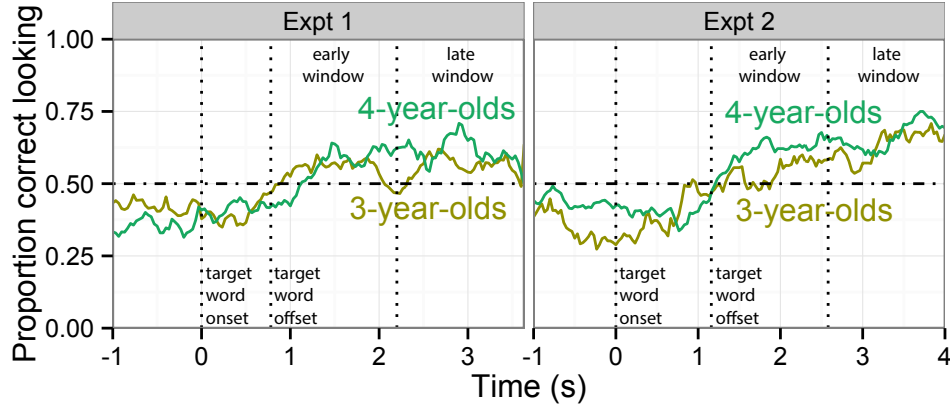


Figure 3: Proportion of 3- and 4-year-old children looking to the target image as the utterance unfolds, comparing across two Experiments.

ference trials in Experiments 1 and 2 did not indicate an interaction between experiment and window. Based on these results, overall, 3-year-olds seem to be more successful when contextual alternatives and explicit stress both support implicature computation, in line with Stiller et al. (2014)’s findings that 3.5-year-olds succeed in a similar behavioral implicature task. Nevertheless, the strength of contrastive stress cue for implicature computation is still uncertain, given the lack of a significant difference between the two Experiments. Looking at the effect of prosodic cue in a behavioral implicature task may help confirm its strength (see General Discussion for discussion on potential paradigmatic divergences).

General Discussion

Are young children able to make online implicature inferences? The current work looked at children’s processing of ad-hoc implicatures with an eye-tracking paradigm, and found that adults and children older than 3 years show robust looking toward the inferential targets, although at slower and overall lower rate compared to semantically unambiguous targets, whereas younger children still struggled with implicature computation. In Experiment 2, we found that prosodic

Table 4: Coefficient estimates from mixed-effects models predicting proportion of looks to target in Inference trials in Experiment 1 and 2.

Predictor	Value (SE)	<i>t</i> -value
Intercept (Expt 1 and Early)	.55 (.04)	14.52
Experiment (Expt 2)	-.01 (.06)	-.22
Age (4-year-old)	-.01 (.05)	-.23
Window (Late)	.03 (.05)	.54
Experiment × Age	.08 (.08)	1.08
Experiment × Window	.07 (.08)	.84
Age × Window	.03 (.07)	.44
Experiment × Age × Window	-.06 (.10)	-.60

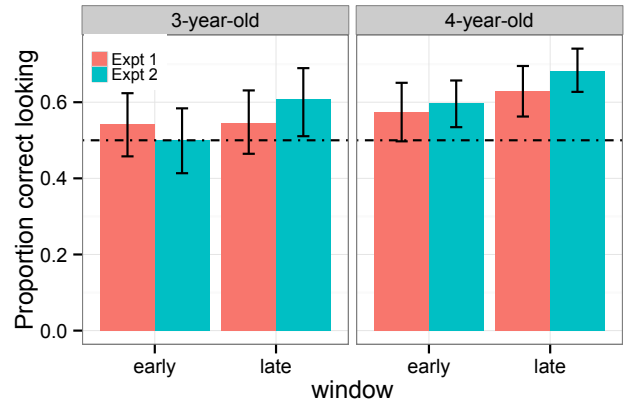


Figure 4: Looking to the target as a proportion of looking to the target and distractor in inference trials, averaged during two time windows (early: first half of the window between offset and end of trial, late: second half of the window).

cues with supportive context can help children as young as 3 to compute implicatures and identify inferential targets.

Our findings are broadly convergent with previous findings (Stiller et al., 2014), suggesting that the ability to make ad-hoc implicatures is fragile but measurable in 3-year-olds (with contrastive prosody) and more robust with 4-year-olds. Nevertheless, the rate of looking to the inferential target was overall lower in the current study than the accuracy rates in Stiller et al. (2014), even though the current paradigm was simpler (with two referent choices instead of three). This difference might be due to divergences between the two paradigms, since accuracy in our paradigm reflected graded patterns of looking rather than a single forced choice.

One unpredicted and intriguing finding was that 2-year-olds not only did not look at the correct inferential target, but looked more toward the distractor. One potential explanation comes from the inhibitory demands of our task. The two potential referents in inference trials differed in saliency:

The distractor was more salient as it contained an extra item (e.g., a carrot and a banana). Inhibitory control is difficult for children and continues to develop throughout the period we studied here (e.g., Davidson, Amso, Anderson, & Diamond, 2006). Congruent with this hypothesis, several recent studies suggest that inhibitory control might potentially affect word recognition in a number of similar eye-tracking paradigms (Yurovsky & Frank, 2014; Nordmeyer & Frank, 2013). Future work should thus address this possibility by explicitly manipulating the salience of potential pragmatic targets.

In sum, we found that preschool children were able to compute ad-hoc implicatures in contexts where there was substantial contextual support, in particular when they had access to lexical alternatives in the context, and when there were signals that emphasize the contrast between the target and its alternative. This supports the view that children's difficulties with SI are caused by their lack of access to linguistic scales (e.g., some-all; Barner et al., 2011).

The current work also sheds light on children's moment-to-moment processing of implicatures. Against claims that children are not able to compute SI's during online language comprehension (e.g., Huang & Snedeker, 2009), the current work suggests another possibility: implicature computation is delayed compared to interpretations of unambiguous utterances; however, with contextual access to scales relevant to implicature computation, children as young as 3 years could spontaneously generate implicature inferences.

Even young children are sensitive to the communicative intentions behind utterances they hear (Clark, 2009; Baldwin, 1993). Our work adds to the body of evidence suggesting that by preschool age they generate sophisticated pragmatic implicatures as well, even though these inferences are easily masked by other processing demands of specific contexts and situations. Overall, our current work takes one step further towards reconciling children's early-emerging communicative abilities with the complex pattern of successes and failures that they show in Gricean pragmatics.

Acknowledgments

We thank the parents, children, and staff at the San Jose Children's Discovery Museum. This work was supported by a Postgraduate Scholarship and Doctoral Program Fellowship provided by Natural Sciences and Engineering Research Council of Canada.

References

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental psychology*, 29(5), 832.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, 118(1), 84–93.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive

functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11), 2037–2078.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing*, 113–132.

Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, 3, 41–58.

Horn, L. R. (1972). *On the semantic properties of logical operators in english*. Unpublished doctoral dissertation, University of California, Los Angeles.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental psychology*, 45(6), 1723.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58(2), 541–573.

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

Kurumada, C. (2014). Contextual inferences over speakers' pragmatic intentions: Preschoolers' comprehension of contrastive prosody. In *Proceedings of the 36th annual meeting of the cognitive science society*.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two- and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in cognitive science*, 4(2), 184–210.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, 27(03), 403–422.

Nordmeyer, A. E., & Frank, M. C. (2013). Measuring the comprehension of negation in 2- to 4-year-old children. In *Proceedings of the 35th annual meeting of the cognitive science society*.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165–188.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, 86(3), 253–282.

Stiller, A., Goodman, N. D., & Frank, M. C. (2014). Ad-hoc

implicature in preschool children. *Language Learning and Development*.

Yurovsky, D., & Frank, M. C. (2014). Beyond naive cue combination: Salience and social cues in early word learning. In *Proceedings of the 36th annual meeting of the cognitive science society*.