Children's Processing of Ad-hoc Implicatures: A comprehensive investigation

Erica J. Yoon, Michael C. Frank

Department of Psychology, Stanford University

Author Note

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 three experiments, we used an eye-tracking paradigm (Experiments 1 and 2) and a tablet paradigm (Experiment 3) to test children's ability to compute implicatures when they have access to contextual alternatives to the target word. We found that by the time children are four years old, they successfully identify the inferential target referent in eye-tracking and tablet paradigms. Younger children still struggle with computation in our simplified tasks, however, and there is little evidence that this struggle is caused by inhibitory demands of the tasks.

Keywords: Pragmatics; implicature; eye-tracking; cognitive development

Children's Processing of Ad-hoc Implicatures: A comprehensive investigation

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 go 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 approximately 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?

One cue that may help children's implicature processing is availability of alternatives to the current term. On standard accounts, implicature involves generating and negating stronger alternatives to a given term. Upon hearing "some," the listener needs to generate a stronger alternative ("all") based on lexical knowledge, and then negate it. One potential cause of children's difficulty with previous SI tasks could be issues generating these alternative terms (Barner, Brooks, & Bale, 2011). If this hypothesis is true, children might succeed on implicature computation if they are given access to alternatives in the context.

Indeed, there is evidence that children can compute *ad-hoc* implicatures, which depend on contextually- rather than lexically-derived scales (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).

Given this initial evidence for preschoolers' pragmatic understanding, there are other important questions still to be answered. First, what happens in children's processing of implicatures, and how do children arrive at their inferential decision? Given the previous task where only one final answer is given at the very end of a trial with no timecourse or reaction time information, it is hard to infer speed or mechanisms

---

[1]These inferences are sometimes known in the pragmatics literature as "particularized" implicatures, in contrast to "generalized" implicatures. Here we use the term "ad-hoc" implicature as a descriptive term and remain agnostic with respect to the reality of this distinction.

of the computation process. Second, even in Stiller et al. (2014)'s simplified task, children younger than 3.5 years failed to choose the pragmatically felicitous referent. What is the cause of younger children's failure, and what makes older children perform better?

In the current work, we ask both about factors underlying the previously-observed developmental trajectory and about the decision-making processes underlying children's implicature computation. In Experiment 1, we measure implicature performance across a wide developmental range with an eye-tracking paradigm, and we replicate Stiller et al. (2014)'s findings that preschoolers compute ad-hoc implicatures. However, there are two surprising findings: younger children (2- to 3-year-olds) consistently fail to compute implicatures and are more biased towards the wrong answer; and children's performance in implicature trials is barely above chance, even for 5-year-olds. Experiments 2 and 3 address these concerns. In Experiment 2, we explore the cause of younger children's difficulty with implicature computation and explore one potential cause: inability to inhibit their response to more salient items. In Experiment 3, we use a tablet paradigm to confirm children's robust implicature computation, and compare their performances - accuracy and speed - across the two methodologies used.

Experiment 1

In Experiment 1, we use an eye-tracking paradigm to look at children's ad-hoc 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 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 in Experiment 1: measure the time-course of ad-hoc pragmatic inference; and identify potential factors that contribute to the developmental differences in implicature computation performance.

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 123 (out of 143 participants): 26 2-year-olds (M = FIXME, range FIXME, FIXME girls), 33 3-year-olds (M = FIXME, range

FIXME, FIXME girls), 29 4-year-olds (M = FIXME, range FIXME, FIXME girls), 35
5-year-olds (M = FIXME, range FIXME, FIXME girls). Children were given a sticker
for participating in the study. We also tested fifteen 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 (shown in Figure 1). 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.
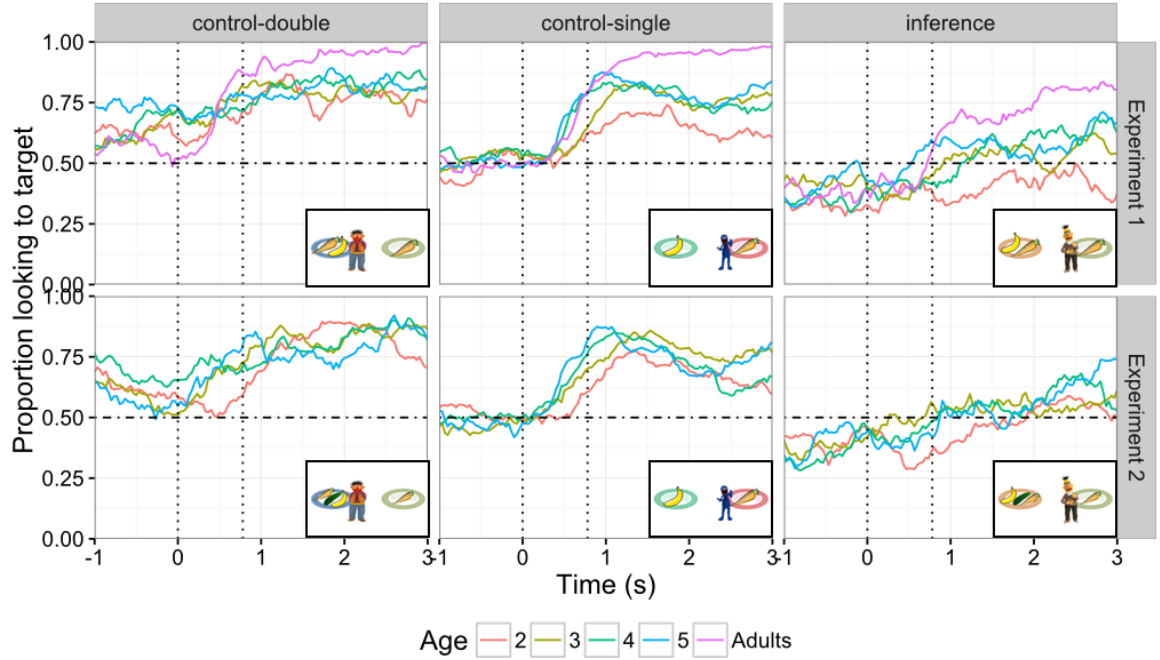
Results and Discussion

Figure 1. Proportion of 2- to 5-year-old children and adults looking to the target image as the utterance unfolds in Experiments 1 and 2 (rows), in different trial types (columns). Time 0 represents the target word onset, and time 0.78 represents the average target word offset. Proportion correct looking is defined by looks to the target divided by the total looks to both the target and the distractor. Example stimuli are shown in the bottom right hand corner for each condition; the named character emerged at the end of the trial to mark the correct target.

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.

For inference trials, children of 4 years and above robustly looked to inferential

targets (for 4-year-olds: $t(28) = 2.55$, $p =0.016$). 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. Although previous studies are not directly comparable due to low-level differences in the task and materials, our 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"; Barner et al. (2011)).

We additionally observed an unpredicted trend in two-year-olds' behavior: they did not disengage from distractors relative to their baseline bias prior to hearing the target word, and were marginally *below* chance in their overall performance ($t(25) =$ -3.66, $p = 0.001$).

Another noticeable pattern in children's responses was that, even though older children's correct look to inferential target exceeded look to distractor, it was lower than expected; in Stiller et al. (2014)'s paradigm, 4-year-olds selected the correct referent at approximately 75%, whereas even 5-year-olds looked to the target barely above 60% in the current paradigm.

We fit a linear mixed-effects model[2] to measure the effects of trial type and age on the proportion of children looking to the target between 0.8 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

---

[2]All mixed-effects models were run using the `lme4` package, version 1.1-10 (D. Bates, Maechler, Bolker, Walker, & others, 2014). The random effects structure for this model was as follows: `(trial type $|$ subid) + (age $|$ item)` All of our data and processing and analysis code can be viewed in the version control repository for this paper at: `https://github.com/ejyoon/FIXME`.

Predictor	Estimate	Std. Error	t value
Intercept	0.60	0.05	13.19
Control-double	0.13	0.06	2.24
Inference	-0.33	0.06	-5.24
Age	0.04	0.01	3.34
Control-double * Age	-0.02	0.01	-1.67
Inference * Age	0.03	0.02	1.94

Table 1

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting accurate looking to target in Experiment 1.

inferential target, given 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.

We next analyzed participants' reaction times (Fernald, Zangl, Portillo, & Marchman, 2008). We selected 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 in inference trials compared to both control trial types across all age groups. We next fit a linear mixed-effects model with the same structure as the previous analysis, but predicting reaction time rather than accuracy. This model again showed significant main effects of trial type ($\beta = 0.52$, $p < .05$) and age ($\beta = -0.08$, $p < .01$) on the average RT, with no interaction (largest $\beta = -0.06$, $p > .24$). Inference trials were generally slower compared to unambiguous

control trials, regardless of the participants' age, and participants reacted faster with increasing age generally across trial types.
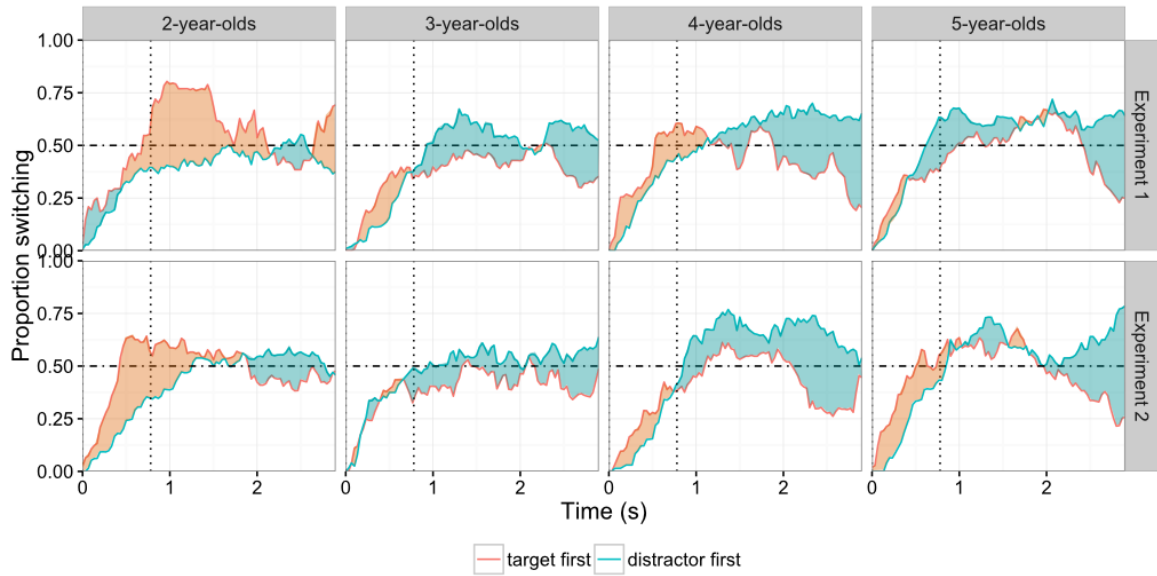

Figure 2. Onset contingency plot showing results from Experiments 1 and 2. Trials were divided depending on where the participant was first looking: the green line indicates trials in which participants looked at distractor first and made switch to target, and orange line target first and switched to distractor. the size of the green shaded region indicates more switches made from distractor-to-target than target-to-distractor. The size of the orange shaded region represents more switches made from target-to-distractor than distractor-to-target.

One question is whether younger children had difficulty shifting correctly from distractor to target, or whether they shifted incorrectly away from target to look at distractor. To explore this question, we looked at target- and distractor-initial trials separately, contingent on which item the child was looking toward from the onset of the target noun (Fernald, Thorpe, & Marchman, 2010). Top panel in Figure 2 shows the

mean proportion of participants that switched from where they started in Experiment 1. Thus, increase in shift on distractor-initial trials is a correct response, whereas increase in shift on target-initial trials is an *incorrect* response.

Across all age groups, there was an initial increase in shift for both target-initial and distractor-initial trials, until the offset of the target word. Then age groups diverged in their looking pattern, and two-year-olds' struggle was evident: whereas older children's switch to distractor decreased and switch to target increased, two-year-olds' switch to distractor continued to *increase* after the target offset, whereas shift from distractor to target stayed the same.

Experiment 2

In Experiment 1, we largely replicated Stiller et al. (2014)'s findings in an eye-tracking paradigm, and showed that adults and older children (4- to 5-year-olds) look toward the pragmatically felicitous based on ad-hoc implicature.

But younger children still struggled to look at the inferential target. Further, 2-year-olds not only did not look at the correct inferential target, but seemed to look if anything more toward the distractor. A potential explanation for this pattern comes from the inhibitory demands of our task. The two items in inference trials differed in salience: Since the distractor item contained an extra referent (e.g., a carrot and a banana), it was likely to be more salient. Supporting this idea, looking to the two-referent item was greater than chance during the baseline period of each trial. Perhaps 2- and 3-year-olds had difficulty disengaging from this more salient (and logically possible) distractor item in favor of the inferentially-correct target item. Inhibitory control is difficult for children and continues to develop throughout the period we studied here (e.g., Davidson, Amso, Anderson, & Diamond (2006)). In addition,

several recent studies suggest that inhibitory control might affect word recognition in similar eye-tracking paradigms (Nordmeyer & Frank, 2013; Yurovsky & Frank, 2014).

Experiment 2 sought to explore the question of whether inhibitory demands of the task caused younger children's failures. We increased the saliency of distractor even more by presenting three instead of two features. If two-year-olds' failures are due to the inhibitory demands, we predict them to *increase* their incorrect look toward distractor even more due to its salience. On the other hand, older children may increase their look toward inferential target, given an extra feature on the distractor that is negated by the implicature (e.g. "carrot *but not banana and not cucumber*").

Method

Participants. Participants were recruited as in Experiment 1. The final sample consisted of 102 (out of 126 participants): 26 2-year-olds (M = FIXME, range FIXME, FIXME girls), 30 3-year-olds (M = FIXME, range FIXME, FIXME girls), 36 4-year-olds (M = FIXME, range FIXME, FIXME girls), 27 5-year-olds (M = FIXME, range FIXME, FIXME girls).

Stimuli. The stimuli were identical to Experiment 1, except for one change: target items in inference trials and distractor items in control-double trials now had three features instead of two (see bottom panels in Figure 1).

Design and Procedure. The design and procedure were identical to Experiment 1.

Results and Discussion

As in Experiment 1, participants robustly looked to the targets in the control trials (Figure 1). Interestingly, age differences were no longer seen in control-double

trials. Importantly, there was no evidence of performance above *or* below chance for any of the age groups in inference trials (largest t: $t(28) = 1.78$, $p = 0.086$).

A linear mixed-effects model predicting accuracy based on age and trial type in Experiment 2 showed a significant main effect of trial type ($\beta = -0.25$, $p < .001$), such that looking at target was lower in inference trials than in control trials. There was no significant main effect of age or interaction between age and trial type (largest $\beta = -0.25$, $p > .19$). Thus, there was no improvement in looking to the correct target with age increase.

A linear mixed-effects model looking at the reaction times of making first switch from distractors to targets as in Experiment 1, found a significant main effect of trial type ($\beta = 0.58$, $p < .05$) and age ($\beta = -0.1$, $p < .05$) on the average RT, with no interaction ($\beta = -0.08$, $p > .23$). Thus, participants' looking was faster with increasing age, and looking at inferential targets was slower and overall lower compared to unambiguous targets, consistent with what was observed in Experiment 1.

Comparison between Experiment 1 and 2.   To determine the effect of saliency contrast on children's inferential processing, we compared looking at targets across both Experiment 1 and 2 for inference trials. A linear mixed-effects model predicting accuracy based on experiment, age, and trial type (Table 2) revealed significant main effects of trial type and age, but no interaction between Experiment 2 and any other variable. Thus, in contrast to our initial predictions, we did not find evidence of the effect of perceptual saliency on children's looking patterns. Why did two-year-olds not look at the distractor more than the target even though the distractor salience was increased? It is possible that even younger children benefitted from the strengthed implicature ("carrot *but not banana and not cucumber*"), or that they simply needed more time to explore and process all that is displayed.

Predictor	Estimate	Std. Error	t value
Intercept	0.60	0.05	12.92
Experiment 2	0.07	0.07	1.00
Control-double	0.13	0.06	2.11
Inference	-0.33	0.07	-4.94
Age	0.04	0.01	3.34
Experiment 2 * Control-double	0.02	0.08	0.25
Experiment 2 * Inference	0.08	0.10	0.86
Experiment 2 * Age	-0.02	0.02	-1.28
Control-double * Age	-0.02	0.01	-1.67
Inference * Age	0.03	0.02	1.88
Experiment 2 * Control-double * Age	0.00	0.02	0.13
Experiment 2 * Inference * Age	-0.02	0.02	-0.79

Table 2

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting accurate looking to target in Experiments 1 and 2.

In both Experiments 1 and 2, we found lower proportion of look toward inferential target than expected from older children, rarely exceeding 65%. One possible explanation is that children needed more time to completely process all the information presented in one trial, especially for Experiment 2. But based on the current findings, children, and even adults to some extent, did not reach the predicted level of referent identification across both Experiments that used eye-tracking. Thus, this could potentially be a methodological issue: eye-tracking may not be capturing the full process of implicature computation. To explore this possibility, Experiment 3 examined

children's implicature computation on another paradigm, using tablet.

Experiment 3

In Experiment 1, we confirmed that 4- and 5-year-olds reliably looked to the pragmatically felicitous targets above chance, and thus saw evidence of their implicature computation. Interestingly, however, the proportion of looking to target was generally lower than expected, never reaching beyond 75%, whereas older children in Stiller et al. (2014)'s paradigm selected the target much more robustly.

There are a few possible sources of this discrepancy: one potential source of this discrepancy is the specific type of stimuli used, such as pictures and utterances. (FIXME: Add a point about how our task was actually probably easier in terms of processing the pictures?) Yet another important difference is a methodological one: Stiller et al. (2014) used a behavioral selection paradigm, whereas we used an eye-tracking procedure. Hence looking time in eye-tracking may not be the most accurate measure of children's selection of the correct target, but eye-tracking does offer the advantage of looking at the reaction time. Is there a paradigm that can show both accuracy and reaction time, to yield comparable data to eye-tracking?

A tablet paradigm is a useful, engaging way to collect data from young children, in that it yields comparable data to that of behavioral paradigms, while making it possible to examine the accuracy and reaction time of children's judgments (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Thus, it can be a more naturalistic and engaging paradigm than eye-tracking, which requires children to sit still and watch similar trials repetitively rather than interact with the stimuli.

In Experiment 3, we examine children's ad-hoc implicature processing using a tablet paradigm, and we revisit the findings from all Experiments to compare the two

methodologies, tablet vs. eye-tracking.

Method

Participants. Participants were recruited as in Experiment 1, except that a partial sample was recruited from a local nursery school. The final sample consisted of 94 (out of 118 participants): 22 2-year-olds (M = FIXME, range FIXME, FIXME girls), 24 3-year-olds (M = FIXME, range FIXME, FIXME girls), 25 4-year-olds (M = FIXME, range FIXME, FIXME girls), 23 5-year-olds (M = FIXME, range FIXME, FIXME girls). FIXME: put this in a table format? Also, put a table of keep rate (how many participants were included in the analysis as a proportion of all who participated)

Stimuli. Items in the visual stimuli used the same set of images as in Experiment 1, presented on a tablet. Same auditory stimuli were used as in Experiment 1.

Design. The design was identical to Experiment 2, except that each participant saw two possible variations of the number of features for each trial type (2-vs-1 and 3-vs-1 for inference and control-double trials, 1-vs-1 and 2-vs-2 for control-single trials). There were no filler trials.

Procedure. An experimenter introduced children to the task using a tablet. Then they completed two practice trials, where they were asked to select an obvious, unambiguous referent (e.g., "cow" as opposed to "rabbit"), followed by 16 test trials.

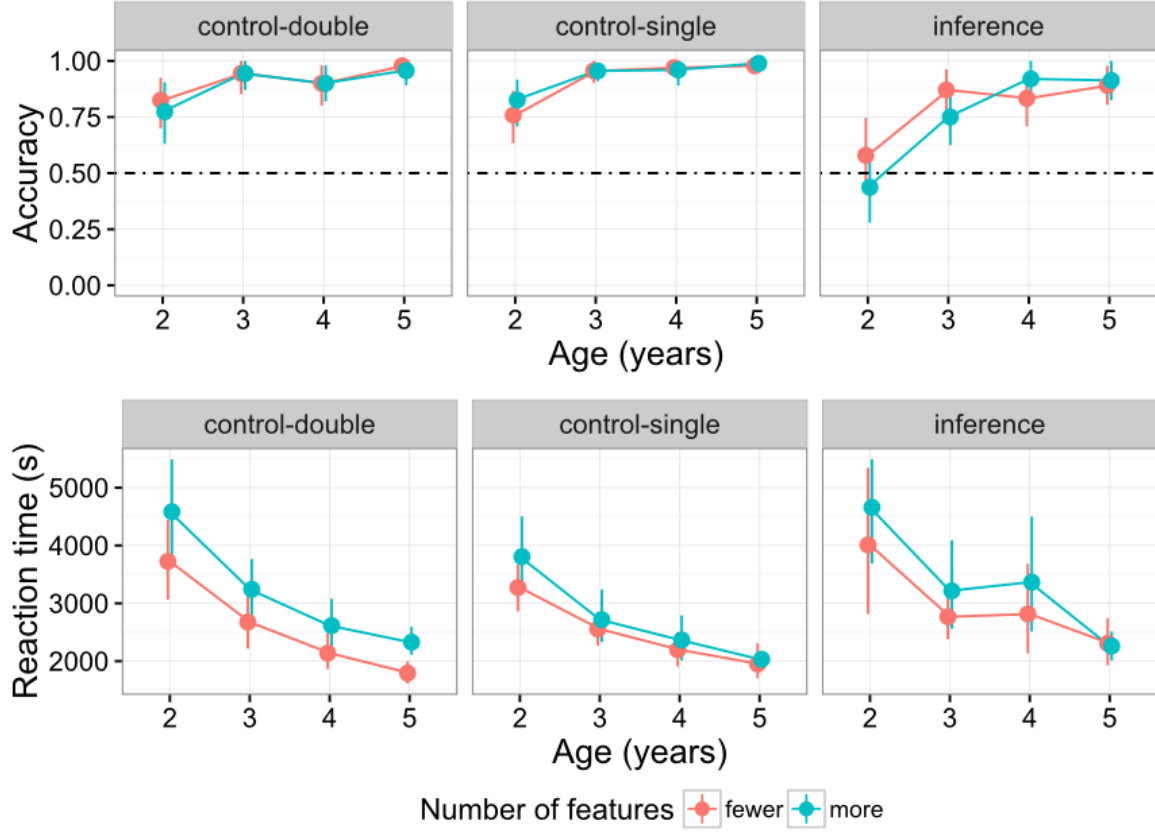**Results and Discussion**

Figure 3. Accuracy rates and reaction times in Experiment 3. Orange lines represent trials in which there were less features present (2-vs-1 for control-double and inference, 1-vs-1 for control-single) and green lines represent trials with more features (3-vs-1 for control-double and inference, 2-vs-2 for control-single).

A linear mixed-effects model predicting accuracy based on age, trial type and number of features present showed a significant negative interaction of inference trial and 3-vs-1 (β = -0.44, $p < .05$), and a significant interaction of inference trial, 3-vs-1, and age ($\beta = 0.1$, $p < .05$), indicating that performances on 3-vs-1 inference trials were poorer, especially for the younger children, two- and three-year-olds (Figure 3). This is

Predictor	Estimate	Std. Error	*t* value
Intercept	0.87	0.11	8.00
Control-double	-0.07	0.14	-0.48
Inference	-0.01	0.15	-0.08
Age	0.02	0.03	0.91
3-vs-1	0.05	0.14	0.33
Control-double * Age	0.01	0.03	0.28
Inference * Age	-0.02	0.04	-0.54
Control-double * 3-vs-1	-0.02	0.20	-0.08
Inference * 3-vs-1	-0.44	0.20	-2.20
Age * 3-vs-1	-0.01	0.03	-0.32
Control-double * Age * 3-vs-1	0.00	0.05	0.03
Inference * Age * 3-vs-1	0.10	0.05	2.05

Table 3

Predictor estimates with standard errors and significance information for a linear mixed-effects model predicting accurate selection of target in Experiment 3.

in line with our initial predictions, namely that younger children were expected to have more difficulty processing implicature when they need to inhibit responding to the greater perceptual salience of the distractor.

A linear mixed-effects model predicting reaction time based on age, trial type and number of features present showed a significant main effect of age ($\beta = -471.34$, $p < .01$). Thus, children identified the correct referent faster increasingly with age across all trial types.
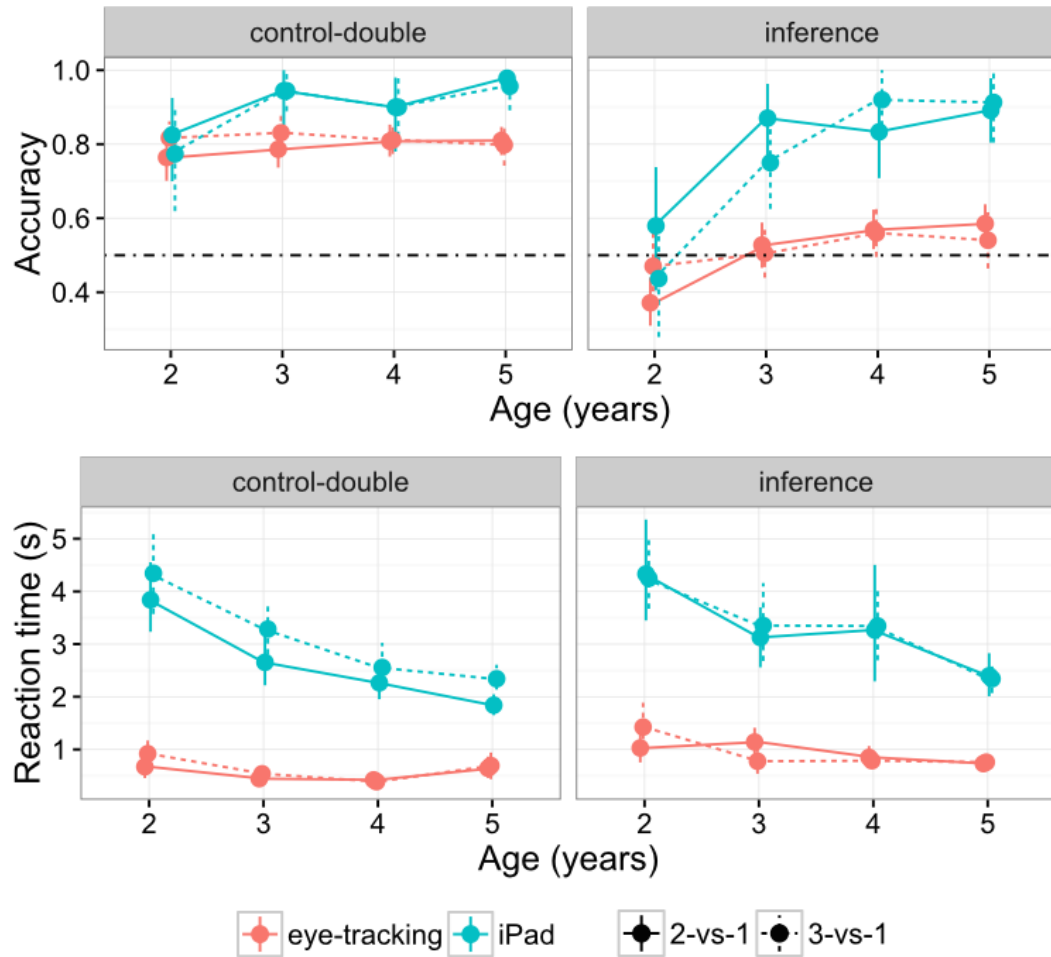
Comparing across Experiments 1, 2, and 3

Figure 4. Accuracy and reaction times rates for control-double and inference trials (columns) in eye-tracking vs. tablet (iPad) paradigm (colors). Solid lines represent 2-vs-1 conditions and dashed lines represent 3-vs-1 conditions.

Comparison of accuracy and reaction time measures from the eye-tracking paradigm (Experiments 1 and 2) and tablet paradigm (Experiment 3) shows that the developmental trajectory of implicature processing and robust implicature computation by older children are much more clearly revealed in the tablet paradigm than in the

eye-tracking paradigm (Figure 4). In fact, accuracy for inferential trials in the tablet paradigm was overall higher than what was reported in Stiller et al. (2014), possibly due to visual stimuli that were even more simplified (two referent candidates instead of three). In the tablet paradigm, both control and inference trials showed improvement in accurate referent identification over age, but much more prominently for inference trials. In eye-tracking, the improvement was observed in inference trials and was much subtler. Similarly, children responded quicker to the correct referent increasingly with age in both control and inference trials in the tablet paradigm, and this gain in efficiency was present in only inference trials in the eye-tracking paradigm.

Thus, the tablet paradigm more clearly showed children's success in implicature computation compared to eye-tracking. There are a few different possible interpretations of this result: eye-tracking may be better at capturing the ambiguity and uncertainty of ad-hoc implicature. Also, children might have needed more time to respond for the eye-tracking paradigm.

General Discussion

In the current work, we sought to take a detailed look at children's ability to compute implicature, and identify factors that contribute to children's success and failures as shown in previous research. We employed two different methodologies, eye-tracking and tablet paradigms, to confirm the developmental trajectories of implicature computation ability, and test working hypotheses about what may cause younger children to fail.

In Experiment 1, we used eye-tracking to replicate previous findings that children of 4 years and above can compute ad-hoc implicatures when they are given access to alternatives to the given term. There were two interesting observations: first,

two-year-olds looked *more* to the distractor than target, rather than look at both equally, as would be predicted if the looks were based on semantic ambiguity only. We thus hypothesized that younger children's failures are due to inhibitory demand of the task: to inhibit look to the more salient item (distractor) to look at the inferential target. Second, all age groups showed lower correct looks to the inferential target than expected based on previous findings on implicature computation (e.g. Stiller et al. (2014)). Given that the main discrepancy between the previous and the current task was methodology – naturalistic referent selection paradigm versus eye-tracking – we wanted to confirm children's behaviors in tablet paradigm that uses referent selection, but also examines speed of processing.

Experiments 2 and 3 tested the first question of whether inhibitory demand drives children's failure to reveal their implicature computation ability. We found partial evidence for this hypothesis: we did not see any significant differences between trials with less (Experiment 1) vs. more prominent contrasts (Experiment 2) between inferential target and distractor in the eye-tracking paradigm; however, in the tablet paradigm (Experiment 3), younger children tended to have more difficulty when distractor was more salient, providing some support for our initial hypothesis.

We compared between eye-tracking (Experiments 1 and 2) versus tablet paradigm (Experiment 3) to investigate the second question about methodological implications. We found that children showed more robust inferential target identification in the tablet paradigm compared to barely-above-chance inferential looking in the eye-tracking paradigm. Thus, the current work demonstrated the importance of employing different paradigms to allow complete investigation of development of an inferential process in question.

FIXME: what we now know about children's pragmatic abilities, based on

previous literature + the current work.

There are other remaining questions. Given partial evidence of inhibitory control as a cause of children's difficulty with implicature computation, more work is needed to identify other potential factors that contribute to children's implicature processing. FIXME: add what those factors might be.

FIXME: implications of methodological discrepancies.

Even young children are sensitive to the communicative intentions behind utterances they hear (Baldwin, 1993; Clark, 2009). Our work adds to the body of evidence suggesting that by preschool age they are able to 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.

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.

Bates, D., Maechler, M., Bolker, B., Walker, S., & others. (2014). Lme4: Linear mixed-effects models using eigen and s4. *R Package Version*, *1*(7).

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

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.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*, 1–17.

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* (PhD thesis). 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.

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

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.*