The trouble with quantifiers: Children's difficulty with "some" and "none"

Alexandra C. Horowtiz

Department of Psychology, Stanford University

Rose M. Schneider

Department of Psychology, Stanford University

Michael C. Frank

Department of Psychology, Stanford University

Address all correspondence to Rose M. Schneider, Stanford University, Department of Psychology, Jordan Hall, 450 Serra Mall (Bldg. 420), Stanford, CA, 94305. Phone: 650-721-9270. E-mail: rschneid@stanford.edu.

Abstract

Using language requires speakers and listeners to make pragmatic inferences that extend beyond the literal sense of the utterance. For example, adult listeners quickly and easily infer from the utterance "*Some* of the cookies are oatmeal raisin" that the others might not be. On the other hand, children have more difficulty with this statement; previous research has suggested that children succeed in making scalar implicatures using quantifiers in some cases but not others. Children's performance varies tremendously across studies and tasks however, limiting the number of possible direct comparisons between datasets. To address this issue, we designed a novel experimental paradigm, which both minimized task demands for participants, and enabled comparisons across different tasks. In Experiment 1, we used this paradigm to explore children's ability to compute both ad-hoc (contextual) and scalar (quantifier) implicatures, and found that while older 4-year-olds performed at ceiling for ad-hoc descriptions, they still performed poorly with scalar descriptions. Intriguingly, we also found bimodal and correlated performance with the quantifiers "some" and "none". In Experiment 2, we attempted to isolate children's sources of difficulty with these terms by including only scalar trials, and found that while performance increased, it was still low. We also observed once again that children's success on "some" and "none" trials were positively correlated. In Experiment 3, we explored possible sources of developmental difficulty in this task, combining our paradigm with tests of inhibitory control and quantifier knowledge. After controlling for age, we found that inhibitory control did not predict the ability to make scalar implicatures, but that children who had difficulty with "some" and "none" in an implicature task also had issues with these scalar terms in the quantifier knowledge task. Taken together, our results provide easily comparable developmental data on the development of pragmatic implicatures, and suggest that sources of difficulty over development

are not confined to making implicatures, but may be affected by quantifier knowledge and other pragmatic and processing demands.

Keywords: pragmatics, development

Introduction

In speech, adult listeners effortlessly make inferences that go beyond the literal sense of an utterance. For example, an adult who hears "I ate *some* of the cookies" would expect that I did not eat *all* the cookies, and left several for them. Similarly, an adult listener who hears "I ate the sugar cookies," would most likely assume that I ate *only* the sugar cookies, and not the other varieties. As implicatures, these two statements use a weaker literal description (in these cases, scalar and contextual) to imply that a stronger alternative is true. The first statement requires the listener to make a scalar implicature, one that relies on lexical scales such as quantifiers ("some" vs. "all") or modals ("possibly" vs. "definitely") (Horn, 1972). The second statement necessitates an ad-hoc implicature, in which a stronger description is negated by a contextually weaker description.[1] Previous research on children's on implicature comprehension has suggested that while children are able to make ad-hoc implicatures, they experience difficulty in computing scalar implicatures, and thus would have trouble inferring from "I ate *some* of the cookies" that cookies still remain. Here, we explore the developmental trajectory of children's ability to make both ad-hoc and scalar implicatures, as well as investigate factors influencing children's performance across scalar descriptions.

Children's processing of scalar implicatures, is an intriguing case study in pragmatic development. In contrast to adults' spontaneous and relatively automatic computation of scalar implicatures along quantifier lexical scales like <SOME, ALL>, children's performance on these same scales is variable even until fairly late in development (Noveck, 2001). However, it is possible that some of the paradigms used to test implicature comprehension in children might

---

[1]While Grice (1975) distinguished "generalized" and "particularized" implicatures, this distinction has been controversial. Here we use "ad-hoc implicature" as a term of convenience to describe contextually-supported inferences while remaining agnostic about the theoretical distinction.

obscure their pragmatic abilities, such as ones that require children to make truth judgments for complex propositions. In fact, children show a graded pattern of successes and failures across different tasks (Guasti et al., 2005; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004). For example, five-year-olds asked to rate the felicity of a statement by selecting the magnitude of a reward (rather than making a binary true/false decision) assigned only mid-sized rewards for true but pragmatically odd descriptions (Katsos & Bishop, 2011), suggesting that they do recognize that weak statements are less felicitous than stronger ones. Additionally, the wide range of methods and measures used to assess children's ability to compute implicatures does not allow for many direct comparisons in children's performance across different tasks (Table 1).

However, research on children's ability to compute scalar implicatures does indicate that their fragile performance may have less to do with their general pragmatic knowledge per se, and more to do with their knowledge of particular scales. The *Alternatives Hypothesis*, proposed by Barner and colleagues (Barner & Bachrach, 2010; ?, ?), posits that children's ability to compute scalar implicatures relies on their recognition of the relevant lexical alternatives (e.g., that use of the weaker term "some" conveys a direct contrast with the stronger alternative "all", thus implying *some but not all*). In other words, children's pragmatic inferences rely on their ability to consider relevant possible alternative word choices that could have been used in place of the ones the speaker chose. So even in supportive paradigms, if children cannot bring to mind "all" when reasoning about "some," they will fail to make an implicature.

Study	Scale(s)	Ages	Measure	Sentence Type	Main Finding
Noveck (2001)	non-necessity–necessity, possibility–impossibility, some–all	5;1–5;11, 7;1–8;0, 9;0–9;5, 10;0–11;7	Truth Value Judgment (*Yes, I agree* or *No, I do not agree*)	"Some giraffes have long necks"	Comprehension matters: Children demonstrate pragmatic competence by age 7 in evaluating a variety of plausible and implausible sentences.
Papafragou & Mussolini (2003)	some–all, two–three, start–finish	4;11–5;11 (Study 1) 5;1–6;5 (Study 2)	Felicity Judgment (*Did Minnie answer well?*)	"Some of the horses jumped over the fence" (when all of the horses jumped over the fence)	Support matters: Children were more likely to reject infelicitous weak descriptions for numbers, and for all types of weak descriptions in the task with more pragmatic support (informativeness training, context of competition, statements about specific events).
Papafragou & Tantalou (2004)	some–all, ad-hoc, encyclopedic	4;1–6;1	Felicity Judgment (Decide whether or not to award a speaker a prize)	*Did you color the stars?* "I colored some (when all were colored)	Scales matter: Children mainly withheld prizes for weak descriptions, and at higher rates for ad-hoc trials than other trial types.
Guasti, Chierchia, Crain, Foppolo, Gualmini, & Meroni (2005)	some–all	7;0–7;7	Truth Value Judgment: *Yes, I agree* or *No, I do not agree* (Studies 1–3), *Did Carolina say the wrong thing?* (Study 4)	"Some giraffes have long necks" (replication of Noveck (2001)), and scene descriptions, e.g. "Some monkeys are eating a biscuit" (when all are)	Context matters: 7-year-olds reliably computer implicatures for "some" after a training that increased their sensitivity to the informativeness of speakers' descriptions (e.g., calling a grape "fruit" instead of "grape") and when contextualized as evaluating a novice speaker novice speaker describing a scene.
Miller, Schmitt, Chang, & Munn (2005)	some–all	4;1–5;5 (Study 1) 3;6–5;10 (Study 2)	Direct Instruction Task (Study 1); Picture Matching Task (Study 2)	"Make some faces HAPPY/Make SOME faces happy/Make some HAPPY faces" (Study 1), "Shoe me where Pete made some faces HAPPY/Show me where Pete made SOME faces happy"	Prosody matters: In both tasks (completing the scene or selecting the referent), children reliably identified only a subset of the faces (out of four) when "some" was stressed, but not when it was unstressed.
Huang & Snedeker (2009)	some–all, two–three	5;2–6;1 (Study 1) 5;5–6;9 (Studies 2 & 3)	Eye-tracking referent selection	"Point to the girl with some of the socks" (when other girls and boys have shares of socks and soccer balls)	Time scale matters: Across studies, children were delayed in identifying the referent for scalar implicature trials, and accept and overlap between the meaning of "some" and "all."
Katsos & Bishop (2011)	some–all, ad-hoc	5;1–6;3	Binary Truth Value Judgment (Study 1); Ternary Truth Value Judgment (Study 2); Sentence-to-picture Matching Task (Study 3)	"The mouse picked up some of the carrots"	Measures matter: While children tended to a accept under-informative scalar and ad-hoc descriptions given a binary decision, they showed sensitivity to weaker statements given a ternary choice or picture matching task.
Barner, Brooks, & Bale (2011)	some–all, ad-hoc	4;0–5;0	Truth Value Judgment	"Are some of the animals sleeping?" (when all are)	Specificity matters: 4-year-olds accept weak ad-hoc and scalar descriptions. When preceded by restrictive "only", they reject ad-hoc descriptions but continue to accept that "only some" can mean *all*.
Skordos & Papafragou (2014)	some–all	4;9–5;8	Felicity Judgment (*Did the puppet answer well?*)	"Some of the blickets have a crayon" (when all of them do)	Comparisons matter: Children were more likely to reject infelicitous uses of "so" if they first heard "all" falsely refer to quantity (only 3/4 blackouts had crayons), but not is "all" referred falsely to the objects (e.g., "all of the blackest have a scarf").

Table 1
Review of previous literature on children's comprehension of implicatures.

Previous research has supported this hypothesis, with children's performance in implicature tasks increasing when they have stronger access to lexical alternatives. Evaluating performance in competitions where alternative outcomes are salient (Papafragou & Musolino, 2003) and using contextually-accessible (ad-hoc) scales (e.g., "the cat and the cow are sleeping" rather than "some animals are sleeping"; ?, ?) both help preschoolers make implicatures. The Alternatives Hypothesis also predicts interactions between the supportiveness of a task and children's performance. For example, even three-year-olds show evidence of computing implicatures for ad-hoc contextualized scales when the task is a referential forced choice between possible interpretations (Stiller, Goodman, & Frank, 2014). Preschoolers also show some preliminary evidence of computing scalar implicatures for quantifiers in a similar forced-choice paradigm, albeit with prosodic support (Miller, Schmitt, Chang, & Munn, 2005).

In sum, the Alternatives Hypothesis appears to provide a promising account of the current patterns of preschoolers' successes and failures in pragmatic implicature tasks. Nevertheless, work to date has varied widely in the particular scales, tasks, and measures that were used, and the developmental samples are relatively small while spanning several years of age. These concerns make it difficult to interpolate across findings and draw strong inferences about contrasts between contextually-supported (ad-hoc) and lexicalized (scalar) implicatures. In the current study, we create a paradigm which fills this gap, allowing exploration of children's processing of both ad-hoc and scalar implicatures with a single task.

In the current work, we designed a simple referent selection task in which children were asked to select which of three book covers they thought the experimenter was describing. Our design allowed us to fully counterbalance the trial types (ad-hoc vs. scalar descriptions crossed with implicature vs. unambiguous control targets) across participants, to examine both

within-subject patterns of responses and between-subject developmental patterns, and to reduce the demands of the task by having children select the implied referent among three visual alternatives.

In Experiment 1, we included both ad-hoc and scalar descriptions with implicature and control trials for each. Four–year-olds were strong on ad-hoc trials (similar to previous work, e.g. Stiller et al., 2014), but their performance on scalar implicature trials was very low. In Experiment 2, we ran the same task but replaced the ad-hoc trials so that all of the descriptions used scalar quantifiers. We found developmental increases in performance for each trial type, and higher performance on implicature trials for 4-year-olds in this scalar-only version of the task. Children's pattern of responses on scalar implicature trials was bimodal and strongly correlated with their performance on "none" (scalar control) trials, providing some clues about the factors underlying success in scalar implicatures. In Experiment 3, we explored two alternatives (inhibitory control problems and lack of quantifier knowledge) that could be contributing to children's difficulty with the quantifiers "some" and "none". Overall, our findings suggest that while preschooler's computation of scalar implicatures difficult for preschoolers can be supported by stronger recognition of the lexical alternatives, that their difficulty with scalar terms may be rooted in an unclear and unestablished quantifier scale.

Experiment 1: Ad hoc and scalar implicature computation in children

Given the difficulty in equating results on children's computation of implicatures across different methods and paradigms, we created a single task that could be adapted to investigate both ad hoc and scalar items in one task. This task involved one set of stimuli presented in the same order to all participants; however, the particular items queried were counterbalanced across participants. Thus, with one set of stimuli we created a novel paradigm wherein we could directly

compare children's performance on both ad hoc and scalar implicatures in a single experimental session. In Experiment 1, we included questions about ad hoc and scalar implicatures within one session, and found that this paradigm was appropriate for ad hoc items, but children still experienced difficulty computing scalar items.

Methods

   Participants. A planned sample of 48 children was recruited from a university preschool. These children were drawn from two age groups: twenty-four 4.0 – 4.5-year-olds (M = 4;2, median = 4.19, SD = 0.14) and 24 4.5 – 5.0-year-olds (M = 4.74, median = 4.73, SD = .16). Two children were excluded from the final sample for not completing the task, and one additional child was excluded due to experimenter error. The final sample was comprised of (XX males and XX females) with English as their primary language. Across all experiments, no child completed more than one session of the task.

   Stimuli. Stimuli for all experiments were created to be appropriate for questions about both ad hoc and scalar implicatures, allowing the experimenter to use one set of stimuli for both kinds of items in one experimental session. The experimental stimuli consisted of a set of printed pictures of three book covers with four familiar items on each cover. In each trial, one book cover contained four items of the same kind (e.g., four cats), another book cover contained four items of another kind (e.g., dogs), and the final cover contained two items of a new set and two items repeated from one of the other book covers (e.g., two birds and two cats). An example of the stimuli can be seen in Figure 1. Experiment 1 consisted of 18 test trials preceded by one test trial, with three book covers containing only one familiar item. All items on the book covers were familiar to children, and were able to be identified. All participants saw the same book covers in the same order.
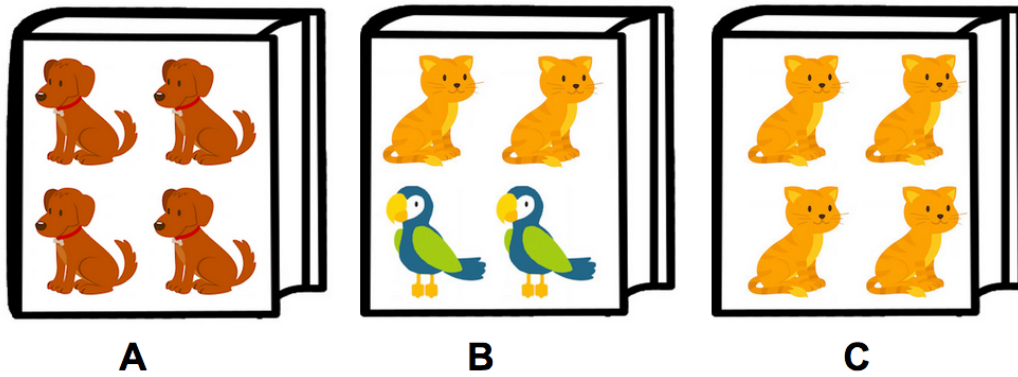
Figure 1. Example trial stimuli used in all experiments. Children received a clue from the experimenter about which book she had in mind and responded based solely on the clue; this was either an ad hoc or a scalar description of a book with either an unambiguous or implicature target.

Procedure. Participants were tested in individual sessions in a quiet room at their nursery school. The experimenter introduced the study as a guessing game, and explained that the child would receive a hint about which book cover the experimenter had in mind. In the instructions for the task, the experimenter emphasized that the child would only receive one clue about what book the experimenter was describing, and they had to use that clue to make their decision. All participants saw the same book covers in the same order; however, scripts with both ad-hoc and scalar trials were counterbalanced across participants. A breakdown of trial types and sample scripts can be seen in Table 2.

Prior to the test trials, children were familiarized to the task with a practice trial. The practice trial consisted of three book covers, each displaying a single unique and familiar item. After children had successfully completed the practice, children saw 18 test trials with stimuli of books containing sets of familiar items. At the start of every trial, the experiment would provide

Condition	Trial type	# trials, Expt. 1	# trials, Expt. 2	Statement: "On the cover of my book, ..."	Target
Scalar	implicature	4	6	"...some of the pictures are cats"	B
	all	2	6	"...all of the pictures are cats"	C
	none	2	6	"...none of the pictures are cats"	A
	unambiguous 'some'	2		"...some of the pictures are birds"	B
Adhoc	implicature	4		"...there are cats"	C
	distractor	2		"...there are dogs"	A
	comparison	2		"...there are birds"	B

Table 2

Study designs for Experiments 1 and 2, using script examples for the trial set pictured in Figure 1.

the child with either an ad-hoc or scalar description of one of the books and instruct the child to point to the book she was describing. If the child pointed to more than one book, or the response was otherwise ambiguous, the experimenter emphasized again that she was talking about just one book, and that they should choose the single book she was describing.

In ad-hoc trials (eight total), the experimenter's descriptions of the target book used names of the pictured objects, providing contextual support for the target. Ad-hoc control trials referred to an unambiguous target (e.g., "On the cover of my book, there are dogs" in Figure 1). while implicature trials required the child to reason about the speaker's meaning given the ambiguous utterance (e.g., "On the cover of my book, there are cats", which could refer to either the book containing only cats or the book containing cats and birds). In these critical trials, children had to understand that the speaker could potentially be talking about either the book with four or two of the named object, but that by opting to describe only one kind of object she was referring to the cover with four of the same object; otherwise, she would have mentioned both kinds of objects, or the ones unique to that cover (i.e., birds).

In scalar trials (ten total), the experimenter described the target book with quantifiers. For
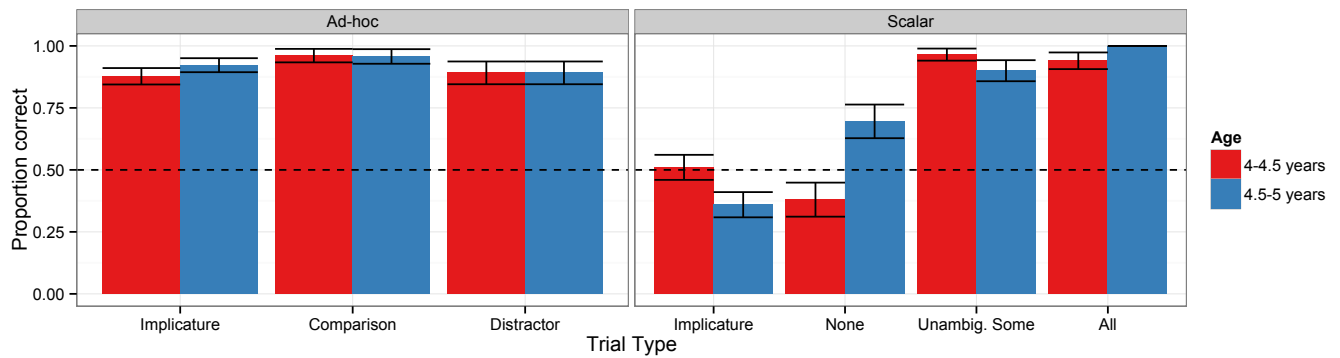
Figure 2. Proportion of correct responses by each age group across all trial types and split by implicature type. Error bars show standard error of the mean. Children performed almost at ceiling in ad-hoc implicature conditions, but struggled in scalar trials, particularly in the "some" and "none" trials.

scalar items, control trials referred to unambiguous targets with the quantifiers "all" and "none" (e.g., "On the cover of my book, *all/none* of the pictures are cats") or an unambiguous referent of "some" (e.g., "On the cover of my book, *some* of the pictures are birds"). On critical scalar implicature trials, the experimenter used the weak quantifier "some" to reference the item pictured across two book covers (e.g., "On the cover of my book, *some* of the pictures are cats"). These trials required the child to reason that because the speaker used the weak quantifier "some", some must be referring to the book picturing only two of the named target, or else she would have used the strong quantifier "all".

All participants saw image sets in the same order; however, these image sets were counterbalanced for target location and book triad position. Description condition and trial-type were further randomized across participants, and were spaced to avoid immediate repeat trial types. Children did not receive feedback after the test trial.

Results

Children's performance on all trial types can be seen in Figure 2. Across all trials, children's performance was coded as correct if they selected the image consistent with either the ad-hoc or scalar description. Children were at ceiling making ad-hoc implicatures, which is consistent with previous research suggesting that children are able to succeed in making such implicatures when they have access to the relevant lexical alternatives (Stiller et al., 2014). Children's performance across ad-hoc trials provides strong evidence that our novel paradigm is an appropriate measure for such items.

In contrast to their success in making ad-hoc implicatures, children struggled in scalar trials. Although they succeeded in "all" and "unambiguous some" trials, children performed near chance on "some" and "none" trials.

We ran a logistic mixed effect model, predicting a correct response as an interaction of age, condition (ad-hoc or scalar) and trial type (implicature or control), with random effects of participant and trial type. We found that performance was slightly lower for scalar trials than ad-hoc trials ($\beta = -8.02$, $p = .09$), and that there was a significant interaction between condition and trial type, such that performance was significantly worse on scalar implicature trials ($\beta = 16.45$, $p < .02$). We also found a significant 3-way interaction between condition, trial type, and age, such that performance on scalar implicature trials decreased with age ($\beta = -4.16$, $p < .01$). There were no significant effects of adding trial order (trials in the first half vs. second half of the experiment), indicating that performance did not change throughout the course of the experiment.

Interestingly we found that when children's performance in scalar trials varied consistently. To examine their patterns of responses more closely, we ran Hartigan's dip test and found significant bimodal distributions for both *some* ($D = .15$, $p < .0001$) and *none* ($D = .20$, $p < .0001$).

This suggests children did not respond at chance in scalar trials, but either consistently correctly or incorrectly. Additionally, children's success on *some* and *none* trials was highly correlated ($r = .47$, $p < .001$), such that children who performed better on some trials also tended to perform better on none trials. Performance on *none* and *all* trials ($r = .11$, $p = .45$) and *some* and *all* trials ($r = .01$, $p = .95$) was not correlated.

Discussion

The results of Experiment 1 indicated that while children were easily able to make ad-hoc implicatures in our task, they had difficulty making scalar implicatures. This pattern of performance was puzzling, given both our efforts to reduce task demands and children's striking success in ad-hoc trials. In spite of having access to both visual alternatives (the three selection choices) within each trial and lexical alternatives across all trials, children were still at chance in making scalar implicatures in our task.

Even more intriguing was the unexpected developmental change we observed on *none* trials. We chose to include "none" as an unambiguous control quantifier, but found that children performed at chance for this scalar term as well. These results are supported by previous work suggesting that even older preschoolers struggle with negation occurring in arbitrary contexts (Nordmeyer & Frank, 2014). Given children's difficulty with "none" and the strong positive correlation between *some* and *none* trials, it is possible that making implicatures necessitates some familiarity with both ends of the quantifier scale (*none – some – all*).

In addition to understanding the extremes of the quantifier scale, two other possibilities – namely, lack of quantifier knowledge, and inhibitory control – may also account for children's decreased performance on scalar items. However, we also wondered if these results were the effect of including both ad-hoc and scalar quantifier descriptions within one experimental session.

It is possible that children's success on ad-hoc trials may have to a misinterpretation of a scalar description (e.g., "On the cover of my book, some of the pictures are cats") to an ad-hoc one ("On the cover of my book, there are cats"). There is some evidence that children will sometimes rely exclusively on one strategy to solve problems, even if another strategy is more appropriate (citing Ann's work from the summer, maybe?).

To further explore the possibility that children's performance on scalar trials might have been influenced by ad-hoc trials, we removed all ad-hoc trials from our task, and ran a scalar-only version of the the study.

Experiment 2: Isolating scalar implicatures

In Experiment 2, we pursued the possibility that children might be failing in making scalar implicatures as a result of competing ad-hoc descriptors contained in the same experimental session. Additionally, we expanded our sample age-range to 3–5 years in order to more fully explore the developmental trajectory of scalar implicature comprehension.

Methods

Participants. We recruited a new sample of 50 participants from Bing Nursery School at Stanford University: 12 3.0–3.5-year-olds (M=3;4), 12 3.5–4.0-year-olds (M=3;8), 14 4.0–4.5-year-olds (M=4;3), and 12 4.5–5.0-year-olds (M=4;8). One additional child was excluded for stopping the task early.

Stimuli. Stimuli were identical to Experiment 1. The only changes made in experimental protocol were to the scripts; ad-hoc trials were dropped, and all 18 test trials were converted to scalar descriptions (Table XX). In Experiment 2, the 18 test trials consisted of six control *all* trials (e.g., "On the cover of my book, *all* of the pictures are cats"), six *none* trials (e.g., "...*none* of the
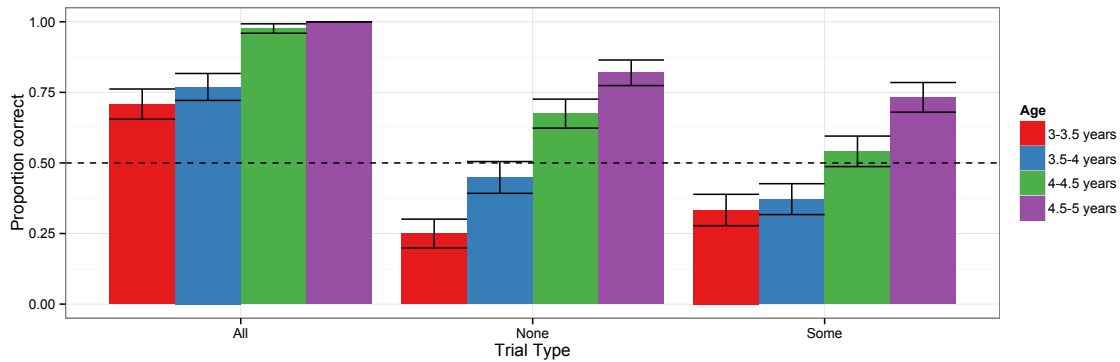
Figure 3. Proportion of correct responses by each age group across all scalar trial types. Error bars show standard error of the mean.

pictures are cats"), and six scalar implicature trials ("...*some* of the pictures are cats"). We removed the unambiguous references to *some* to more effectively counterbalance trials; in scalar implicature trials, the quantifier always referenced the item pictured across two book covers (e.g., in Figure 1, children heard references to *none, some,* or *all* cats). As in Experiment 1, image sets were presented in a fixed order, counterbalanced for both target location and triad order. Participants were randomly assigned to one of three scripts, with a pseudo-randomized trial order such that every book set was referred to by each quantifier type, and the same trial type never immediately repeated.

Procedure. The procedure was identical to Experiment 1.

Results

In Experiment 2, children's performance increased with age (3. Performance was highest in *all* trials, with all age groups significantly above chance. However, performance was still low in

both *none* and *some* trials, with only the oldest age group performing above chance for *none* trials ($t = 3.09$, $p = .01$), but only marginally above chance for *some* trials ($t = 1.85$, $p = .09$).

We ran a logistic mixed effects model, predicting correct responses as an interaction of age and trial type (*all, some,* or *none*), with random effects of trial type by participant. However, we found that the only significant effect that emerged was age, such that performance increased across trials as children got older ($\beta = 20$, $p < .001$). Adding trial order (first or second half of the experiment) to the model did not interact with any of the variables, indicating the performance did not change over the course of the experiment.

Consistent with the findings from the mixed effects model, we again found significant bimodal patterns of responses for both *some* ($D = .12$, $p < .0001$) and *none* ($D = .15$, $p < .0001$) trials. And again, these trial types were highly correlated with one another ($r = .52$, $p < .001$).

As a exploratory analysis, we ran another version of the mixed effects model removing the random effect of trial type. In addition to a main effect of age ($t = 1.88$, $p < .01$), this model revealed that performance on *some* trials was lower than *all* trials ($t = -7.69$, $p < .01$), and marginally reduced from *none* trials ($t = 3.03$, $p = .09$). We also found interactions between trial type and age, such that there was a greater difference between younger children's performance on *some* and *all* trials ($t = 2.84$, $p < .001$), and *some* and *none* trials ($t = 0.90$, $p = .05$).

Discussion

In Experiment 1, we observed success in children's computation of ad-hoc, but not scalar implicatures. To explore whether children's performance in making scalar implicatures might be hindered by the presence of both ad-hoc and scalar items in the same session, leading them to misinterpret all descriptions as contextual. In Experiment 2, we excluded ad-hoc items, and replaced them with scalar descriptions.

When faced with only scalar implicatures, children's performance was marginally better than in Experiment 1, with success in making a scalar implicature positively correlated with age. However, we still observed decreased performance in both the *some*, as well as the unambiguous *none* trials. Additionally, we again found bimodal and correlated patterns of responses in these two trials, with children making errors in implicature trials also failing on *none* trials.

The results of Experiment 2 indicated that children have difficulty with making scalar implicatures beyond dealing with competing contextual descriptors, but the cause of this developmental change is not clear. One possible explanation for this effect is that children's knowledge of the full quantifier scale is not yet adult-like in their preschool years, and that the meanings of these scalar terms may yet be established. Another possibility is that children's performance might suffer on these task from developing inhibitory control, with children not being able to overcome the impulse to select the target noun, regardless of the quantifier user.

Based on our findings of bimodal and correlated performance and patterns of errors, we decided to explore these two alternatives in Experiment 3, and include measures of both inhibitory control and quantifier knowledge in the same session.

Experiment 3: Inhibitory control and quantifier knowledge measures

In an effort to explore the particular sources of difficulty for children in making scalar implicatures, in Experiment 3 we supplemented our implicature task with an inhibitory control task (Zelazo, 2006), and a quantifier-knowledge task (Barner, Chow, & Yang, 2009) in a within-subject experiment.

Methods

Participants. We recruited a new planned sample of 72 children from Bing Nursery School at Stanford University. Once again, we included children from 3–5 years: 21 3–3.5-year-olds (M = 3;3), 15 3.5–4-year-olds (M = 3;8), 19 4–4.5-year-olds (M=4;2), and 17 4.5–5-year-olds (M = 4;7). Twelve children were excluded from the final sample for having participated in either Experiments 1 or 2. Nine children completed one of the tasks in a different testing session.

Stimuli. The stimuli for the implicature task were identical to Experiments 1 and 2. Once again, we excluded ad-hoc items, and ran a scalar-only version of the task. Our measure of inhibitory control was the Dimensional Change Card Sort (DCCS, (Zelazo, 2006)), with materials drawn from that study. The 14 laminated sorting cards used in the study were 11cm x 7 cm, and consisted of 7 red rabbits and 7 blue boats. These cards were put into plastic sorting trays, one with a red boat, and the other with a blue rabbit. To assess quantifier knowledge, we used the Give-Quantifier task (Barner et al., 2009). Stimuli for this task consisted of three different sets of plastic fruits (8 oranges, 8 bananas, and 8 strawberries) and a red plastic plate. These fruits were grouped together by kind at the start of each trial.

Procedure. The procedure for our implicature task was identical to Experiment 2, the scalar-only version of the task. Tasks were counterbalanced across participants, and individual scripts for each task were also counterbalanced to avoid any possibility of order effects. The tasks were done in a small room apart from the main classroom in individual sessions. The experimenter asked the child before the start of every task whether she would like to play the game or return to the classroom. As the all three tasks took around 15 minutes in total, only 9 children opted to return to the classroom and finish an individual task on another day.

We drew our protocol for DCCS directly from the original methods paper (Zelazo, 2006).

At the start of the game, the experimenter randomly chose a dimension (color or shape) that would guide the rules of the pre-switch phase of the task. The experimenter then told the child that they would be playing a game about the relevant dimension (e.g., the shape game), in which they would sort the cards by that particular dimension. Two training trials preceded the testing phase: in these training trials, the experimenter labeled the two target cards (the blue rabbit and the red boat), and then displayed a card to the participant, identified it by the relevant dimension, and then explained to the child where the card should be sorted and why. This process was repeated with another playing card before 6 pre-switch test trials. The 6 pre-switch trials each began with the experimenter holding up a card, labeling it by the target dimension, and asking the child where it should go (e.g., "Here is a boat. Where should it go?"). The child then placed the card in the sorting tray face-down. If the child placed the card face-up, the experimenter turned it over.

After the 6 pre-switch trials, the experimenter stopped the game briefly, and explained to the child that now they were playing a game about the other dimension (e.g., the color game), and that now they would be sorting the cards by the other dimension. The experimenter proceeded immediately into 6 post-switch trials. The first post-switch trial started with a reminder about the shifted rules of the game, then a label of the relevant dimension, and finally a question as to where it should go (e.g., "Remember, we're playing the color game now; if it's blue it goes here, if it's red it goes here. Here is a red one; where does it go?). In the remaining test trials, the experimenter only labeled the target dimension, and asked the child where the card should go. Again, if the child did not place the card face-down, the experimenter turned it over. The pre-switch dimension was chosen randomly at the start of the task, and cards were shuffled after the two training cards.

In running the Give-Quantifier task, we followed the methods of the original task fairly

closely, with the exception of a few minor modifications. At the start of the task, the experimenter arranged the piles of fruit (8 each of strawberries, bananas, and oranges), separated by fruit kind, around a small plastic circle. Before the testing trials, the experimenter asked the child to identify each kind of fruit, but did not use any of the target quantifiers in the study (*some, all, none* and *most*). After the child identified the tree different kinds of fruit, the experimenter started the 8 test trials. In each test trial, the experimenter made sure all fruits were taken out of the plastic circle and separated by kind before asked the child to place a certain amount of fruit (denoted by the particular quantifier used) on the red plate. The quantifiers used in this task were *some, none, all*, and *most*. The experimenter used the partitive construction and prosodically emphasized the quantifier equally across all trials (e.g., "Can you put *all* of the bananas into the circle?"). Quantifiers were presented in two different orders between participants, and fruit-quantifier pairings were quasi-randomized such that the same pairing was not repeated within a session. If the child request clarification, the experimenter repeated the prompt, and added that the child should put however many pieces of fruit she felt should go on the plate.

In coding the results of the Give-Quantifier task, we relied on the original coding scheme (Barner et al., 2009); *all* and *none* trials were coded as correct for 8 and 0 pieces of fruit given respectively, whereas *some* trials were coded as correct if the child gave between 2 and 7 pieces, and *most* trials were correct if the child gave between 5 and 7 pieces.

Results

We again replicated children's performance on our implicature task 4; we ran independent sample t-tests in each age-group, and did not find that they significantly differed. Once again, we found that performance increased with age, with the oldest children the only group performing significantly above chance in both implicature and *none* trials. As in Experiment 2, we again
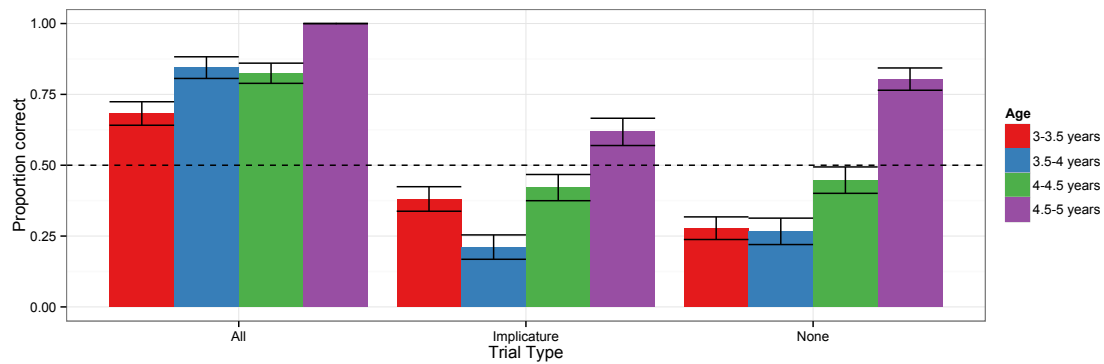
Figure 4. Proportion of correct responses by each age group across all scalar trial types. Error bars show standard error of the mean.

found a significant bimodal distribution of performance in both *some* ($D = .07$, $p <$¡ $.001$) and *none* trials ($D = .15$, $p < .0001$). After controlling for age, we again found that these performance on these trials were also significantly correlated ($r = .3$, $p < .01$).

We next turned to whether children's lower and correlated performance on implicature and *none* trials was a result of difficulty with inhibitory control. As seen in 5, performance on our inhibitory control measure, the Dimensional Change Card Sort task, was significantly correlated with age ($r = .28$, $p = .018$), with only 4–5-year-olds performing significantly above chance. After controlling for age, we did not find any significant correlation between inhibitory control and performance on either *some* trials ($r = .13$, $p = .26$) or *none* trials ($r = - .01$, $p = .93$) in our implicature task. Additionally, we were unable to predict performance on either of these trials with DCCS success when accounting for age.

Another hypothesis for the pattern of results we observed in our implicature task was that children tested had not yet established a consistent quantifier scale; we therefore included the
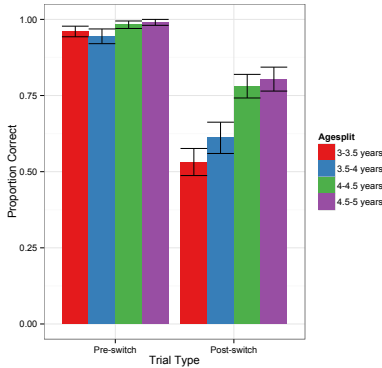
Figure 5. Proportion of correct responses by each age group in both pre- and post-switch trials. Error bars show standard error of the mean.

Give-Quantifier task in our experimental sessions. Children's responses on the Give-Quantifier task were coded as correct or incorrect for all analyses based on the coding scheme established by Barner and colleagues (Barner et al., 2009). Children's performance on this task was very similar to performance on our implicature task, with all age groups performing at ceiling for the prompt *all*, and only older children succeeding on with the prompts *most, none* and *some*. 7 shows the breakdown of how children responded to these scalar terms. Younger children showed a bimodal pattern of response to the prompt *none*, with the majority of children giving 0 or 8 items in these trials, gradually shifting to a correct response by 4.5–5 years. Similarly to performance with implicature trials in our storybook task, we found a large proportion of younger children giving all 8 objects in response to the prompt *some*, and the oldest age groups giving a more adult-like response. After controlling for age, we found that performance in both these trials were significantly correlated ($r = .61$, $p < .0001$).

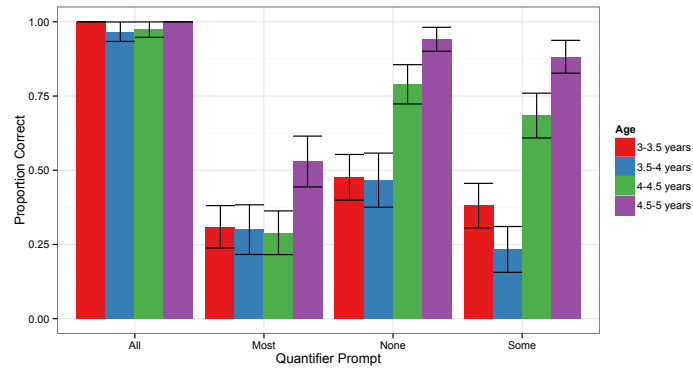In a further exploration of the relationship between quantifier knowledge and scalar

Figure 6. Proportion of correct responses by each age group for quantifier prompts *all, most, none* and *some*. Error bars show standard error of the mean.
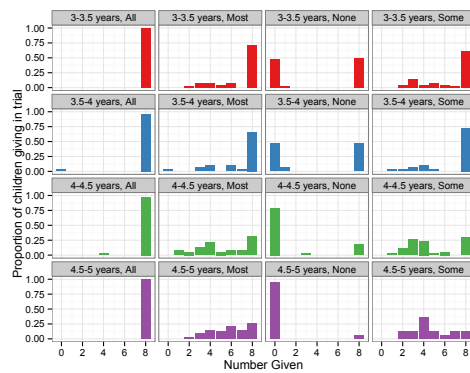

Figure 7. Proportion of children giving numbers of items faceted by age group and quantifier prompts *all, most, none* and *some*.
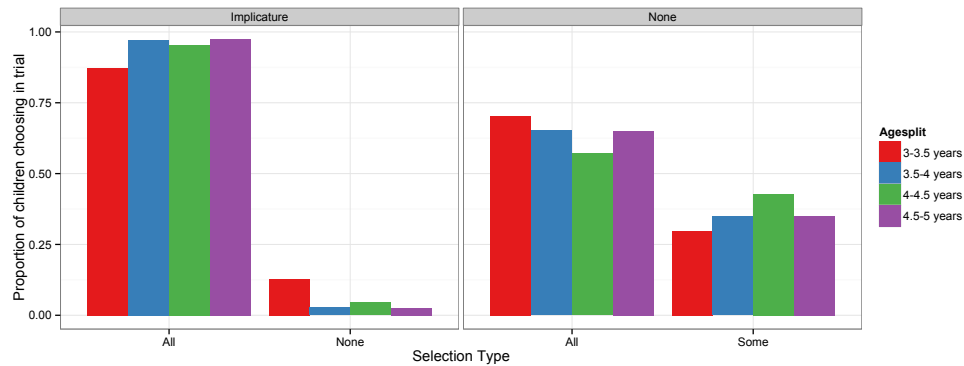
Figure 8. Proportion of children choosing an alternative on incorrect trials, faceted by trial type (*some, none*) and split by age group.

implicature performance, we examined both the particular kinds of errors that children made across both tasks, and how they were related. 8 shows the breakdown of children's performance in our Scalar Implicature task on incorrect trials. We found that children showed evidence of "listening through the quantifier", and making their selection based on the target noun rather than the scalar description. This result closely mirrors children's performance in our Give-Quantifier task (7), in which younger children responded in a binary fashion for the items *some* and *none*. A closer look at the individual differences on a very correlated item (*some*) across both tasks can be seen in 9. In a mixed-effects model, we found a main effect of age in performance across these two tasks, but only found that these successes across these tasks were significantly predictive of each other when age was removed from the model.

However, in partial-correlations controlling for age, we found that performance on our Give-Quantifier and implicature tasks were correlated: children who tended to struggle with scalar terms in the context of implicatures also had lower performance when asked to produce a a
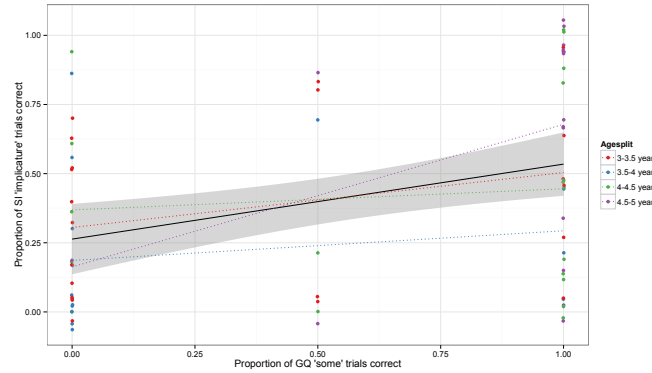
Figure 9. Proportion of children choosing an alternative on incorrect trials, faceted by trial type (*some, none*) and split by age group.

number of items in response to a quantifier prompt, specifically *some* and implicature trials (r = .27, p = .02) and *none* trials in both tasks r = .52, p <.0001). We did not find a significant correlation with performance on *some* implicature and *none* quantifier trials (r = .18, p = .14), although we did find a relation between performance on *none* implicature and *some* quantifier trials (r = .35, p = .002).

Discussion

In Experiment 3, we turned our attention to investigating potential factors behind children's difficulty with making scalar implicatures. We combined two tasks targeting specific hypothesized areas of difficulty (inhibitory control and lack of quantifier knowledge) with our implicature paradigm in a within-subjects design to explore the relationship amongst these abilities.

While we found that younger children did have difficulties in our inhibitory control task, we did not find a significant relation between performance in this task and our other tasks on the basis of DCCS. However, a measure of inhibitory control that we did not capture in our study

were participants' reactions times: it is possible that successful performance in both making implicatures and exerting inhibitory control might incur additional costs in processing, resulting in longer response latencies. Therefore, while we did not find evidence for the role of inhibitory control in the current study, it is possible that it may be captured by a reaction time measure.

We did find that children's performance on the Give-Quantifier and Implicature tasks were related, and children who failed with scalar items tended to do so across both tasks. This result indicates that processing quantifiers is difficult for children even removed from making implicatures. While it is clear that children struggle with quantifier comprehension, the source of this developmental difficulty is less clear. It is still possible that children's definitions of these scalar terms are not solidified in the pre-school years, and that successfully understanding *some* necessitates understanding both *none* and *all*. It is also possible that the pragmatic contexts in which these scalar items occur are particularly difficult to process, and may require excessive processing time and effort on behalf of the child.

Conclusion

We designed a simple task to test children's sensitivity to a variety of word choice cues in a single paradigm, allowing us to investigate patterns of pragmatic development both within- and between-subjects. We minimized task demands by asking participants to select the speaker's intended referent from among three visual alternatives. In Experiment 1, we replicated the finding that preschoolers can compute ad-hoc implicatures, though we found poor performance on scalar implicatures. In Experiment 2, preschoolers' comprehension of all scalar quantifier terms in the task increased with age, and removing the ad-hoc trials increased older children's performance on scalar implicature. In Experiment 3, we explored two possible sources of difficulty in implicature computation, inhibitory control and quantifier knowledge, and found that even when removed

from an implicature context, quantifier comprehension is difficult for children. Our findings suggest that 4-year-olds are able to compute scalar implicatures with support from contextual cues, but their performance is fragile and may be inhibited by an unestablished quantifier scale.

Our work contributes to the existing literature in a number of ways. First, it offers a novel paradigm that is less complicated than many other implicature tasks, leading us to feel more confident that our results reflect children's true sensitivity rather than inadvertent task demands. Each test set remained visible to children, and they were merely asked to select which picture they thought was the referent of the speaker's description. Second, the relatively high number of trials both helped strengthen our analytical power and also offered the possibility for children to identify lexical alternatives as the study progressed (although we did not find that their performance significantly changed over the course of either experiment, cf. Skordos & Papafragou, 2014). Third, we were able to not only compare performance across age groups, but also examine individual patterns of responses across the different trial types. This design helped us determine that preschoolers' performance on scalar implicature trials was bimodal and highly related to their performance on *none* trials, which we would have been unlikely to uncover in a purely between-subjects implicature design without controls. Finally, we were able to explicitly test two hypotheses about the sources of children's difficult with scalar implicatures, and compare their performance with a controlled design.

Our findings support the Alternatives Hypothesis (Barner & Bachrach, 2010; ?, ?). First, our ad-hoc trials in Experiment 1 show that preschoolers had no difficulty generally making inferences about contextual descriptions when they are obvious from the context. Performance on scalar trials also appeared to be related to the recognition of a broad set of lexical alternatives, due to both preschoolers' increasing ability to compute scalar implicatures with age (presumably a

proxy for familiarity with scalemates) and due to the difference in performance across Experiments 1 and 2. Overall, these patterns of results support the idea that children's ability to compute implicatures relates to their ability to reason about what other possible utterances a speaker could have used instead. The correlated responses for *none* and *some* trials in both Experiments 1 and 2 presented an interesting puzzle which we addressed in Experiment 3. *None* is not typically considered part of the same Horn scale as *some* and *all* (because "all" entails *some*, but "some" does not entail *none*—and in fact entails the opposite), but it is nonetheless a lexical contrast along the same quantifier scale. One possibility is that children's knowledge of the whole quantifier scale plays a role in scalar implicature, though knowledge of logically-false alternatives is not involved in the computations outlined by most theoretical accounts of (e.g. ?, ?). Another is that performance on *none* and *some* trials may be correlated because both scalar implicature and negation comprehension might require inhibiting another response—the positive alternative in the negative case, and the stronger alternative in the implicature case. In our exploration of these two alternatives, we found support for the possibility that quantifier knowledge, rather than inhibitory control, drives the pattern of performance that we observed in our studies. However, it is possible that a more nuanced measure of inhibitory control (response latencies) might yield more information about the particular processing and inhibitory demands that these implicatures entail. While our results suggest that children's failure in scalar implicature tasks may have a root in absent quantifier knowledge, it is still very likely that the sources of developmental difficulty with implicatures come from several areas.

One pattern in our data is more difficult to reconcile with the Alternatives Hypothesis: Children's performance did not change over the course of either experiment. We had expected that, if children's difficulties with scalar implicature were due to a lack of recognition of the

contrastive relationship between "some" and "all," that this relationship would be revealed by the two words' consistent use in contrasting references over the course of the many trials that each child completed (Skordos & Papafragou, 2014). The lack of trial order effects we observed could indicate that children in our task did not yet have strong enough comprehension of these terms for contrastive use to matter, or alternatively that our referent-selection task eliminated the problem of summoning the contrasting term to mind and instead foregrounded some other inferential challenge (perhaps that of inhibitory control).

In sum, our work suggests that children can draw implicatures based on some lexical choices—such as in the case of ad-hoc implicatures—but they still struggle with quantifier-based scalar implicatures until relatively late. Their computation of scalar implicatures increases in supportive contexts, but their inferences are fragile and depend on their knowledge of lexical items.

References

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*(1), 40-62.

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, *58*(2), 195-219.

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

Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667-696.

Horn, L. (1972). The semantics of logical operators in english. *Unpublished doctoral dissertation, University of California Los Angeles*.

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

Miller, K., Schmitt, C., Chang, H., & Munn, A. (2005). Young children understand some implicatures..

Nordmeyer, A., & Frank, M. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, *77*, 25-39.

Noveck, I. (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.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*, 71-82.

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

Stiller, A., Goodman, N., & Frank, M. (2014). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*(2), 176-190.

Zelazo, P. D. (2006). The dimensional change card sort (dccs): a method of assessing executive function in children. *Nature Protocols*, *1*, 297-301.