# Lexical alternatives improve 5-year-olds' ability to compute scalar implicatures

**Dimitrios Skordos and Anna Papafragou**
**University of Delaware**

## 1. Introduction

A *scalar implicature* (SI) is a pragmatic inference triggered in the presence of certain lexical items like quantifiers, where the use of a certain item (*some*) is taken to implicate that a logically stronger item (*all*) would be false. For example the statement in (1a) below is usually taken to implicate (1b).

(1) a. Megan ate some of the cupcakes.
b. Megan did not eat all of the cupcakes.

The term *scalar* comes from the idea that linguistic terms like *some* and *all* form ordered sets of alternatives based on informational strength, typically described as scales. In our case, the term *some* can be thought of as belonging to the following scale < *some, …, most, all*> (see Horn, 1972).

An account of how scalar implicatures are derived was described by Paul Grice (1975). He suggested that communication is a co-operative effort largely governed by rational expectations ('Maxims') about how a conversation should proceed. According to Grice's maxims, other things being equal, communicators are expected to offer contributions which are truthful, informative, relevant and appropriate to the goals of the conversation. These expectations about rational conversational conduct guide the inferences which hearers usually entertain when interpreting utterances. When these expectations seem to be violated, the assumption that this was done on purpose creates a variety of effects. According to Grice's theory, in producing (1a), the speaker violated the maxim of Quantity (or informativeness):

Quantity maxim
i. Make your contribution as informative as is required.
ii. Do not make your contribution more informative than is required.

In (1a), the speaker has violated the submaxim (i) by using a relatively weak term from a set ordered according to informational strength (< *some, …, all* >). The speaker is expected to say as much as he/she truthfully can, in a way relevant to the exchange. The choice of the weaker term is reason to believe that the speaker is not willing to commit to an informationally stronger statement

("Megan ate all of the cupcakes."). Therefore, as far as the speaker is willing to share, the stronger statement does not hold, thus (1b) (see also Horn, 1972). The quantifier *some* has lower-bounded semantics ('at least some and possibly all'). The upper-bounded meaning is a pragmatic enrichment derived from the implicature which can also be explicitly canceled without logical contradiction ("In fact, Megan ate all of the cupcakes"). SIs can also be derived from non-logical scales, based on contextual information (Hirschberg, 1985). For instance, the response in (2) implicates that the action was not completed.

(2) Q: Did you change the oil?
A: I opened the hood.

**2. Developmental Evidence**

How do children fare with SIs? Early studies designed to investigate children's knowledge of quantification and propositional connectives provide relevant evidence. Smith (1980) found that children up to the age of 9 usually treated *some* as 'some and possibly all', and Braine and Rumain (1981) reported that children seemed to prefer a logical inclusive interpretation of the disjunction *or* ("p or q and possibly both") rather than the pragmatic, exclusive one that the adults tended to prefer ("either p or q but not both").

Relatedly, Noveck (2001) showed that children of ages up to 9 would overwhelmingly treat the modal term *might* logically, while adults seemed to be ambivalent between the logical and pragmatic interpretations. In the same study, French speakers interpreted the French existential quantifier *certains* ( "some") in statements like *Some giraffes have long necks* as compatible with *tous* ("all") at ages up to 10, while adults were again equivocal between the logical and the pragmatic interpretations.

Are linguistically competent children simply incapable to engage in the computations required to derive pragmatic inferences linked to scalar terms like modals and quantifiers? As Noveck and others noted, it could be that the failure observed was due to task demands, since the tasks described above require a non-trivial amount of effort on the part of the participant, who has to evaluate the truth of an out-of-context statement against world knowledge. This hypothesis leaves open the possibility that children's ability to calculate scalar implicatures would improve under certain experimental circumstances.

A series of studies by Chierchia, Crain, Guasti, Gualmini and Meroni (2001) and Gualmini, Crain, Meroni, Chierchia and Guasti (2001) provide evidence in support of this hypothesis. The studies investigated preschoolers' interpretation of the disjunction operator *or*. They found that while adults were sensitive to the implicature of exclusivity from the use of disjunction in statements like *Every boy chose a skateboard or a bike* (i.e., they interpreted the statement as meaning "either a skateboard or a bike"), children, once again

seemed oblivious to the exclusive interpretation of disjunction. Crucially, in a follow-up task, when children were presented with two statements produced by two puppets and were asked to reward the puppet who "said it better", they overwhelmingly chose to reward the puppet who produced a stronger/more informative statement with *and* (*Every farmer cleaned a horse and a rabbit*) over a puppet who offered a weaker/less informative statement with *or* (*Every farmer cleaned a horse or a rabbit*) under conditions that made the stronger statement true.

In another study, Ozturk and Papafragou (under review) present very similar results with modal expressions (*may, have to*). In their first experiment, children and adults showed a clear preference for logical (weak) interpretations of the modal *may* in a reasoning task that involved guessing about the location of a hidden animal. Even when according to the available evidence, the cow absolutely *had* to be in the orange box, both groups of participants accepted the statement *The cow may be in the orange box*. However, in a second experiment, when given a choice between two statements under the same conditions, both adults and 5-year-olds preferred statements with *have to* (strong) over statements with *may* (weak).

Along with other studies that explored the effects of training and relevant context on children's generation of SIs (Papafragou & Musolino, 2003; Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni, 2005), these findings confirm the hypothesis that children have the ability to make pragmatic inferences. More specifically, these findings suggest that children's problem lies with generating scalar alternatives when faced with a weak scalar term (Gualmini et al., 2001; Papafragou & Skordos, to appear). As we saw above, when children are explicitly given a weak and a strong statement in environments that make the stronger statement true and are asked to choose the best statement, they tend to overwhelmingly favor the stronger alternative (Gualmini et al, 2001; Chierchia et al, 2001; Ozturk & Papafragou, under review). This indicates that children do not have a problem in comparing alternatives to the weak scalar term, or assessing the relative informativeness of alternatives, but accessing alternatives in the first place.

Evidence in support of this view comes from Barner, Brooks and Bale (2011). They tested 4-year-old children in a task that involved answering questions about the behavior of a group of three animals. In critical trials, all three animals (a dog, a cat and a cow) were sleeping and children were asked whether "…some/only some of the animals are sleeping". Children seemed quite happy to respond affirmatively (about 66% of the time) regardless of the form of the question. Children's affirmative response to the question with bare *some* is expected since questions do not give rise to SI generation. Children's failure to respond with *No* to the question with *only some*, however, was taken to indicate that children have difficulty with generating scalar alternatives even when this is predicted to be triggered by the grammar (*only* is a focus element requiring the

generation and negation of relevant alternatives). Interestingly, a different group of children performed much better when members of the set of animals were explicitly individuated within the same displays, thereby making the set of relevant alternatives more salient. Specifically, when asked whether "only the cat and the dog are sleeping", children correctly gave *No*-responses 86% of the time. When simply asked whether "the cat and the dog are sleeping", children accurately responded with an affirmative answer 93% of the time. Barner et al. (2011) interpreted these findings as strong evidence that children's problem with SIs lies mainly in realizing what terms can come together to form a scale. When scalemates are provided for them, children's generation of SIs improves significantly.

One question that remains open is how exactly children come to organize lexical alternatives in the form of scales and use such alternatives to compute SIs. Obviously providing the alternatives explicitly and asking participants to choose the best among them seems to make alternatives accessible since they are being compared and contrasted and the participant is given an option between the two (Chierchia et al., 2001; Gualmini et al., 2001; Ozturk & Papafragou, under review). Individuation of set members also seems to make alternatives accessible, since it highlights possible set/subset relations (Barner et al., 2011).

In our study we test whether implicit lexical contrast between strong (*all*) and weak (*some*) scalar alternatives, without contextual assistance or training, can improve 5-year-olds' SI generation. By 'implicit contrast', we mean that more and less informative scalar terms (*some, all*) will not be explicitly compared in the form of two statements that the participant has to evaluate in order to choose the more informative one. Rather, the scalar terms will become available to the participant in the form of statements that need to be evaluated independently of each other, based on visual evidence. Accessibility to the alternatives (and thus implicit contrast) is going to be manipulated through the order of *some-all* statements. Of interest is whether the mere mention of a stronger scalar alternative (*all*) in the course of the experiment can encourage the participant to generate a SI from the use of a weak alternative (*some*).

## 3. Experiment

### 3.1 Method

#### 3.1.1 Participants

We tested 61 typically developing 5-year-old children (4;10 – 5;11, M=5;3) and 36 adult controls, all monolingual speakers of English. The children were recruited from daycare centers in Newark, DE and the English speaking adults were college students recruited from the University of Delaware, who received course credit for their participation. An additional group of 3 children were tested but excluded from the analysis for justifying their *No* answers (rejection

of the puppet's statement) by referring to the items featured in the trials and not the quantifiers[1].

### 3.1.2 Materials

The stimuli consisted of 16 MS PowerPoint slides, each depicting 4 novel creatures, the "blickets". We used a unique set of novel creatures in order to restrict the universe of discourse to the visual content provided for each trial. Blickets had several everyday items (crayons, flashlights, paintbrushes, etc.). In half of the slides 4 out of 4 blickets would have an item each (*full set* scenes) and in the other half 3 out of 4 blickets would have an item each (*subset* scenes). Each slide was paired with a statement containing a quantifier (*some* or *all*). Scene type (full set vs. subset) was crossed with quantifier type (*some* vs. *all*) to provide 4 types of trials. In *True All*-trials, 4 out of 4 blickets had the item and participants heard: "All of the blickets have an X.". In *False All*-trials3 out of 4 blickets had the item and participants heard: "All of the blickets have an X." In *Felicitous Some*-trials, 3 out of4 blickets had the item and participants heard: "Some of the blickets have an X.". Finally, in *Infelicitous Some*-trials, 4 out of blickets had the item and participants heard: "Some of the blickets have an X." The first three types of trials tested participants' semantic judgments about *some* and *all*. The last type of trial (where *some* is used despite it being evident that all 4 blickets have the item) tested participants' pragmatic judgment (i.e., their ability to generate SIs). Examples of the visual stimuli and statements for each trial type (with the same item for ease of comparison) can be found in Figure 1.

These 4 types of trials were repeated 4 times each with different items so that no subject saw the same scene paired with more than one statement, for a total of 16 test trials. Pairings of scenes with statements were rotated to create 4 different batteries so that each scene was paired with a different statement type in each battery.

[1]For example these children would reject a statement like: "All of the blickets have a flashlight" and would justify their answer by saying: "these are not blickets", or "they don't have flashlights, they have lightsabers", instead of focusing on whether the quantifier (*all*) was used appropriately to describe the scene.

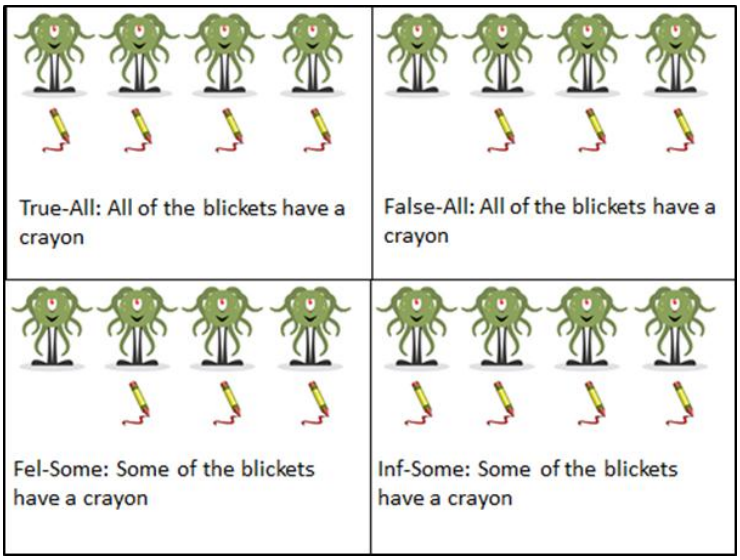


**Figure 1. Types of trials (scene types and corresponding statements).**

### 3.1.3 Procedure

The task was an Acceptability Judgment Task similar to that in Papafragou and Musolino (2003). Participants sat in front of a laptop PC computer and were shown the slides depicting the experimental stimuli. The experimenter introduced the task to the children by explaining that they would see some silly pictures on the computer together. A puppet, Max the silly gorilla, was introduced and it was explained that he says silly things sometimes. Participants were told that the puppet would describe the pictures and that they would have to evaluate the puppet's statement by answering whether the puppet "said it well or not". They would also have to justify their answer in case they rejected the puppet's statement. An assistant animated the puppet and provided the appropriate statements, while the experimenter wrote children's answers down in an answer sheet. Adults were tested in a very similar way with the only differences being that (a) they had to write down their own responses, and (b) they were tested in groups without the presence of a puppet (they were shown a cartoon character, Max the silly gorilla, that supposedly provided the statements the experimenter read).

Participants first went through 4 pre-test trials. These consisted of slides depicting cartoon animals or objects (e.g., a cow, an ice cream cone). Two of the pre-test trials were erroneously described by the puppet and two of them were correctly described, so that participants would have evidence that the puppet was

capable of providing both 'silly' and accurate statements. For pre-test trials, participants were also provided with feedback when they failed to reject a false statement. For example, if participants agreed with the puppet when it described the cow as an "elephant", the experimenter would explain that the puppet "didn't say it well', and that in fact the picture depicted a cow.

After the pre-test trials were concluded, participants were introduced to a cartoon character, Ben the Wizard. Ben was shown to use his magic wand to create the 4 blickets and participants were informed that these are the only blickets "in the whole world". The 16 test trials followed. The order of presentation of the test trials was manipulated across three between-subjects conditions. In the *Mixed* condition, *some*- and *all*- statements were intermixed in a pseudorandom order so that lexical contrast between the stronger (*all*) and weaker (*some*) scalar terms could be established. The pseudorandom order ensured alternations of *some* and *all* at least every three trials. In the *Some-First* condition, *some*- and *all*- statements were presented in blocks, with the *some*-block always first so that lexical contrast to the stronger alternative (*all*) was eliminated (within each block, the order of statements was pseudorandom). Finally, in the *Infelicitous Some-First* condition, the *some*-block of the previous condition was further split into two blocks, with *Infelicitous some*-trials always first and *Felicitous some*-trials always last, so that even the contrast between the two uses of *some* was eliminated.

Adult participants were evenly distributed across the 3 conditions. For children, n=21 were assigned to the *Mixed* condition, n=20 were assigned to the *Some-First* condition and n=20 were assigned to the *Infelicitous Some-First* condition.

### 3.2 Predictions

If lexical contrast between stronger and weaker scalar terms can make alternatives more accessible and therefore lead to scalar inference, children's rejection of *Infelicitous Some*-statements will improve when a contrast between strong (*all*) and weak (*some*) scalar alternatives is available. Therefore children's performance with these pragmatic trials should be better in the *Mixed* than in the *Some-First* or *Infelicitous Some-First* conditions.

It is not clear whether any differences in the performance of children in SI generation should be expected between the *Some-First* and *Infelicitous Some-First* conditions: if SI generation is facilitated only through contrast to the strong scalar alternative (*all*), then no difference between the *Some-First* and *Infelicitous Some-First* conditions is expected; if, however, contrast between the felicitous and infelicitous uses of *some* also facilitates SI generation (by promoting consideration of uses of *some* in different visual contexts), then those conditions are expected to differ, with children performing better in the *Some-First* than in the *Infelicitous Some-First* condition. No difference in children's

performance between conditions is predicted for the semantic trials (*True-All, False-All, Felicitous-Some*). Finally, no difference in adult performance is expected between conditions for either the semantic or the pragmatic trials.

### 3.3 Coding

*Yes* answers were coded as correct in the case of felicitous or true statements. *No* answers were coded as correct in the case of false or infelicitous statements. A mean of correct answers from 0 to 1 was calculated for each participant for each of the 4 trial types (*True-All, False-All, Felicitous-Some, Infelicitous-Some*). Those scores were used to categorize participants according to their performance on each trial type as either Passers (if they had a score of .75 or greater), or Failers (if they had achieved a score of .50 or less). For example, someone who had at least .75 correct on *Felicitous-Some* trials was categorized as a Passer for that trial type.

### 3.4 Results

Adult performance was practically at ceiling for all conditions and trial types. Table 1 below summarizes adult performance. Fisher's Exact test analyses on 2x3 contingency tables for each trial type revealed no significant difference in the numbers of Passers vs. Failers across conditions (*True All*-trials, p=1; *False All*-trials, p=1; *Felicitous Some*-trials, p=1; *Infelicitous Some*-trials, p=0.31).

| | | **Mixed** | **Some-First** | **Inf-Some-First** |
|---|---|---|---|---|
| **T-All** | Passers | 12 | 12 | 12 |
| | Failers | 0 | 0 | 0 |
| **F-All** | Passers | 12 | 12 | 12 |
| | Failers | 0 | 0 | 0 |
| **Fel-Some** | Passers | 12 | 12 | 12 |
| | Failers | 0 | 0 | 0 |
| **Inf-Some** | Passers | **12** | **12** | **10** |
| | Failers | **0** | **0** | **2** |

**Table 1. Adult performance.**

Table 2 below summarizes child performance. Children overall appeared to have no major problems with the 3 semantic trial types. Fisher's Exact Tests on 2x3 contingency tables did not reveal significant differences in the numbers of Passers vs. Failers across the 3 conditions for either the *True All*-trials (p=0.541), *False All*-trials (p=0.92), or *Felicitous Some*-trials (p=0.44).

Turning to the critical *Infelicitous-Some*-trials, children appeared to be pragmatic in the *Mixed* condition, and logical (non-pragmatic) in the *Infelicitous Some-First* condition (children seemed to be divided between a pragmatic and a logical interpretation of the *Infelicitous Some*-statements in the *Some-First* condition). A Fisher's Exact test on a 2x3 contingency table revealed a highly significant difference (p=0.0002) between the numbers of Passers and Failers for *Infelicitous-Some* trials across the 3 conditions. This effect was further explored by running Fisher's Exact Test on 2x2 contingency tables comparing each condition to the others. We found that the *Mixed* condition had significantly more Passers than either the *Some-First* condition (p=0.0148) or the *Infelicitous Some-First* condition (p=0.00001). We also found that the *Some-First* condition had significantly more Passers than the *Infelicitous Some-First* condition (p=0.0484).

| | | **Mixed** | **Some-First** | **Inf-Some-First** |
|---|---|---|---|---|
| **T-All** | Passers | 21 | 19 | 19 |
| | Failers | 0 | 1 | 1 |
| **F-All** | Passers | 16 | 16 | 17 |
| | Failers | 5 | 4 | 3 |
| **Fel-Some** | Passers | 14 | 17 | 14 |
| | Failers | 7 | 3 | 6 |
| **Inf-Some** | Passers | **19** | **11** | **4** |
| | Failers | **2** | **9** | **16** |

**Table 2. Child performance.**

We can be quite confident that children rejected the *Infelicitous-Some* statements for the correct reason, namely because they generated the appropriate SI. When asked to justify their rejections, children overwhelmingly referenced either the stronger scalar term ("*not all* of them have an X"), or used the focus element *only* ("*only some* of them have an X"). This shows that children who appear to be generating SIs in our task truly demonstrate an ability to compute SIs.

These results support our hypothesis that lexical contrast facilitates scalar implicature generation by children as evidenced by the significant differences between the performance of children in the *Infelicitous-Some* trials between the *Mixed* and the *Some-First* conditions. It also appears that there might be a facilitating effect of a 'usage' contrast between felicitous and infelicitous uses of *some*, as evidenced by the difference in *Infelicitous-Some* trials between the *Some-First* and *Infelicitous Some-First* conditions. However, it is worth pointing out that some of the children performed poorly in the *False-All* and *Felicitous-Some* trials, failing to reject the former and failing to accept the latter. This

raises doubts as to whether these children have fully acquired the semantics of the quantifiers. If this is the case, it is not clear that one can look at children's performance and derive conclusions about their pragmatic competence with quantifiers.

To address this, we conducted a second analysis excluding children who had under .75 correct in either the *True All*, *False All*, or *Felicitous Some* statements. This resulted in n=7 children being excluded in the *Mixed* condition, n=7 in the *Some-First* and n=9 in the *Infelicitous Some-First* condition. This new analysis included 14 children in the *Mixed* Condition, 13 children in the *Some-First* condition, and 11 children in the *Infelicitous-Some First* condition (see Table 3). All of these children can safely be assumed to have the correct semantics for *some* and *all.* A Fisher's Exact test on the 2x3 contingency table in Table 3 revealed a highly significant difference between the numbers of Passers vs. Failers for the 3 different conditions (p=0.001), confirming the results of the first analysis. This effect was further explored by running Fisher's Exact Test on 2x2 contingency tables comparing each condition to the others. Comparing the *Mixed* and the *Some-First* condition we once again found a significant difference (p=0.0407), with the *Mixed* condition having significantly more Passers than the *Some-First* condition. Comparing the *Mixed* and the *Infelicitous Some-First* condition we found again a significant difference (p=0.0007), with the *Mixed* condition having significantly more Passers than the *Infelicitous Some-First* condition. However, comparing the *Some-First* condition to the *Infelicitous Some-First* condition we did not find any significant difference (p=0.217).

| | | **Mixed** | **Some-First** | **Inf-Some-First** |
|---|---|---|---|---|
| **Inf-Some** | Passers | **14** | **9** | **4** |
| | Failers | **0** | **4** | **7** |

**Table 3. *Some/all*-knowers' performance in *Infelicitous-Some* trials.**

Even when we look only at children that seem to have a solid grasp of the semantics of the quantifiers ('*some/all* knowers', as evidenced by their performance in the semantic trials), the results strongly support our hypothesis that lexical contrast facilitates scalar implicature generation in children. The possible effects of a 'usage' contrast seem to disappear, although it is not clear at this point whether this might be a result of losing statistical power due to excluding children from the analysis (we address this point in ongoing work).

## 4. Discussion

Our study was motivated by the hypothesis that children's problem in generating SIs lies primarily with generating scalar alternatives when faced with

a weak scalar term (Gualmini et al., 2001; Papafragou & Skordos, to appear). We hypothesized that implicit lexical contrast between weak and strong scalar terms makes alternatives more accessible for children and thus facilitates SI generation. This hypothesis is supported in our data, at least as far as quantifiers are involved. In the *Mixed* condition, where *some*- and *all*- statements were intermixed so that contrast between the weak and strong scalar terms could be established, children were very successful at generating the appropriate scalar inference by rejecting the infelicitous statements. Crucially, children's performance fell significantly when lexical contrast to *all* was eliminated (in the *Some-First* and *Infelicitous Some-First* conditions).

An important question that remains is whether lexical contrast and the accessibility of alternatives in general is in any way linked to questions of informativeness or relevance (Sperber & Wilson, 1986). If that is the case, the assumption that communicators need to be informative or relevant might be driving the effects of contrast ("the speaker said X, where he/she might have also said Y, therefore I should infer that for some reason the speaker did *not* want to commit to Y"). Alternatively, informativeness or relevance might not be involved and contrast might function more or less as a simple, bottom-up mechanism that highlights semantically related alternatives to scalar terms. More work is necessary to address these points and clarify the role of contrast in pragmatic development.

**Acknowledgements**

We would like to thank the Early Learning Center at the University of Delaware and all the other preschools for their cooperation, Ann Bunger, Megan Johanson, Alyssa Kuchta, Adrienne Pinto, and Leah Diamant for giving life to Max the Silly Gorilla and the members of the Language & Cognition Lab at the University of Delaware for their comments and suggestions.

**References**

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. Cognition, 188, 87-96.

Braine, M., & Rumain, B. (1981). Children's comprehension of "or": evidence for a sequence of competencies. Journal of Experimental Child Psychology, 31, 46-70.

Chierchia, G., Crain, S., Guasti, M., T., Gualmini, A., & Meroni L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. Proceedings from the Annual Boston University Conference on Language Development, 25, 157-168.

Grice, H., P. (1975). Logic and conversation. In P. Cole, and J., L. Morgan (eds.) Syntax and semantics: Speech acts, Volume 3, 41–58. New York: Academic Press.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M., T. (2001). At the semantics/pragmatics interface in child language. Proceedings of Semantics and Linguistic theory XI. Ithaca, NY: CLC Publications, Department of Linguistics, Cornell University.

Guasti, M.,T., 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, 667-696.

Hirschberg, J. (1985). A theory of scalar implicature. Doctoral diss., University of Pennsylvania.

Horn, L., R. (1972). On the semantic properties of the logical operators in English. Doctoral diss., UCLA.

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicatures. Cognition, 78, 165-188.

Ozturk, O., & Papafragou , A. (under review). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. Cognition, 86, 253-282.

Papafragou, A. & Skordos, D. (to appear). Implicature. In J. Lidz, W. Snyder and J. Pater (eds.), The Oxford Handbook of Developmental Linguistics. Oxford: Oxford University Press.

Sperber, D., & Wilson, D., (1986/1995). *Relevance: Communication and Cognition*, 2nd ed. 1995. Cambridge, MA: Harvard University Press.