SCALAR IMPLICATURES IN CHILDREN:

ALTERNATIVES AND RELEVANCE

by

Dimitrios Skordos

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

Spring 2014

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

SCALAR IMPLICATURES IN CHILDREN:

ALTERNATIVES AND RELEVANCE

by

Dimitrios Skordos

Approved: _____
Benjamin Bruening, Ph.D.
Chairperson of the Department of Linguistics & Cognitive Science

Approved: _____
George Watson, Ph.D.
Dean of the College of Arts & Sciences

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Anna Papafragou, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Gabriella Hermon, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Satoshi Tomioka, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Paul Quinn, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to thank my advisor Anna Papafragou for the training, advice and support she has been offering over several years. It goes without saying that this work would not have been possible without her. I would also like to thank the Language and Cognition lab members who have helped and supported me in my graduate career: Ann Bunger, Megan Johanson, Joshua Viau, Francis Wilson and Ercenur Unal.

I would like to extend my gratitude to the members of my dissertation committee, for their valuable feedback, general advice and support.

I would like to thank the Office of Professional and Graduate Education at the University of Delaware for their support in the form of a Dissertation Fellowship that allowed me to work focused solely on my doctoral research, and the College of Arts and Sciences at the University of Delaware for a Dean's Doctoral Summer Scholar award.

Many thanks to all the preschools and parents who have helped me with participant enrollment for my studies, and particularly the Early Learning Center and Bianca Graves at the University of Delaware. A big "thank you" to the children who were interested in playing a game with "Max the silly gorilla". Seeing sparks fly in those bright eyes when they feel they have figured out "the game", has never failed to put a smile to my face to this day.

Last but not least, I would like to thank my friends and family for their love and support. Angeliki Athanasopoulou deserves a very special mention in that respect

as she had (and continues to have) a special and rather demanding role as a significant other, friend, therapist, colleague and editor.

TABLE OF CONTENTS

LIST OF TABLES viii
LIST OF FIGURES ix
ABSTRACT x

Chapter

1 INTRODUCTION 1

1.1 Meaning in Natural Language 1
1.2 Scalar Implicature 7
1.3 How Children Calculate Sis 11
1.4 Open Questions 20
1.5 The Present Studies 24

2 EXPERIMENT 1 27

2.1 Method 28

2.1.1 Participants 28
2.1.2 Materials and Procedure 29

2.2 Predictions 32
2.3 Coding 33
2.4 Results 33
2.5 Discussion 37

3 EXPERIMENT 2 40

3.1 Method 41

3.1.1 Participants 41
3.1.2 Materials and Procedure 41

3.2 Predictions 43
3.3 Coding 44
3.4 Results 44
3.5 Discussion 48

4 EXPERIMENT 3 50

4.1 Method 50

4.1.1 Participants 50

4.1.2 Materials and Procedure 51

4.2 Predictions 52
4.3 Coding 53
4.4 Results 54
4.5 Discussion 57

5 GENERAL DISCUSSION 58

5.1 Goals and Summary of Findings 58
5.2 Relevance, Conversational Goals and the QUD 64
5.3 Alternatives in Pragmatics and Lexical Learning 67
5.4 SIs and Pragmatic Development 69

REFERENCES 72

Appendix

UNIVERSITY OF DELAWARE IRB APPROVAL FOR RESEARCH ON HUMAN SUBJECTS 83

LIST OF TABLES

Table 1. Adult performance in Experiment 1. The numbers represent Passers vs. Failers in corresponding trials. 34

Table 2. Child performance in Experiment 1. The numbers represent Passers vs. Failers in corresponding trials. 34

Table 3. *Some/all*-knowers' performance on *True-and-Infelicitous-Some*-trials of Experiment 1. 36

Table 4. Adult performance in Experiment 2. 45

Table 5. Child performance in Experiment 2. 46

Table 6. *Some/all*-knowers' performance in *True-and-Infelicitous-Some*-trials of Experiment 2. 47

Table 7. Adult performance in Experiment 3. 54

Table 8. Child performance in Experiment 3. 55

Table 9. *Some/none*-knowers' performance in *True-and-Infelicitous-Some*-trials of Experiment 3. 56

LIST OF FIGURES

Figure 1. Types of trials for Experiment 1. 31

Figure 2. Types of trials for Experiment 2. *False-All*-trials differ between the *Quantity* and *Object* condition. 42

Figure 3. Types of trials for Experiment 3. 52

ABSTRACT

Utterances like "Megan ate some of the cupcakes" are typically interpreted as "Megan ate some but not all of the cupcakes". Such an interpretation is thought to arise from a pragmatic inference called Scalar Implicature (SI). Preschoolers typically fail to spontaneously generate SIs without the assistance of training or context that makes the stronger alternative salient. However, the exact role of alternatives in generating SIs remains contested. Furthermore, it is not clear whether children take into account the goal of the conversation in considering scalar alternatives and generating SIs. We present three studies with English-speaking preschoolers and adults designed to address these questions. We show that the presence of stronger alternative is important for SI generation (Experiment 1), that the stronger alternative leads to SI generation only when it is relevant to the goal of conversation (Experiment 2) and that relevance does not simply mediate, but in fact drives accessibility of the alternatives (Experiment 3). We discuss the implications of these findings for pragmatic inference and the study of pragmatic development in general.

Chapter 1

INTRODUCTION

1.1 Meaning in Natural Language

A central goal for linguistic research is to arrive to a theory of language with explanatory adequacy. Such a theory would be a model of linguistic behavior that accounts for the structure and variability of language and in doing so has predictive power (Chomsky, 1964). According to current linguistic theory, a distinction needs to be made in natural language between what is explicitly included in a sentence (what is encoded) and what is meant to be communicated by a speaker uttering the same sentence in discourse.

(1) a. "Are you joining us for drinks tonight?"

b. "I have to submit a grant application by Monday…"

In the example above, the speaker by (1b) means to communicate that she cannot join the others for drinks because she needs to work, even though this is not explicitly mentioned in the utterance. What is encoded in the linguistic message belongs to the field of Semantics, which studies natural language sentences and their truth conditions. What is actually intended to be communicated by the speaker (even though it is often left unsaid) belongs to the field of Pragmatics. Pragmatics studies how sentences are used in actual utterances to communicate information within a contextual frame that the communicative partners share (see Stalnaker, 1972; Kempson, 1988; among others for an overview).

The distinction between semantic and pragmatic contributions to meaning is particularly useful for linguistic theory. It allows for abstract compositional semantics, and thus a simpler grammar, while at the same time it captures the crucial generalization that the way people actually use language to communicate affects the interpretation of the linguistic message (Papafragou, 2006). This distinction somewhat parallels the sharp distinction in generative linguistics between competence (the linguistic system per se) and performance (the way the system is used in communication) made by Chomsky (1965). Perhaps for that same reason generative linguists were generally content to study the structure and meaning of words and sentences (syntax and semantics) in isolation, without taking into account pragmatic content, a practice that has generally yielded valuable insights into natural language. However, there is an argument to be made about possible limitations that such approaches place on our understanding, since they isolate their object of study (natural language) from the environment within which it is placed (human communication). Evidently humans are capable of pragmatic inferences, otherwise (1b) above would make no sense as a response: semantic content is often coupled with pragmatic enrichment; words and utterances in real life are always interpreted within context and never in isolation.

The above assumption is even more important for the study of how children acquire language. As the above arguments show, it would be unwise to study semantic development while ignoring pragmatic inference. On the contrary, the joint study of semantic and pragmatic development can provide crucial insights into how meaning as a whole is understood and acquired in natural language. Nevertheless, the picture from the early development of pragmatics in children is somewhat mixed.

To begin with, there is a wealth of empirical evidence that shows the presence of abilities necessary for pragmatic reasoning in prelinguistic infants as young as 10 months of age. Joint attention has been reliably observed for 10 month-olds (Corkum & Moore, 1998), and by 18 months of age children show evidence of employing joint attention systematically in language acquisition and language use (Tomasello, 1995). In addition, children show some evidence of understanding the intentions of others from pretty early on: One and 2-year-olds seem to comprehend the communicative intentions of their confederates in non-linguistic games (Behne, Carpenter & Tomasello, 2005; see Tomasello, Carpenter, Call, Behne & Moll, 2005 for discussion). Even more strikingly, 12-month–old infants seem to adjust their communicative gestures to be appropriate to the level of knowledge of others in an effort to help them find items that they sought (Lizskowski, Carpenter & Tomasello, 2008). Finally, at around the same age children show some evidence of precursors to powerful metarepresentational abilities, by using gestures to point to absent entities (Liskowski, Shäfer, Carpenter & Tomasello, 2009).

There is also evidence that pragmatic inference is involved from very early on in word learning (see Bloom, 2000 for a review and Bale & Barner, 2013 for related discussion). For example, when 2-year-old children are presented with a novel label (*dax*) together with an item with a known label (e.g. a toy car) and a novel unlabeled item, they are able to use the contrast between known (*car*) and novel (*dax*) labels to infer that a speaker using the novel label must refer to the previously unlabeled item (Clark, 1990, Diesendruck & Markson, 2001). Other studies show that pragmatic inference might be involved when children learn counting and the semantics for number words (Barner & Bachrach, 2010; Brooks, Audet & Barner, 2013, but also

Hurewitz, Papafragou, Gleitman, & Gelman, 2006; for a different view). Two-year-olds have also recently been shown to use disfluencies in speech to infer a speaker's intended referent (Kidd, White & Aslin, 2009) and 3-year-olds have been known to use pragmatic inference in referent disambiguation (Stiller, Goodman & Frank, 2011).

Nevertheless, other tests have suggested that the pragmatic abilities of children up until the age of 5 might be rather limited. Classic Theory-of-Mind tasks (Wimmer & Perner, 1983) designed to test children's ability to monitor the mental states of others (generally thought to be involved in pragmatic inference) found that it was only after the age of 6 that children could consistently attribute false beliefs to others, while at around 5 years of age a majority of children could pass the "appearance-reality" or "smarties" task (Gopnik & Astington, 1988, see Wellman, Cross & Watson, 2001 for an in depth review). In addition, other early attempts to estimate the pragmatic sophistication of young children resulted in generally pessimistic conclusions (see discussion in Papafragou, 2006). For example children have been described as unable to reason about and attempt to influence the mental states of others, and therefore as unable to have an adult-like understanding of the communicative process (Shatz, 1983; Shatz & O'Reilly, 1990; but see also Grosse, Behne, Carpenter & Tomasello, 2010 for discussion). Other studies have demonstrated that children as old as 5 are unable to use pragmatic cues available through discourse to resolve syntactic ambiguity and recover from garden path effects (Trueswell, Sekerina, Hill & Logrip, 1999). In other studies that investigated uses of language typically attributed to pragmatic competence, preschoolers' abilities to comprehend metaphor and irony appear rather limited (Markman & Seibert, 1976; Vosniadou, 1986) and epistemic

modality seems to have a rather complex pattern of acquisition that is not completed until after the preschool years (see Papafragou, 1998 for discussion).

This apparent paradox can be resolved if we think about extrinsic factors that might affect the performance of children in the linguistic and classic theory of mind tasks. For instance, it has been demonstrated that those tasks impose heavy loads on other cognitive mechanisms, particularly inhibitory control (Bloom & German, 2001, Carlson & Moses, 2000; Carlson, Moses & Claxton, 2004). This was supported by more recent findings that have cast doubt on the previous evidence regarding the lack of any appreciable amount of pragmatic competence in children until the age of 5 or 6: When tested under circumstances that imposed a lesser cognitive load, children as young as 3 have been shown to be quite successful at theory of mind tasks (Carpenter, Call & Tomasello, 2002) and when tested with non-verbal paradigms the age barrier has been reduced to as young as 15 months of age (Onishi & Baillargeon, 2005; Surian, Caldi & Sperber, 2007). Despite some criticism regarding whether those tasks genuinely show attribution of goals and false beliefs to others, or higher order cognitive and social abilities (Haith, 1998; Perner & Ruffman, 2005; Sirois & Jackson, 2007, but see also Aslin, 2000 for a response), newer studies designed to address those criticisms provide evidence of false belief attribution by at least 18 months of age if not sooner (Buttelmann, Carpenter & Tomasello, 2009). To sum up, children appear to have significant metarepresentational competence in place from very early on; however, their ability to show their competence suffers when they are asked to do so through tasks that impose additional cognitive load such as explicit linguistic tasks.

This reinterpretation leaves open the possibility that preschool children have problems with the pragmatics governing language use (linguistic pragmatics), such as

those governing metaphor, irony, syntactic disambiguation and so on (Markman & Seibert, 1976; Vosniadou 1986; Trueswell et al., 1999). The goal of the present thesis is to gain insights into pragmatic development by studying a central area of linguistic pragmatics, namely conversational inference (Grice, 1975; see also the example in (1)).We focus on a sub-type of such inferences called 'scalar implicature'. Scalar implicature is generally considered a paradigm case of pragmatic inference (Papafragou & Musolino, 2003). This is supported by the following facts: (a) In utterances with scalar implicatures there are clear contributions of semantic and pragmatic content (see next section). (b) Importantly, adult hearers seem to take into account the speaker's knowledge and mental state in deriving a SI (Bergen & Grodner, 2012), which links SIs nicely with pragmatic precursor abilities we reviewed previously (theory of mind, knowledge and intentions of others, etc.).

In what follows, we will present a series of developmental studies on scalar implicature aimed to show that children's well attested problems in deriving these inferences should ultimately be attributed to limitations on their ability to quickly decipher the conversational goal of a communicative exchange and consequently the appropriate inferences to be made. We will argue that this conclusion has significant explanatory power for the development of pragmatic inference in general and that it is exactly this ability that develops over time and transforms children into fully competent adult communicators.

In the remainder of Chapter 1, we will provide some theoretical background on scalar implicature (section 1.2). We will also review the relevant developmental literature and identify open questions (sections 1.3, 1.4 and 1.5). In the main body of the thesis, we will present our series of studies (Chapters 2, 3 and 4) designed to

address those open issues. We will close this thesis by discussing the implications of our results for current theories of pragmatic development (Chapter 5).

1.2 Scalar Implicature

A *scalar implicature* (SI) is a pragmatic inference triggered in the presence of certain lexical items like quantifiers. Use of a proposition containing a certain item (*some*) is taken to implicate that another one containing a logically stronger item (*all*) would not hold. For example the statement in (2a) below can be used to implicate (2b).

(2) a. Megan ate some of the cupcakes.

b. Megan did not eat all of the cupcakes.

The term scalar comes from the fact that linguistic terms like *some* and *all* form an ordered set of alternatives (*scale*) based on informational strength. (< *some, ..., most, all*>; Horn, 1972). Informational strength is based on asymmetrical logical entailment where the informationally stronger term (*all*) logically entails the weaker one (*some*) but not vice versa.

On this account, the quantifier some has lower-bounded semantics ('at least some and possibly all'; Horn, 1972). The upper-bounded meaning ('some but not all') corresponds to the scalar implicature and is therefore a pragmatic enrichment of the semantic content of the quantifier. The conclusion that the upper-bounded meaning is a pragmatic, not a semantic, contribution is further supported by the fact that this meaning can be explicitly canceled without logical contradiction ("Megan ate some of the cupcakes. In fact, she ate all of them"). Other logical scales include ones based on disjunction <*or, and*> or modals <*might, must*>. For instance, the statements in (3a) and (4a) below are taken to implicate (3b) and (4b) respectively.

(3) a. Megan ate a cupcake or a cookie.

b. Megan did not eat a cupcake and a cookie.

(4) a. Bert might be in his lab.

b. It is not the case that Bert must be in his lab.

Scalar implicatures can also be derived from non-logical scales, based on contextual information (Hirschberg, 1985). For instance, the response in (5b) implicates that the action of changing the oil was not completed.

(5) a. Did you change the oil?

b. I opened the hood.

The first 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 about how a conversation should proceed. These expectations were formalized in his now famous Maxims:

Maxim of Quantity:

i. Make your contribution as informative as is required.

ii. Do not make your contribution more informative than is required.

Maxim of Quality:

i. Do not say what you believe to be false.

ii. Do not say that for which you lack adequate evidence.

Maxim of Relation (or Relevance):

i. Be relevant.

Maxim of Manner:

i. Avoid obscurity of expression.

ii. Avoid ambiguity.

iii. Be brief.

iv. Be orderly.

The above maxims are thought to 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 summarized in the steps (I.) to (IV.) below (see also Horn, 1972):

I. In (2a), the speaker has violated the Quantity maxim by using a relatively weak term from a set ordered according to informational strength (< *some, ..., all* >)

II. The speaker is expected to say as much as he/she truthfully can, while observing the Relevance maxim, i.e. in a way *relevant* to the communicative exchange. An utterance can be thought of as 'relevant' if it can be understood as contributing to the goals of the communicative exchange (Leech, 1983).

III. The choice of the weaker term is reason to believe that the speaker is not able to commit to an informationally stronger statement ("Megan ate all of the cupcakes.").

IV. Therefore, as far as the speaker is able to say, the stronger statement does not hold, thus (2b).

Apart from the standard Gricean account, several newer theories of scalar implicature generation have been put forward (Chierchia, 2004; Chierchia, Fox &

Spector, 2009; Gazdar, 1979; Geurts, 1998, 2010; Fox, 2007; Harnish, 1976; Horn, 1972, 2005; Sauerland, 2004; Spector, 2007). Among the most influential of these is a group of neo-Gricean accounts that considers SIs based on logical scales such as <*some, all*> to be automatically available when a weak scalar term is encountered and canceled only if they are not supported by context (Levinson, 2000). On these accounts, such *generalized* scalar implicatures are different from context-driven, *particularized* implicatures such as (5). Another group of accounts analyze SIs as the result of a local grammatical process within the sentence (Chierchia, 2004; Chierchia et al., 2009). In such accounts, SIs contribute to the truth conditions of an utterance and, as in neo-Gricean models (Levinson, 2000), logical scales are treated as part of the lexicon, activated every time a weak scalar item is encountered. Finally, in Relevance-theoretic accounts, pragmatic inference, including SIs, is seen as an optimization process between the projected cognitive gain from computing pragmatically enriched meaning and the cognitive effort required to derive it (Sperber & Wilson, 1985/1996; Carston, 1995; Noveck & Sperber, 2007). Relevance-theoretic accounts treat all SIs as the result of context-driven inference and do not draw a distinction between generalized and particularized implicatures.

The psycholinguistic literature has shown that adults are very adept at deriving scalar inferences (e.g., Huang & Snedeker, 2009a; Breheny, Katsos & Williams, 2006; Breheny, Ferguson, & Katsos, 2013; Noveck, 2001; Bott, Bailey & Grodner, 2012). Several studies have shown, however, that children, even if they are otherwise competent language users, seem to face difficulties with SIs more broadly (Noveck, 2001; Chierchia et al., 2001; Papafragou & Musolino, 2003; Huang & Snedeker, 2009b; see next section).

1.3 How Children Calculate Sis

Early studies designed to investigate children's knowledge of quantification and propositional connectives show that children typically fail to generate scalar implicatures. Smith (1980) found that children up to the age of 9 usually treated *some* as 'some and possibly all'. Similarly, 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'). More recently, 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 between the ages of 5 and 10 interpreted the French existential quantifier *certains* ("some") in statements like "Some giraffes have long necks" as compatible with *tous* ("all"), while adults were equivocal between the logical and the pragmatic interpretations. Several studies that followed confirmed that children typically display non-adult behavior when interpreting scalar statements (Papafragou & Musolino, 2003; Guasti et al., 2005; Pouscoulous, Noveck, Politzer, & Bastide, 2007; Huang & Snedeker, 2009b; Katsos & Bishop, 2011; Barner, Brooks, & Bale 2011).

One interpretation of these results would be that linguistically competent children are simply incapable of engaging in the computations required to derive pragmatic inferences linked to scalar terms like modals and quantifiers. As Noveck and others noted, however, it could well be that the failure observed was due to task demands, since they 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 encyclopedic knowledge. For instance, in Noveck's (2001) study, the task of

evaluating the truth of the out-of-context statement "Some giraffes have long necks" against world knowledge, arguably requires a non-trivial amount of effort on the part of the participant. This leaves open the possibility that children's performance could improve under experimental circumstances that reduce those task demands.

A series of studies by Chierchia, Crain, Guasti, Gualmini and Meroni (2001) and Gualmini, Crain, Meroni, Chierchia and Guasti (2001) support this possibility. These studies explored how preschoolers interpret the disjunction operator *or*. While adults were shown to be 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 seemed oblivious to the exclusive interpretation of disjunction. Crucially, in a follow-up task, 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 (to appear) 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. For instance, even when according to the available evidence, a 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). These findings corroborate the evidence from Chierchia et al (2001) and Gualmini et al (2001) and collectively show that children at the very least have knowledge of the relative logical strength of *or* vs. *and*, or *may* vs. *have to*, and they can use that as a criterion for choosing stronger sentences over weaker ones under conditions which make the stronger sentences true. Whether this could also be interpreted as showing that children demonstrate knowledge of the information strength of the relevant expressions as opposed to simply logical strength (in Gricean terms as explained previously), remained to be seen.

Evidence for children's ability to recognize informational strength differences between otherwise identical conversational contributions came from Papafragou and Musolino (2003). In their first experiment, Greek-speaking 5-year-olds and adults were tested on three types of scales: the quantificational scale <*all, some*>, the numerical scale <*three, two*> and the aspectual scale <*finish, start*>, using a variation of the Truth Value Judgment task (Crain & McKee, 1985) called the Acceptability Judgment task. Their participants watched a series of acted out stories along with a silly puppet. After a story was concluded, the puppet was asked to say "what happened" and participants had to evaluate the statement and answer whether the puppet "answered well". On critical trials, the puppet produced a true but underinfomative statement: While the story clearly showed *every* horse in a group of three horses jump over the fence, the puppet would state that "Some/Two of the horses jumped over the fence". Children, as opposed to adult participants who were

pragmatic, were overwhelmingly logical, accepting the logically true but underinformative statement.

The authors hypothesized that this behavior on the part of children might be due to the difficulty of reconstructing the experimenter's goal in this task: to answer the question of whether the puppet "answered well", children may have been more likely than adults to base their judgments on truth (as a more powerful criterion for judging a statement) than pragmatic infelicity. Therefore, in a second study Papafragou and Musolino modified their procedure in several ways. First, to enhance awareness of the goals of the experiment, they initially trained children to detect pragmatically infelicitous statements produced by a "silly puppet" (e.g., children were encouraged to say that the statement "This is a small animal with four legs" is "silly" and the puppet should simply say "This is a dog")[1] . Second, to ensure there was a salient informativeness threshold, the experimental scenarios and test question were modified to focus on a character's performance in a task of some sort, where exactly how well someone did would be clearly important (e.g., in one of the stories, Mr. Tough brought back all three horses that had run away; when asked how Mr. Tough did, the puppet gave the response "He caught some/two of the horses"). Under these conditions, 5-year-olds were more likely to compute scalar implicatures, even though still not at adult-like levels.

---

[1] Crucially, while the term "dog" is more informative than the term "small animal with four legs", it is not necessarily a logically stronger term as is the case with quantifiers (*some/all*). Therefore any effects of training should be due to sensitivity towards the relative informativeness strength and not merely the logical strength of the scalar terms.

The effects of training and context in older children's ability to calculate SIs have been confirmed by studies following Papafragou and Musolino (2003). Guasti, Chierchia, Crain, Foppolo, Gualmini and Meroni (2005) tested Italian-speaking 7-year-olds with underinformative but true statements of the type "Some giraffes have long necks" and replicated Noveck (2001), showing that children accepted the statements, overwhelmingly favoring the logical, underinformative interpretation. In subsequent experiments, however, when the very same statements were preceded by training in rejecting underinformative statements or when the critical underinformative statements were embedded within a background story rather than presented out of context, children appeared adult-like.

To sum up, children do in fact show some ability to make pragmatic inferences. Therefore, their inability to do so in some tasks must be due to other factors. But what could possibly limit children's pragmatic abilities in those other tasks? There are two main factors that have been explored to an extent and independently claimed to be a source of children's difficulty with SIs: The first is the difficulty of deciphering the conversational goal within the experiment (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Guasti et al., 2005). Nearly all of the early studies documenting children's limitations with regard to the generation of SIs employed judgments about the acceptability of weak scalar terms in situations that made a stronger term true. Such tasks might have underestimated children's pragmatic abilities for the following reasons: First, the experimental conditions did not make it clear whether the SI should even be entertained in the first place as part of the speaker's actually intended interpretation of the message. While in ordinary cases of intentional communication the speaker clearly intends the addressee to compute the

implicature, in the studies discussed so far, the computation of SIs was not similarly facilitated by the speaker's clear intention. In Papafragou and Musolino (2003) a "silly" puppet uttered an underinformative statement without a clear intention towards a SI, while in Noveck (2001), underinformative statements appeared out of context, necessitating comparisons against encyclopedic knowledge to evaluate their truthfulness. It is not clear under such conditions that any participant would (or should for that matter) arrive to the conclusion that the generation of an SI is warranted.

Second, there is recent evidence that regardless of their final acceptance of a logically true but underinformative statement, adults weigh both the weak and strong interpretations against the perceived experimenter's goal, as evidenced by their longer response times and eye-gaze patterns when they were provided with underinformative statements (Tavano & Kaiser, 2010). Given adults' mixed pattern of responses in many studies and the fact that arguably both weak and strong interpretations are entertained and weighed before a commitment is made, children's tendency to accept underinformative statements in the same environments should not necessarily be interpreted as inability to make the necessary pragmatic inference. If so, adults who commit to a logical choice in similar tasks should also be considered unable to make the relevant pragmatic inference.

Support for the hypothesis that children might not be completely oblivious to pragmatic infelicity despite failing to reject under-informative statements comes from Katsos and Bishop (2011). Preschoolers aged 5 to 6 were presented with stimuli depicting cartoon characters having moved a number of items from one end of the screen to the other. Another cartoon character who supposedly was not fluent in English would provide statements describing what happened. Children were asked to

reward a speaker with a strawberry, the size of which would be either small, big, or huge (a 3-point Likert scale) depending on how good the speaker's statements were. It was found that fully informative responses (e.g. "The boy lifted all the books" when the boy had lifted 5 out of 5 books) were rewarded with huge strawberries 85% of the time, underinformative statements (e.g. "The mouse picked up some of the carrots", when the mouse had picked up all of the carrots) with big strawberries 85% of the time, while false responses (e.g. "The dancer picked up some of the yellow flowers", when the dancer had picked up some of the red flowers) with small strawberries 95% of the time. Children's behavior did not differ significantly from that of adults who were at ceiling. Strikingly, in a standard *Yes-No* judgment task with the same materials, children of the same age overwhelmingly accepted underinformative statements (Katsos & Bishop, 2011). Therefore, there is evidence that while children may notice pragmatic infelicity (entertaining the necessary pragmatic inferences), they may nevertheless be unwilling to reject infelicitous statements for independent reasons.

The above results indicate that task goals and the way children (as opposed to adult communicators) may understand those goals are likely to be the factors that limit children's performance in SI tasks. One would expect that if this indeed the case, then studies featuring tasks with requirements and goals that are more transparent to children should have more success at uncovering children's elusive pragmatic abilities. Papafragou and Tantalou (2004) provide such evidence, testing Greek-speaking 5-year-olds pragmatic inferences. The children were shown stories where animals were asked to perform different tasks. The animals would supposedly perform the tasks without the children being able to directly evaluate how well they did

because the animals went 'off-stage'. When the animals returned and were asked if they performed the task, they would sometimes answer with underinformative statements (e.g. Experimenter: "Did you color the stars?" Animal: "I colored some."). If the children interpreted the statements pragmatically, they would conclude that the animals failed the task and would not give them a reward, while if they interpreted the statements logically, they would conclude that the animals might have performed the task and they would reward them. The children overwhelmingly denied the animals the reward, as was expected for a pragmatic interpretation of the underinformative statements. What is even more striking is the fact that children's own reports as to why they denied the animals the reward made reference to the stronger scalar alternative (e.g. "The animal did not color ALL the stars") exactly as expected from a communicator who demonstrates a full understanding of the inferential process.

Even though the "conversational goals" explanation is both theoretically motivated, at least within context-based accounts for scalar inference (Grice, 1975; Sperber & Wilson, 1986/1995; Carston 1995; Noveck & Sperber, 2007), and backed up by empirical evidence (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Guasti et al., 2005), what is exactly meant by "conversational goals" has never been made explicit enough or formalized to the extent that it could systematically and unambiguously be manipulated within an experimental paradigm. Perhaps for this same reason, and despite the general plausibility of such an account, more recent research moved to link the source of children's difficulty to a more constrained factor that seems to be motivated to an extent by grammatical accounts of SI generation (Chierchia, 2004; Chierchia et al., 2009).

This more recent hypothesis that has been put forward is that a major part of children's problem lies with generating scalar alternatives when faced with a weak scalar term (Chierchia et al., 2001; Gualmini et al., 2001; Barner et al., 2011, Bale & Barner, 2013). 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, to appear). This indicates that children can compare alternatives to the weak scalar term and assess their relative informativeness. This fact, however, does not guarantee that children can independently access the relevant scalar alternatives on their own when needed.

A recent study by Barner, Brooks and Bale (2011) offers more direct evidence for the role of the accessibility of unspoken lexical alternatives on children's SI calculation. Barner et al. 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 responded affirmatively about 66% of the time regardless of the form of the question. Children's affirmative response to the question with bare *some* was 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 (e.g., when the experimenter listed the animals that were supposed to be sleeping), children's generation of SIs improves significantly.

While there is evidence that the accessibility of scalar alternatives is one source of children's difficulties with SIs (Chierchia et al. 2001; Gualmini et al., 2001, Barner et al., 2011), the precise role and potency of lexical alternatives, as well as the factors that determine their accessibility, remain imprecise. For instance, direct manipulations of the accessibility of alternatives in the context of a single task are lacking (even though see Bagassi, D'Addario, Macchi, & Sala, 2009, where 7 and 10-year-olds were shown to be more likely to interpret the quantifier "some" as meaning "only some" if they are first exposed to "all").

1.4 Open Questions

The developmental evidence reviewed above sketches a rather complex picture with regard to children's abilities in scalar inference. While they are not altogether oblivious to pragmatic inferences in situations involving scalar inference, children seem to have considerable problems in generating SIs spontaneously, although those problems are mitigated with different manipulations within different tasks. The field appears fragmented, with different studies attempting to explain children's difficulties

by looking at different factors that might influence their performance in scalar inferences. Individually, all of the studies cited above succeed to a greater or lesser extent to attribute some part of the difficulty children face in scalar inference to the factors they investigate. What is still missing, however, is a unifying hypothesis incorporating findings from different studies and able to account for the complex pattern we are observing. Such a hypothesis should also identify the pragmatic ability that is developing over time and turns children into adult communicators. We identify two hypotheses that could achieve this:

One possibility, referred to from now on as the *lean hypothesis*[2] , is that scalar alternatives feed into a simple, bottom-up contrastive mechanism similar to what has been suggested for mutual exclusivity in the context of lexical learning (see de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011 for recent findings and discussion). The presence of the weak alternative negates the stronger alternative in a context-independent way. For example, in an utterance "Some of the gnomes have a pie", the presence of the proposition *some(P)* automatically negates the stronger scalar alternative *all(P)* that corresponds to "All of the gnomes have a pie", regardless of contextual considerations. Since this would potentially lead to an overgeneration of SIs, unwarranted SIs are later filtered through context and if necessary discarded. In this account, accessibility of the alternatives is not mediated by conversational relevance and the factors of accessibility of the alternatives and conversational goals, as explored in prior research are independent of one another.

---

[2] It should be noted here that the lean hypothesis (at least as presented here) has some a priori difficulties in explaining away the fact that context does seem to influence the generation (or lack thereof) of SIs by adult speakers (Bergen & Grodner, 2012; Breheny et al., 2013; Degen 2013; Zondervan, 2010).

For instance, such a minimalist computation is the following, adapted here from Bale and Barner (2013)[3]: On hearing an utterance U containing a scalar term SCL the hearer computes the semantic content of the sentence corresponding to the utterance that is typically understood as the literal meaning (Step A). The hearer then generates a stronger alternative U_{alt} (Step B). The hearer adds a pragmatic enrichment to the semantic content of U by negating the stronger alternative (Step C).

A. Compute semantic content of utterance U containing scalar term SCL

   ⟦**Some of the gnomes have a pie.**⟧ = Some and possibly all of the gnomes have a pie.

B. Generate all alternative propositions (p1, p2,…, pn) to U (U_{alt}), that are not logically entailed by U by replacing SCL with its stronger scalar alternatives.

   all of the gnomes…

C. Strengthen the semantic meaning of U by negating all stronger alternatives.

   ⟦**Some of the gnomes have a pie.**⟧ → Some but not all of the gnomes have a pie.

In the lean hypothesis account, children's failures with SIs would be attributed to failures in accessing the necessary stronger alternative altogether (step B of the

---

[3] Bale and Barner's account is arguably richer than the bare-bones version we present here. They make explicit reference to "alternatives that might have been uttered" which introduces Gricean considerations, something that they acknowledge in the paper. In addition, they reference an additional step in the computation, namely negating weaker scalar alternatives to the original, that we subsume in B) by positing a restriction to generate only stronger alternatives.

computation). As children develop, experience with scales and greater cognitive resources, such as working memory, allow them to access the stronger alternative more easily, and eventually they mature into the typical adult behavior of generating the SI when necessary.

Another possibility, the *rich hypothesis*, is that lexical alternatives may lead to SI generation through a richer, fundamentally pragmatic, inferential mechanism that includes reasoning about the conversational goal of the discourse. While conversational goals and accessibility to scalar alternatives have until now been explored as independent factors affecting SI performance in children, they are potentially very closely linked: Accessibility of a more informative term would be mediated by conversational relevance, such that it would push children to SI generation only when the stronger alternative to be generated is relevant to the goals of conversation as those are defined by Gricean pragmatics. Going back to the classic Gricean formulation of SI derivation that we saw in section 1.2, it is evident that in this account, children's difficulty seems to correspond to step (II) of the computation:

> *(II). The speaker is expected to say as much as he/she truthfully can, while observing the Relevance maxim, that is in a way relevant to the communicative exchange. An utterance can be thought of as 'relevant' if it can be understood as contributing to the goals of the communicative exchange (Leech, 1983).*

A difficulty in evaluating relevance quickly and efficiently can disrupt the rest of the computation and result in failure to generate a SI even when one would be typically expected. As children develop, their ability to exploit contextual information becomes

more efficient, and like pragmatically savvy adults, they draw pragmatic inferences (like SIs) when those are warranted.

1.5 The Present Studies

In the present studies we attempt to explain what is responsible for children's apparent failures in linguistic pragmatics. To do that we test the lean and rich hypotheses in order to see whether the previously identified factors of accessibility of the scalar alternatives and conversational goals can be reconciled. The experimental question at the heart of this inquiry is whether the accessibility of alternatives is mediated by pragmatic considerations of relevance and conversational goals, or not. The studies test the rich hypothesis described above systematically, within a single experimental paradigm, by disentangling the issues of the accessibility and relevance of the alternatives for the first time in the field. If confirmed, the rich hypothesis unifies the "conversational goals" and the "accessibility of alternatives" approaches under a unified account based on the notion of relevance. Moreover, it constrains the somewhat vague "conversational goals" account into a more specific formulation (conversational relevance mediates accessibility of scalar alternatives) and at the same time it provides a theoretically motivated and empirically tested account of how exactly scalar alternatives become accessible.

I focus on a well-known quantificational scale <*some, all*>. In Experiment 1 I test whether the presence of the 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*). The availability of the stronger alternative is going to be manipulated through the order of *some*- and *all*-statements. If this is a limiting factor

in children's computation of SIs, then children's computation of SIs will improve when a contrast between strong (*all*) and weak (*some*) scalar alternatives is available.

In Experiment 2, I explore the nature of the mechanism that uses lexical alternatives in an attempt to gain insights into how such alternatives lead to SI generation. If it is the case that lexical alternatives lead to SI generation via a non-pragmatic, bottom-up constraint, then children's computation of SIs will improve by the mere availability of the stronger alternative *all* regardless of whether the scalar alternatives can be seen as relevant to the conversational goal. If, however, alternatives lead to SI generation through a pragmatic inferential process linked to relevance, then the availability of the stronger alternative *all* should have no effect on children's computation of SIs if the alternatives do not appear to be relevant to the conversational goal.

In Experiment 3, I ask whether the generation of SIs in children can be driven purely by conversational goals that clearly make scalar alternatives relevant even in the lexical absence of the stronger scalar term itself within the experimental paradigm.

The results of these studies will have broad significance for the field of pragmatic development: If the rich hypothesis is confirmed, we will have strong evidence that the apparent discontinuity between children and adult communicators is primarily due to factors linked to a communicator's ability to efficiently exploit context in order to interpret what is relevant in a communicative situation. Thus, the notion of conversational relevance (Grice, 1975), will be placed at the center of future research on pragmatic development. On the other hand, if the lean hypothesis is confirmed, we will have evidence that conversational relevance might not have a central role in children's pragmatic development and we should perhaps focus instead

on other factors that affect linguistic pragmatics, such as the role of grammar and the lexicon.

Chapter 2

EXPERIMENT 1

Experiment 1 uses an Acceptability Judgment Task (AJT) similar to that used in prior literature (Chierchia et al., 2001; Gualmini et al., 2001; Papafragou & Musolino, 2003; Guasti et al., 2005; Katsos & Bishop, 2011). It tests whether the stronger scalar alternative "all" will encourage children to generate a SI when they encounter they weak scalar term "some". Accessibility of the stronger alternative is manipulated through the ordering of the statements the participants will have to judge. Despite some recent criticism for such tasks (Papafragou & Tantalou, 2004; Pouscoulous et al., 2007; Huang & Snedeker, 2009b; Katsos & Bishop, 2011), AJT tasks have well demonstrated strengths and allow our results to be more easily comparable to those of prior research. In our task, scalar terms are embedded within statements that need to be evaluated based on visual evidence. Critical trials designed to assess children's generation of scalar inferences consist of *True-and-Infelicitous-Some*-statements that need to be rejected if one derives a scalar implicature. Such statements will be of the form *some(P)*, in the context of a scene where *all(P)* will be evidently true. If *all(P)* is true, then *some(P)* is necessarily true as well, but at the same time, underinformative and thus pragmatically infelicitous. Rejections need to be appropriately justified so that they provide convincing evidence that a scalar inference was drawn.

Our design has some noteworthy differences from prior studies: First, we include semantic controls[4] to test for children's understanding of the actual semantics

[4] Noveck (2001) included semantic trials but both that study and subsequent studies have not used semantic trials systematically as controls.

of the quantifiers. While there is evidence that children show some understanding of the quantifiers *all* and *some* from around the age of two, children's use of *some* is not really consistent until the age of 5 and even then it is not completely error free (Barner et al., 2009). We want to make sure to look at the pragmatic competence of children who have already mastered the relevant semantics of the quantifiers. The semantic trials will allow us to do this by excluding participants whose performance in the semantic trials is low.

Second, we restrict the universe of discourse. In previous studies like Noveck (2001) statements like "Some giraffes have long necks" can be completely open about the set of entities that should be considered. If one is to consider all giraffes in the world, what about baby giraffes? Do they have sufficiently long necks, or not? In other studies like Papafragou & Musolino (2003), statements like "Some of the horses jumped over the fence", if not properly restricted, can be ambiguous between the set of horses referred to by the experimenters and the set of horses in the whole world. We used a unique set of 4 novel creatures ("blickets") in order to restrict discourse to the visual content provided for each trial.

2.1 Method

2.1.1 Participants

We tested 90 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. The adults were college students recruited from the University of Delaware, and received course credit for their participation. An additional group of 7 children were tested but excluded from the

analysis for failure to follow instructions (n = 3) or for misidentifying objects in the displays as made evident by their responses (n = 4).

2.1.2 Materials and Procedure

The task was an Acceptability Judgment Task similar to that in Papafragou and Musolino (2003). Children sat in front of a laptop PC computer and were shown the slides depicting the experimental stimuli. A first experimenter introduced the task to the children by introducing a hand-held puppet, Max the silly gorilla, "who says silly things sometimes", and explaining that they would see some pictures on the computer together. Participants were told that the puppet would describe the pictures and that they would have to say whether the puppet "said it well or not". They would also have to justify their answer in case they rejected the puppet's statement. A second experimenter animated the puppet and provided the appropriate statements, while the first experimenter wrote children's answers down on 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 in answer sheets, with the options *Yes/No* and space to justify *No*-answers 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 that 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 on an introductory slide. Ben was shown to use his magic wand to create the 4 "blickets", novel animate creatures that would appear on all test slides. Participants were informed that these are the only blickets "in the whole world".

Blickets were paired with 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 (*partial set* scenes). For each slide Max offered a statement containing a quantifier (*some* or *all*). Scene type (full set vs. partial set) was crossed with quantifier type (*some* vs. *all*) to provide 4 types of trials: a) In full set/*All*-trials, 4 out of 4 blickets had the item and participants heard: "All of the blickets have an X.". I will refer to these as *True-All*-trials. b) In partial set/ *All*-trials, 3 out of 4 blickets had the item and participants heard: "All of the blickets have an X." These were the *False-All*-trials. c) In partial set/*Some*-trials, 3 out of 4 blickets had the item and participants heard: "Some of the blickets have an X." (*True-Some*-trials). Finally, in full set/*Some*-trials, 4 out of 4 blickets had the item and participants heard: "Some of the blickets have an X." (*True-and-Infelicitous-Some*-trials). The first three types of trials tested participants' semantic judgments about *some* and *all*. The last type of trial (where *some* was used despite the fact that all 4 blickets have the item) tested participants'

pragmatic judgment (i.e., their ability to generate SIs). Examples of the visual scenes and statements for each trial type can be found in Figure 1.

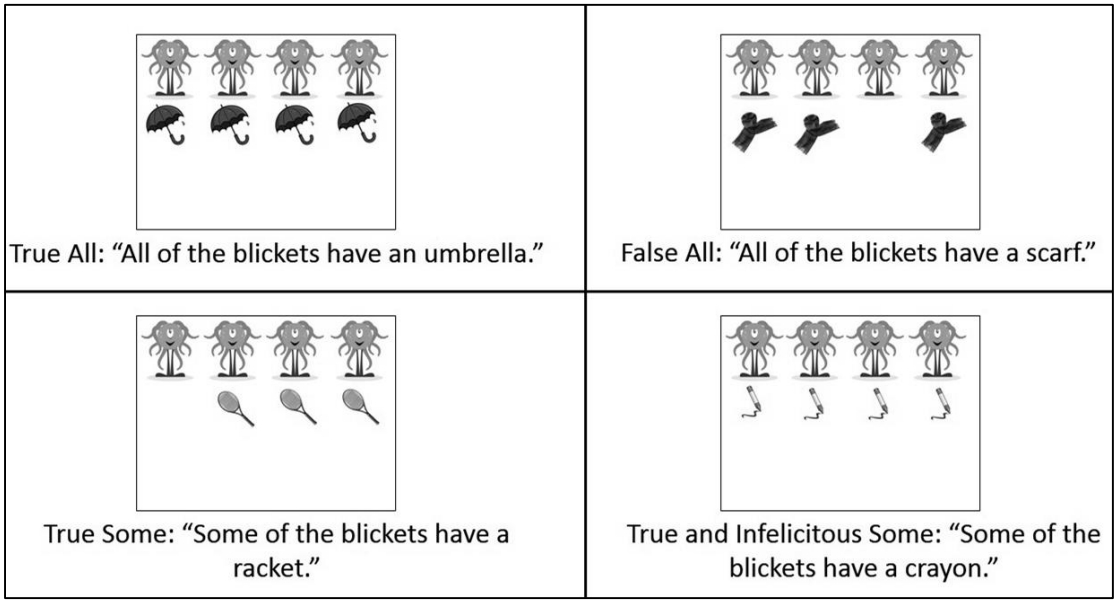

Figure 1. Types of trials for Experiment 1.

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.

The internal order of statements within each battery was manipulated across 3 between-subjects conditions: In the *Mixed* condition, *some*- and *all*- statements were intermixed in a pseudorandom order so that the stronger (*all*) alternative would be lexically present and made available to children when they would have to judge the statements with the weaker (*some*) scalar term. 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 the stronger alternative (*all*) would not be made available to children when they had to judge the *True-and-Infelicitous-Some*-statements. Finally, in the Infelicitous *Some-First* condition, the *some*-block of the previous condition was further split into two blocks, with *True-and-Infelicitous-Some*-trials always first and *True-Some*-trials always last, so that the felicitous and infelicitous uses of *some* would not be made available as alternatives to the children.

2.2 Predictions

If the stronger scalar term has a role in scalar inference, children's rejection of *True-and-Infelicitous Some*-statements should improve when both the strong (*all*) and weak (*some*) scalar alternatives are made available to children. Therefore children's performance with these pragmatic trials should be better in the *Mixed* than in the *Some-First* or *Infelicitous-Some-First* conditions, since the relevant stronger alternative *all* will be available to the children before they reach a *True-and-Infelicitous-Some*-statement in the *Mixed*, but not in the other two conditions. If the accessibility of the stronger term does not affect scalar inference, then children's performance in terms of SI generation (rejection of *True-and-Infelicitous-Some*-trials) should be comparable across the three conditions.

Depending on what can function as an alternative, there may or may not be differences in the performance of children in SI generation between the *Some-First* and *Infelicitous-Some-First* conditions: if SI generation is facilitated only through the accessibility of the strong scalar alternative (*all*), then no difference between the *Some-First* and *Infelicitous-Some-First* conditions is expected; if, however, the

felicitous and infelicitous uses of *some* can also function as alternatives, 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*, *True-Some*). Finally, no difference in adult performance is expected between conditions for either the semantic or the pragmatic trials.

2.3 Coding

Yes answers were coded as correct in the case of true statements. *No* answers were coded as correct in the case of false or *True-and-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*, *True-Some*, *True-and-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 *True-Some*-trials was categorized as a Passer for that trial type.

2.4 Results

Adult performance was 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, $p = 1$; *True-Some*-trials, $p = 1$; *True-and-Infelicitous-Some*-trials, $p = .31$).

Table 1. Adult performance in Experiment 1. The numbers represent Passers vs. Failers in corresponding trials.

TRIAL TYPE		**CONDITION**		
		Mixed	**Some-First**	**Inf-Some-First**
True-All	Passers	12	12	12
	Failers	0	0	0
False-All	Passers	12	12	12
	Failers	0	0	0
True-Some	Passers	12	12	12
	Failers	0	0	0
True-and-Infelicitous-Some	Passers	12	12	10
	Failers	0	0	2

Table 2. Child performance in Experiment 1. The numbers represent Passers vs. Failers in corresponding trials.

TRIAL TYPE		**CONDITION**		
		Mixed	**Some-First**	**Inf-Some-First**
True-All	Passers	30	29	29
	Failers	0	1	1
False-All	Passers	24	24	24
	Failers	6	6	6
True-Some	Passers	26	26	23
	Failers	4	4	7
True-and-Infelicitous-Some	Passers	23	14	7
	Failers	7	16	23

Table 2 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 ($p = 1$), *False-All*-trials ($p = 1$), or *True-Some*-trials ($p = .52$). Turning to the critical *True-and-Infelicitous-Some-*

trials, children appeared to be pragmatic in the *Mixed* condition with 23 Passers and 7 Failers, and logical (non-pragmatic) in the *Infelicitous-Some-First* condition with only 7 Passers and 23 Failers. In the *Some-First* condition children seemed to be divided between a pragmatic and a logical interpretation of the *True-and-Infelicitous Some*-statements with 14 Passers and 16 Failers. A Fisher's Exact test on a 2x3 contingency table revealed a significant difference (p = .0001) between the numbers of Passers and Failers for *True-and-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. It was found that the *Mixed* condition had significantly more Passers than the *Infelicitous Some-First* condition (p = .001) and the *Some-First* condition (p = .03). The *Some-First* condition did not have significantly more Passers than the Infelicitous *Some-First* condition (p = .103).

When asked to justify their rejections of *True-and-Infelicitous-Some*-statements, children overwhelmingly referenced either the stronger scalar term (e.g. "*All* of them/the blickets have an X"; 38 out of 44, 86% of justifications), or mentioned the number of blickets and items available (e.g. "There is 4 blickets and 4 crayons"; 5 out of 44, 11% of justifications). This shows that children rejected the *True-and-Infelicitous Some*-statements for the correct reason, namely because they generated the appropriate SI.

As is obvious from Table 2, some of the children performed poorly in the *False-All* and *True-Some*-trials. 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 derive conclusions about these children's pragmatic competence with quantifiers. To address this concern, we conducted a second analysis excluding children who had

under .75 correct in either the *True-All*, *False-All*, or *True-Some*-statements. This resulted in n = 8 children being excluded in the *Mixed* condition, n = 10 in the *Some-First* and n = 13 in the *Infelicitous-Some-First* condition.

This new analysis examined only performance on the *True-and-Infelicitous-Some*-trials (see Table 3) in children who can be safely 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 significant difference between the numbers of Passers vs. Failers for the 3 different conditions (p = .0007), 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 again revealed a significant difference (p = .009), with the *Mixed* condition having significantly more Passers than the *Some-First* condition. Comparing the *Mixed* and the *Infelicitous-Some-First* condition there was again a significant difference (p = .0006), with the *Mixed* condition having significantly more Passers than the *Infelicitous-Some-First* condition. Finally, comparing the *Some-First* condition to the *Infelicitous-Some-First* condition once again revealed no significant difference (p = .33).

Table 3. *Some/all*-knowers' performance on *True-and-Infelicitous-Some*-trials of Experiment 1.

TRIAL TYPE		**CONDITION**		
		Mixed	**Some-First**	**Inf-Some-First**
True-and-Infelicitous-Some	Passers	21	12	7
	Failers	1	8	10

Finally, the performance of the last group of children was compared with that of adults on the *True-and-Infelicitous-Some*-trials with Fisher's Exact Test on 2x2 contingency tables. There was no difference between age groups in the *Mixed* condition ($p = 1$), a significant difference in the *Some-First* condition, with the adult group having significantly more Passers than the child group ($p = .014$), and finally a trend towards a significant difference in the numbers of Passers vs. Failers for the *Infelicitous Some-First* condition ($p = .053$), with adults having more Passers than the child group.

2.5 Discussion

Experiment 1 was conducted to confirm that the presence of the stronger scalar alternative *all* would facilitate children's generation of SIs in our paradigm (Barner et al., 2011; Gualmini et al., 2001). This hypothesis was supported by our data. In the *Mixed* condition, where *some*- and *all*-statements were intermixed so that the strong scalar term was made available to children by the time they had to judge the underinformative *True-and-Infelicitous-Some*-statements, children were very successful at generating the appropriate scalar inference by rejecting infelicitous statements with *some*. Children's performance fell significantly when the stronger alternative *all* was not made lexically accessible to them (in the *Some-First* and *Infelicitous-Some-First* conditions). There was no difference in *Infelicitous Some*-trials between the *Some-First* and *Infelicitous-Some-First* condition.

Our task contained an important internal control in the form of the semantic trials. Some children did not appear to have a firm command of the quantifiers involved (*some/all*). 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 support the hypothesis that the stronger alternative facilitates scalar implicature generation in children.

As we can see from the comparisons of child to adult performance, the stronger alternative (*all*) in the *Mixed* condition makes children's behavior adult-like in our task, something that previous research using Acceptability Judgment Tasks (e.g., Guasti et al., 2005; Papafragou & Musolino, 2003) had achieved only through training or the use of contextual support. As expected, the performance of children and adults differs when the stronger alternative is not present (*Some-First* condition; see also trend in *Infelicitous-Some-First* condition); adults do not seem to be affected by the absence of lexical contrast, presumably because the task is easy enough to be transparent to them.

The results of Experiment 1, while instructive, cannot adjudicate between the lean and rich hypotheses mentioned earlier: The experiment manipulates only the availability of the alternatives, however the availability of the alternatives can also affect their relevance. While children who succeeded evidently did recover the conversational goal and accessed the relevant alternatives, we simply cannot know from this experiment what the reason for failing was for the children who did not derive the SI. The reason for failures could be attributed to either of two possibilities. One possibility for children's failure is consistent with the lean hypothesis. In this case, children could not spontaneously recover the stronger alternative in the *Some-First* and *Infelicitous-Some-First* conditions, and therefore they were unable to generate the SI (regardless of the conversational goal). The other possibility, consistent with the rich hypothesis, is that children could simply not identify the conversational goal in these conditions.

In order to appreciate the potential role of alternatives in recovering the goal of the task, we need to think about the role of all-statements in considering relevant alternatives for *some*-statements: Any mention of *all* in a very similar, minimally different sentence (e.g. *All of the blickets have a crayon* vs. *Some of the blickets have an umbrella*) will encourage participants to think that quantifiers constitute relevant alternatives. In the *Mixed* condition, the contrast between the *all*- and *some*-statements was available for children, and helped them think that quantifiers are relevant alternatives. In the *Some-First* and *Infelicitous-Some-First* conditions, however, children encountered the critical *True-and-Infelicitous-Some*-statements before they had a chance to encounter an *all*-statement. Without the contrast between *all(P)* and *some(P)*, a substantial number of children might have been unable to identify what constitutes a relevant scalar alternative.

Chapter 3

EXPERIMENT 2

The results of Experiment 1 provide evidence that accessibility of the stronger alternative is a factor in SI generation for children. However, the precise nature of children's difficulty with SIs and the way lexical alternatives help them remain unclear. Recall that one possibility is that children do not know what alternatives enter into a scalar relationship (the lean hypothesis). On this hypothesis, the difference between adult and child performance in SI tasks can be explained by relative experience with scales and scalemates. It is this experience that develops as children mature into having an adult-like performance. If this is the case, one would expect that, other things being equal, simply providing alternatives in an environment where SI generation is possible should lead children to derive SIs.

The second hypothesis is that children do not always realize when alternatives are relevant in order to generate a SI at a given time (the rich hypothesis). This hypothesis is consistent with the notion that accessibility of scalar alternatives is mediated through conversational relevance. If this is correct, the fundamental difference between children and adult communicators is their ability to home in on the conversational goal at any given time and consequently identify scalar alternatives that are relevant within the identified conversational goal. In this case, simply providing scalar alternatives should not necessarily lead children to derive SIs. What is needed instead for SI generation is the crucial realization that the alternatives provided are relevant for the goal(s) of discourse.

Experiment 2 directly tests these two hypotheses. We manipulate the degree to which alternatives can be easily recognized as relevant by children. This is achieved

by comparing a case where the conversational goal remains stable (in a way that helps establish the relevance of the stronger alternative) with another case where it switches during the experiment (in such a way that it renders the stronger alternative irrelevant). In both cases, the scalar alternatives remain lexically accessible to the children (i.e. they are presented to them within the experimental battery in exactly the same way). If the lean hypothesis is correct, the presence of a single vs. multiple conversational goals should not affect SI computation (since accessibility of the stronger scalar alternative is guaranteed). If the rich hypothesis is correct, 5-year-olds should benefit from the accessibility of the stronger scalar term only when this term is relevant to the conversational goal.

3.1 Method

3.1.1 Participants

We tested a group of 50 typically developing 5-year-old children (4;9 – 5;8, M=5;0) and 24 adult controls, all monolingual speakers of English. None of these participants had taken part in Experiment 1. The children were recruited from daycare centers in Newark, DE. The adults were college students recruited from the University of Delaware and received course credit for their participation. An additional group of 4 children were tested but excluded from the analysis for failure to follow instructions.

3.1.2 Materials and Procedure

The materials and procedure were very similar to those in Experiment 1 with the following differences: The *all*-statements were always presented in a first block, and the *some*-statements in a second block so that lexical contrast between the stronger (*all*) and weaker (*some*) scalar terms could be established.

There were two between subjects conditions that differed in a single aspect of the *all* block: In the *Quantity* condition, the *False-All*-trials would involve 3 out of 4 blickets having the correct item (for example a scarf) and be accompanied by a statement such as "All of the blickets have a scarf". In the *Object* condition the same trial would involve 4 out of 4 blickets having an item different from the one mentioned in the statement. For example, each of the 4 blickets would have a shovel and the statement would be: "All of the blickets have a scarf". The statement was still false in the context of the visual scene, but the reason was different from the previous condition: here, the object was wrong (in the *Quantity* condition, the quantity of the objects was wrong). Figure 3 shows example scenes and statements for Experiment 2.

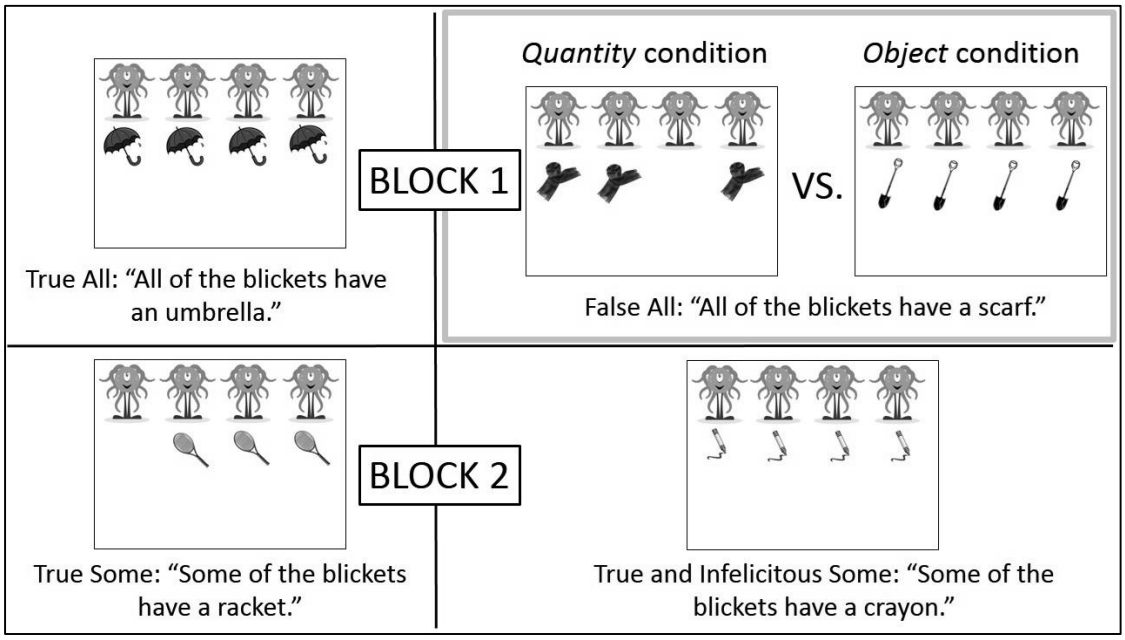

Figure 2. Types of trials for Experiment 2. *False-All*-trials differ between the *Quantity* and *Object* condition.

The *False-All*-statements provided information for the conversational goal in the way described earlier for Experiment 1: By uncovering the dimension of the trials that the puppet is likely to err in (object quantity vs. object identity), the *False-All*-statements implicitly pointed to a distinct conversational goal in each condition: In the *Quantity* condition, the implicit basis upon which the participants were called to evaluate each statement was whether the quantity of blickets shown to possess an item in the scene matched the quantity of blickets mentioned in the statement. This conversational goal remained stable throughout the experiment: it was established in the first (*all*) block through the *False-All*-trials and could later be brought to bear on judgments of the *True-and-Infelicitous-Some*-statements. In the *Object* condition, however, the conversational goal changed between the first and the second block. The first (*all*) block, especially the *False-All*-statements, should arguably lead participants to identify object identity as the conversational goal (i.e. whether the blickets possessed the stated object kind or not). In the second (*some*) block, however, if participants were to detect the infelicity of the *True-and-Infelicitous-Some*-trials, they would have to recover a different conversational goal (namely, whether the quantity of blickets in possession of a certain object was as stated in the sentence or not).

3.2 Predictions

If the lean hypothesis is correct, then children should successfully reject *True-and-Infelicitous-Some*-statements in both conditions, since *all* is lexically accessible in both conditions by virtue of being present throughout the first block. However, if the rich hypothesis is correct, then children should be more successful in the *Quantity* than the *Object* condition, since the stronger alternative *all* is only relevant in the *Quantity*

condition where it is integrated in the same conversational goal with the weak counterpart *some* that should trigger the SI.

Consider the true/infelicitous "Some of the blickets have a racket" (Fig.1) uttered when all of the blickets have a racket. If children believe that the conversational goal is to evaluate whether the puppet got the quantity of blickets right (*Quantity* condition) and already have access to the stronger *all* term, they should easily reject the *some*-statement (since all of the blickets have a racket). But if children believe that the conversational goal is to evaluate whether the puppet got the object owned by the blickets right (*Object* condition), even if they have access to the stronger *all* term, they might not reject the statement (since some of the blickets indeed have a racket).

Adult performance is not expected to differ between the two conditions as adult communicators should in principle be able to identify the different conversational goals in each block of the *Object* condition and still derive the corresponding SI.

3.3 Coding

The coding scheme was identical to the one used for Experiment 1. Participants' correct scores were used to categorize them 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).

3.4 Results

Adult performance was very high for all conditions and trial types. Table 4 summarizes adult performance. Fisher's Exact test analyses on 2x2 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, $p = 1$; *True-Some*-trials, $p = 1$; *True-and-Infelicitous-Some*-trials, $p = 1$). Adults were overwhelmingly pragmatic in the *True-and-Infelicitous-Some*-trials.

Table 4. Adult performance in Experiment 2.

TRIAL TYPE		**CONDITION**	
		Quantity	**Object**
True-All	Passers	12	12
	Failers	0	0
False-All	Passers	12	12
	Failers	0	0
True-Some	Passers	12	12
	Failers	0	0
True-and-Infelicitous-Some	Passers	11	12
	Failers	1	0

Children performed well with the 3 semantic trial types (see Table 5). Fisher's Exact Tests on 2x2 contingency tables did not reveal significant differences in the numbers of Passers vs. Failers across the two conditions for either the *True-All*-trials ($p = 1$) or *False-All*-trials ($p = .357$). We did observe a difference approaching significance in the numbers of Passers vs. Failers in the *True-Some*-trials ($p = .05$). Turning to the critical *True-and-Infelicitous-Some*-trials, children appeared to be more pragmatic in the *Quantity* condition (17 Passers and 9 Failers) and more logical (non-pragmatic) in the *Object* condition with only 8 Passers and 16 Failers (see Table 5). A Fisher's Exact test on a 2x2 contingency table revealed a significant difference ($p = .046$) between the numbers of Passers and Failers for *True-and-Infelicitous-Some*-trials between the two conditions.

As in Experiment 1, when asked to justify their rejections of *True-and-Infelicitous-Some*-statements, children referenced either the stronger scalar term ("*All* of them have an X" 23 out of 26, 88% of justifications), or the number of blickets that possessed an item ("Because there is 4 blickets and 4 X's" 3 out of 26, 12% of justifications).

Table 5. Child performance in Experiment 2.

TRIAL TYPE		**CONDITION**	
		Quantity	**Object**
True-All	Passers	26	24
	Failers	0	0
False-All	Passers	22	23
	Failers	4	1
True-Some	Passers	19	23
	Failers	7	1
True-and-Infelicitous-Some	Passers	17	8
	Failers	9	16

After this initial analysis and for the same reasons as in Experiment 1, we conducted a second analysis excluding children who were Failers (had under .75 correct) in either the *True-All*, *False-All*, and *True-Some*-statements. This resulted in n = 9 children being excluded in the *Quantity* condition, and n = 1 child excluded in the *Object* condition[5]. All of these children can be assumed to have the correct semantics

[5] The exclusion criterion for the *Object* condition was based solely on the *True-Some*-trials, since the *True-All* and *False-All*-trials had to be evaluated based on the identity of the items present and not on quantification and thus would not necessarily offer any evidence of a participant's knowledge of quantifiers. That said, even if we included the *True-All* and *False-All*-trials in the exclusion criteria for the *Object* condition for the sake of uniformity, only one additional participant would have been excluded, and the analyses would not be affected.

for *some*, and *all*. A Fisher's Exact test on the 2x2 contingency table in Table 6 revealed a significant difference between the numbers of Passers vs. Failers for the two different conditions (p = .000007) with the *Quantity* condition having significantly more Passers than the *Object* condition, confirming the results of the first analysis.

Table 6. *Some/all*-knowers' performance in *True-and-Infelicitous-Some*-trials of Experiment 2.

TRIAL TYPE		**CONDITION**	
		Quantity	**Object**
True-and-Infelicitous-Some	Passers	17	7
	Failers	0	16

We then compared the performance of the children who have a solid grasp of quantifier semantics with that of adults, with Fisher's Exact Test on 2x2 contingency tables. No difference was found between age groups in the *Quantity* condition (p = .414), but we did find a significant difference in the *Object* condition, with the adult group having significantly more Passers than the child group (p = .00006).

Finally, to complete our analysis of children's performance in the *True-and-Infelicitous-Some*-statements after the exclusions (Table 6), we ran Fisher's Exact Test on 2x2 contingency tables comparing the *Quantity* condition to each of the three conditions of Experiment 1. No difference was found between the *Quantity* and the *Mixed* condition (p = 1). However significant differences were found in the numbers of Passers vs. Failers between the *Quantity* and the other two conditions (*Some-First*, p = .004; *Infelicitous-Some-First*, p = .0003) with the *Quantity* condition having significantly more Passers than either of the other two conditions. Finally, we

compared the *Object* condition to the three conditions from Experiment 1. We found no difference between the *Object* condition and either the *Some-First* (p = .07) or the *Infelicitous-Some-First* condition (p = .521). However, we did find a significant difference between the *Object* and the *Mixed* condition from Experiment 1 (p = .000008) with the *Mixed* condition having significantly more Passers than the *Object* condition.

3.5 Discussion

The starting point of Experiment 2 was the idea that at least part of the problem children face in SI generation lies in failing to access the appropriate stronger alternative when a weak scalar term is used (Experiment 1; see also Barner et al., 2011, Papafragou & Skordos, in press, among others). Of interest was whether accessibility of the stronger alternative could bear the explanatory burden of children's failure with SIs alone (lean hypothesis), or whether accessibility of alternatives was sensitive to the role of alternatives within the conversational goals of the exchange (rich hypothesis).

We presented children (and adults) with contexts in which they had to judge a true but infelicitous *some*-statement (e.g., "Some of the blickets have a shovel" in a scene in which all of the blickets had a shovel). We ensured that these infelicitous statements were always preceded by a block of *all*-statements (i.e., the stronger scalar alternative was always made accessible). However, we manipulated the degree to which the scalar alternative was relevant to the conversational goal. In the *Object* condition, the first block that contained the stronger alternative all placed emphasis on object identity ("Do some/all of the blickets have X, or not?"). In the *Quantity* condition, the first, *all* block highlighted the quantity of blickets ("Do X of the blickets

have an item, or not?") and was therefore better aligned with the judgment for the infelicitous *some*-statements in the second block. More simply, any mention of '*all*' in a similar, minimally-different sentence like the one we used, will encourage participants to think that quantifiers form relevant alternatives. *False-All*-statements do this even more forcefully than *True-All*-statements, since they also clearly reveal the dimension across which a judgments needs to be made (Blicket Quantity vs. Object identity). But *False-All*-statements that suggest that other alternatives are relevant (e.g. object types) will encourage participants to think that objects form relevant alternatives and thus generate different results.

The findings are quite striking: Despite the fact that the stronger alternative (*all*) was present in exactly the same way in both conditions, children's performance with infelicitous *some*-statements was much lower in the *Object* compared to the *Quantity* condition. We take this as strong evidence for the rich hypothesis that states that scalar alternatives need to be viewed as relevant within a conversational goal in order to lead to SI generation. In fact, to the extent that they realized that the stronger alternative was relevant, children who had a strong grasp of the semantics of the quantifiers performed like adults (*Quantity* condition).

Chapter 4

EXPERIMENT 3

Experiment 2 established that children benefit from the accessibility of the stronger alternative only when the alternatives provided are relevant for the goal(s) of discourse. Experiment 3 tests the expectation (which is compatible with the rich hypothesis) that children can be shown to spontaneously generate SIs triggered by the scalar term *some* when the stronger alternative *all* is made relevant, even if the lexical term *all* is never mentioned. To achieve this we replicate the *Quantity* condition of Experiment 2 without the presence of the stronger scalar alternative *all*. Specifically, we replace *all*-trials in the first block of the experiment with corresponding *none*-trials. The rationale is that *none* being a quantifier should also help children identify relevant alternatives and in turn draw a SI upon encountering the weak scalar term *some*.

4.1 Method

4.1.1 Participants

We tested a new group of 60 typically developing 5-year-old children (4;6 – 5;8, M= 4;11) and 24 adult controls, all monolingual speakers of English. None of these participants had taken part in Experiments 1 or 2. The children were recruited from daycare centers in Newark, DE. The adults were college students recruited from the University of Delaware and received course credit for their participation. An additional group of 6 children were tested but excluded from the analysis for failure to follow instructions (n = 4), failure to complete the experiment (n = 1) and experimenter error (n = 1).

4.1.2 Materials and Procedure

Materials and procedure were almost identical to the *Quantity* condition of Experiment 2 but the *all*-statements in the first block were replaced by *none*-statements. The scenes for the first block were also changed to produce *True-None*-trials (where none of the blickets had the item) and *False-None*-trials (of two types, see below). Order of statements within blocks was pseudo-randomized so as to ensure alternations of trial types at least every three trials.

Participants were equally distributed across two between-subjects conditions. The two conditions differed in the scenes accompanying the *False-None*-statements in the first block: In the *Quantity-Full-Set* condition, a statement such as "None of the blickets have a scarf" would be false because all 4 blickets had a scarf; in the *Quantity-Partial-Set* condition, the same statement would be false because some (always 3 out of 4) blickets had a scarf. (Thus the scenes used for the *False-None*-trials in the *Quantity-Partial-Set* condition were identical to the scenes used for the *False-All*-trials in the *Quantity* condition of Experiment 2.) Examples of the *True-None* and *False-None* scenes and accompanying statements for each condition can be seen in figure 4 below.

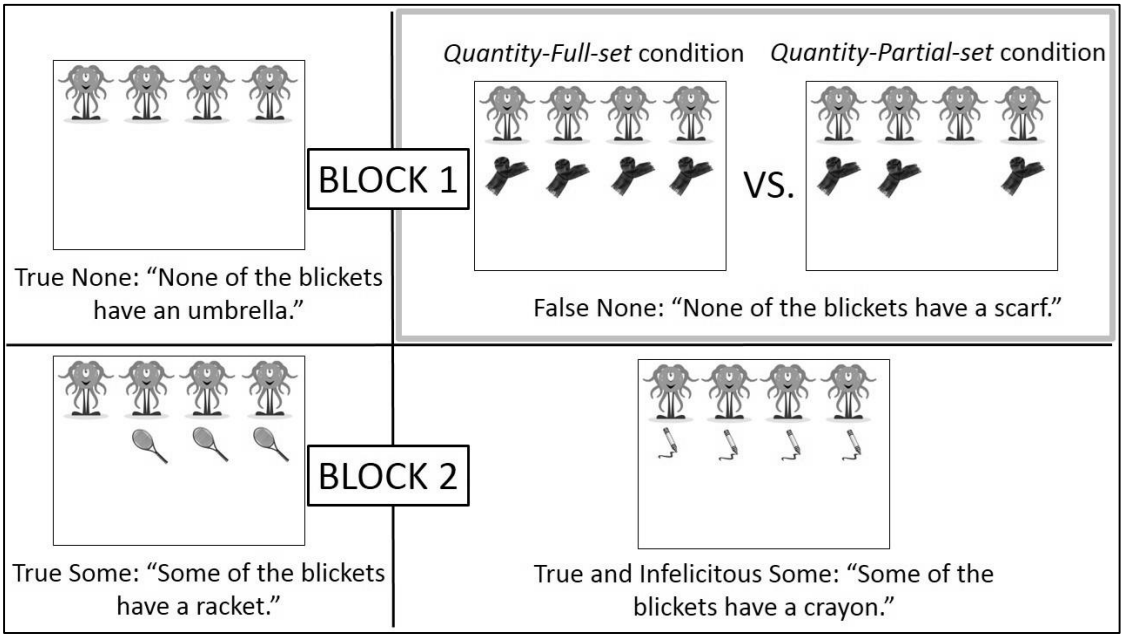

Figure 3. Types of trials for Experiment 3.

4.2 Predictions

If the nature of the *Quantity* task makes assessing the quantity of blickets relevant, then (in accordance with the rich hypothesis) children might be able to retrieve the stronger relevant scalar alternative *all* when faced with underinformative *some*-statements. That is, even if the stronger alternative is not lexically present in either of the two conditions in this study, children might be able to generate the relevant SI (*some* → *not all*) and reject the *True-and-Infelicitous-Some*-statements in the second block. In this case, the performance of children in rejecting the *True-and-Infelicitous-Some*-statements might be similar to the performance of children in the *Quantity* condition of Experiment 2. This would suggest that relevance drives (rather than simply mediates) the accessibility of the alternatives in SI implicature generation.

The two conditions of this study tested whether the implicit presence of *all* in the scenes described by the *False-None*-statements could help children by making the stronger alternative more accessible to them. In the *Quantity-Full-Set* condition, the *False-None*-trials were accompanied by scenes in which *all* was implicitly available (4 out of 4 blickets had the stated item). The *False-None*-statements in these trials were false exactly because all of the blickets had the item mentioned. In the *Quantity-Partial-Set* condition, however, *all* is completely absent from the scenes of the *False-None*-trials. If the performance of children in rejecting the *True-and-Infelicitous-Some*-statements is similar across the two conditions, the implicit presence of *all* in the *Quantity-Full-Set* condition could not have helped children. If, however, children were more successful in rejecting the *True-and-Infelicitous-Some*-statements in the *Quantity-Full-Set* condition rather than in the *Quantity-Partial-Set* condition, that would be evidence that even the implicit presence of *all* in the scenes described by the *False-None*-statements can help children by making the stronger alternative more accessible to them. The performance of adults is expected to be once close to ceiling levels, since their ability to spontaneously retrieve relevant stronger alternatives in the presence of weak scalar terms is not in doubt.

4.3 Coding

The coding scheme was identical to the one used for Experiments 1 and 2. Participants' correct scores were used to categorize them 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).

4.4 Results

Adult performance was consistently very high for all conditions and trial types. Table 7 summarizes adult performance. Fisher's Exact test analyses on 2x2 contingency tables for each trial type revealed no significant difference in the numbers of Passers vs. Failers across conditions (*True-None*-trials, $p = 1$; *False-None*-trials, $p = 1$; *True-Some*-trials, $p = 1$; *True-and-Infelicitous-Some*-trials, $p = 1$). Adults were overwhelmingly pragmatic in the *True-and-Infelicitous-Some*-trials.

Table 7. Adult performance in Experiment 3.

TRIAL TYPE		**CONDITION**	
		Quantity-Full-Set	**Quantity-Partial-Set**
True-None	Passers	12	12
	Failers	0	0
False-None	Passers	12	12
	Failers	0	0
True-Some	Passers	12	12
	Failers	0	0
True-and-Infelicitous-Some	Passers	10	11
	Failers	2	1

Children performed quite well with all of the trial types (see Table 8). Fisher's Exact Tests on 2x2 contingency tables revealed no significant difference in the numbers of Passers vs. Failers across the two conditions for any of the trial types: *True-None*-trials ($p = 1$), *False-None*-trials ($p = .748$), *True-Some*-trials ($p = .492$). More importantly in the *True-and-Infelicitous-Some*-trials, children appeared to be equally pragmatic in the *Quantity-Full-Set* condition (19 Passers and 11 Failers) and in *Quantity-Partial-Set* condition (19 Passers and 11 Failers), with no difference between

the numbers of Passers and Failers between the two conditions (Fisher's Exact test, two-tailed, $p = 1$).

As in previous experiments, when asked to justify their rejections of *True-and-Infelicitous-Some*-statements, children provided reasonable justifications. They typically referenced the stronger scalar term ("*All* of them have an X"; 32 out of 38 times, 84% of justifications), or mentioned the number of blickets with an item ("Because four blickets have an X" 5/38 times, 13% of justifications).

Table 8. Child performance in Experiment 3.

TRIAL TYPE		**CONDITION**	
		Quantity-Full-Set	**Quantity-Partial-Set**
True-None	Passers	27	27
	Failers	3	3
False-None	Passers	23	25
	Failers	7	5
True-Some	Passers	30	28
	Failers	0	2
True-and-Infelicitous-Some	Passers	20	19
	Failers	10	11

After this initial analysis and for the same reasons as in the previous experiments, we conducted a second analysis excluding children who were Failers in either the *True-None*, *False-None*, or *True-Some*-statements. This resulted in n = 10 children being excluded in the *Quantity-Full-Set* condition, and n = 9 children excluded in the *Quantity-Partial-Set* condition. All of these children can be assumed to have the correct semantics for *some* and *none*. A Fisher's Exact test on the 2x2 contingency table in Table 9 revealed no difference between the numbers of Passers

vs. Failers for the two different conditions (p =.697), confirming the results of the first analysis.

Table 9. *Some/none*-knowers' performance in *True-and-Infelicitous-Some*-trials of Experiment 3.

TRIAL TYPE		**CONDITION**	
		Quantity-Full-Set	**Quantity-Partial-Set**
True-and-Infelicitous-Some	Passers	17	16
	Failers	3	5

Following that, we compared the performance of the children who have a solid grasp of quantifier semantics with that of adults, with Fisher's Exact Test on 2x2 contingency tables. No differences were found between age groups in either the *Quantity-Full-Set* condition (p = 1) or the *Quantity-Partial-Set* condition (p = .379).

Finally, to complete our analysis of children's performance in the *True-and-Infelicitous-Some*-statements after the exclusions (Table 9), we ran Fisher's Exact Test on 2x2 contingency tables comparing the two conditions of Experiment 3 to the *Quantity* condition of Experiment 2. There was no difference in the numbers of Passers and Failers between the *Quantity* (17 Passers, 0 Failers) and the *Quantity-Full-Set* condition (17 Passers, 3 Failers; p = .234). A difference approaching significance was found in the numbers of Passers and Failers between the *Quantity* and the *Quantity-Partial-Set* condition (16 Passers, 5 Failers, p = .053), with the *Quantity* condition having marginally more Passers and fewer Failers than the *Quantity-Partial-Set* condition.

4.5 Discussion

Experiment 3 set out to push the boundaries of the rich hypothesis by exploring the possibility that accessibility of alternatives is mediated by whether such alternatives are relevant in furthering the goals of conversation. To that end, we modified our experimental paradigm, so that the stronger scalar alternative *all* would be lexically absent and instead replaced it with the quantifier *none*. The rationale was that, *none* being a quantifier and encoding the quantity of blickets possessing an item, it should also help children identify relevant alternatives and in turn draw a SI upon encountering the weak scalar term *some*.

The results are striking: In both the *Quantity-Full-Set* and the *Quantity-Partial-Set* condition children were overwhelmingly pragmatic, generated the relevant SI and rejected the *True-and-Infelicitous-Some*-statements even though the stronger scalar alternative *all* was lexically absent throughout the experiment. In fact, after controlling for knowledge of semantics, children in both conditions were pragmatic at levels comparable to adults. Thus even if *all* is not explicitly (or implicitly; see *Quantity-Partial-Set* condition) present, children can be led by a manipulation that establishes the relevance of alternatives to spontaneously generate *all* and generate a SI in this experiment.

Chapter 5

GENERAL DISCUSSION

5.1 Goals and Summary of Findings

This series of studies was motivated by the sometimes paradoxical picture in the field of developmental studies regarding the development of pragmatic inference: Preschool-age children have been repeatedly shown to perform poorly on pragmatic inferences related to language (Markman & Seibert, 1976; Vosniadou, 1986; Trueswell, Sekerina, Hill & Logrip, 1999; see Papafragou, 1998; 2006 for discussion), while at the same time, there is evidence that their pragmatic abilities are well-developed by the age of 2 in other, non-linguistic tasks (Corkum & Moore, 1998; Tomasello, 1995; Behne, Carpenter & Tomasello, 2005; Lizskowski, Carpenter & Tomasello, 2008; see Tomasello, Carpenter, Call, Behne & Moll, 2005 for discussion). In an effort to investigate children's pragmatic abilities in inferences closely related to language, we used scalar implicature (SI) as a case study, since it has been shown to be particularly suited to an empirical investigation of linguistic pragmatics. Moreover, there is a considerable, if fragmented, body of work developed in the last two decades that shows children's failures and successes in scalar inferences within different paradigms.

We identified two major factors that have been claimed in the past to bear part of the responsibility for children's difficulties with SIs and that have been explored independently: conversational goals (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004) and the accessibility of scalar alternatives (Gualmini et al., 2001; Barner et al., 2011, Bale & Barner 2013; Papafragou & Skordos, in press).

We systematically tested two open hypotheses about how alternatives might (or might not) be related to conversational goals and relevance. On the lean hypothesis, use of the alternatives (and SI calculation) is not mediated by relevance/conversational goals in children. On the rich hypothesis, conversational relevance mediates use of alternatives and SI calculation. We used an AJT paradigm and independently manipulated a) the lexical accessibility of a quantifier scale (<*some, all*>) and b) the relevance of scalar alternatives. In Experiment 1 we found that the accessibility of the stronger scalar term *all* indeed seemed to facilitate SI generation in children. However, the factors that made alternatives accessible in Experiment 1 also made them relevant. In Experiment 2 we independently manipulated the relevance of scalar alternatives, while keeping the stronger alternative lexically accessible. We found that relevance of the alternatives mediates their use for SI generation: Even when alternatives were lexically accessible, children generated the corresponding SI only when the alternatives were relevant to the goals of discourse. Finally in Experiment 3, we investigated the potency of relevance by completely removing the presence of the strong scalar alternative *all* from our paradigm. We found that even when we did not provide the strong scalar alternative to children, they were consistently able to generate the SI at adult-like levels based on the relevance of the alternatives.

These results provide evidence that relevance does not simply mediate but in fact drives the accessibility of the scalar alternatives. Our results thus strongly support the rich hypothesis: the problem that young children face with SIs should not be sought in their knowledge of scales and appropriate scalar alternatives (Barner et al., 2011; Bale & Barner, 2013) but rather in limitations to their ability to quickly and

efficiently identify conversational goals and conversational relevance. In addition, the findings of Experiment 3 lend further credence to the idea that at least part of the failures we saw in Experiment 1 were not due to the fact that the stronger alternative was not made lexically accessible for children; instead at least some of the children who failed to generate the appropriate SI probably did so because they failed to identify the alternatives as being relevant. To our knowledge, these studies are the first in the literature to empirically show that the mechanism through which alternatives facilitate SI generation in children cannot be simply reduced to a non-pragmatic, lexically-driven contrast effect (lean hypothesis).

Unsurprisingly, adults did not seem to be affected by the present experimental manipulations, presumably because our task was easy and transparent enough for them to quickly adjust their expectations about relevance when that was necessary (for instance going from the first into the second block of the *Object* condition in Experiment 2). This lends further support to the idea that the developmental difference between children and adults in SI-calculations might be that children sometimes find it more difficult to assess relevance in discourse as compared to adults.

Recall the standard Gricean SI computation schema we presented earlier:

I. In (2a), the speaker has violated the Quantity maxim by using a relatively weak term from a set ordered according to informational strength (< *some, …, all* >)

II. The speaker is expected to say as much as he/she truthfully can, while observing the Relevance maxim, i.e. in a way *relevant* to the communicative exchange. An utterance can be thought of as 'relevant'

if it can be understood as contributing to the goals of the communicative exchange (Leech, 1983).

III. The choice of the weaker term is reason to believe that the speaker is not able to commit to an informationally stronger statement ("Megan ate all of the cupcakes.").

IV. Therefore, as far as the speaker is able to say, the stronger statement does not hold, thus (2b).

If so, children's problems are very likely to correspond to step II of the standard Gricean computation, regarding the relevance of a contribution to the conversational goals. In other words, children have difficulty in homing in to the conversational goal and consequently inferring expectations of informativeness and relevance in a given situation. This in return causes difficulties in evaluating a linguistic stimulus with respect to possible alternatives that the speaker could have selected (cf. Papafragou & Tantalou, 2004; Papafragou, 2006), as well as difficulties in accessing the relevant alternatives themselves. These difficulties result in children's well attested failure to generate a SI even when one is expected. Development leads to changes in children's ability to locate and incorporate useful contextual information that allows one to efficiently identify the goals of a communicative exchange. The result is that they become fully adult-like in that they draw pragmatic inferences (and SIs) when those appear to further the goals of discourse.

This is consistent with studies that have shown that relevance implicatures (example (1) in the introduction section), a kind of pragmatic inference based predominantly on conversational relevance, pose significant difficulties for children up to the age of 6 (Bucciarelli et al., 2003; de Villiers, de Villiers, Coles-White, &

Carpenter, 2009; Loukusa, Leinonen, & Ryder, 2007; Loukusa, Ryder & Leinonen, 2008; Verbuk & Shultz, 2010). Interestingly, in an observation that seems to parallel our own, when the tasks employed are simplified in order to reduce the burden on other cognitive systems, children's performance improves (de Villiers et al., 2009). More strikingly, when the relevance inference to be made is further restricted and tested with an act-out task that does not require a verbal response, even 3-year-olds show some evidence of being able to assess conversational relevance and make the appropriate inferences, although their performance still falls rather short of being adult-like (Shulze, Grassman & Tomasello, 2013).

It is quite possible that the results of previous studies that have shown improved SI generation can be re-interpreted using our approach, based on the role of relevance-driven accessibility of alternatives. For instance, in studies that provided an under-informative and a fully informative alternative and asked children to choose between the two, both alternatives offered were relevant (Chierchia et al., 2001, Gualmini et al., 2001). In other work, contextual support in the form of background information essentially drew attention to a contextually relevant stronger alternative (Papafragou & Musolino, 2003; Guasti et al., 2005).

Our data are also consistent with prior studies showing that the alternation of *some*- and *all*-statements leads to successful SI generation by children (Bagassi et al., 2009; Foppolo, Guasti & Chierchia, 2012). In these studies, however, as well as in most previous studies (Gualmini et al., 2001; Chierchia et al., 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Pouscoulous et al., 2007; Katsos & Bishop, 2011) the role of communicative goals and scalar alternatives was conflated, in a way similar to our Experiment 1. For instance, Foppolo and colleagues (2012)

presented children with stories that showed characters acting on items (e.g. a dwarf picking up carrots). Children were presented with a succession of 3 stories described by a *True-All*, *False-All* and an *Infelicitous-Some*-statement respectively and had to evaluate the description (statement) and "correct it if it was wrong". Moreover, children's attention was drawn to quantity as the relevant attribute for judging the statements as they were prompted during each story with a corresponding statement (e.g. "Look! This dwarf has a lot of carrots! Let's see *how many* he will pick in the end."). Since alternatives were made *both* relevant *and* accessible, unsurprisingly, 5-year-olds were quite good at rejecting the infelicitous *some*-statements (see also Papafragou & Tantalou, 2004). Consequently, relevance, the factor that lies at the heart of children's difficulty in scalar inference, was never systematically tested until Experiments 2 and 3 of this study.

Some prior studies also used an alternation of *some*- and *all*-statements in tasks similar to our own (e.g. Noveck, 2001; but cf. Feeney, Scrafton, Duckworth, & Handley, 2004 for somewhat different results) and failed to show that the stronger alternative benefited SI generation. In our opinion, those studies were more open ended, and might have not have made clear that stronger alternatives for each statement would be relevant. For example, when children are asked to judge a statement like "Some giraffes have long necks" that is preceded in the experiment by a statement such as "All chairs tell time", it is not immediately clear whether the communicative partner (experimenter in this case) is focusing on truth conditions (in which case no alternatives need be considered), or pragmatic felicity (in which case the newly heard statement needs to be considered with respect to other things the speaker could have said but did not). In addition, the statements used in those studies

typically had to be evaluated against encyclopedic knowledge, something that might have caused additional complications (e.g. difficulty to appropriately restrict the universe of discourse). Our paradigm was clearly focused on quantification and provided all the evidence necessary to evaluate the statement in front of the child's eyes.

5.2 Relevance, Conversational Goals and the QUD

The present data provide strong evidence that the idea of the accessibility of alternatives needs to be considered within context-driven accounts of SIs (Grice, 1975; Sperber & Wilson, 1985/1996; Carston, 1995; Noveck & Sperber, 2007). As we can see from Experiment 2, the stronger alternative all did not seem to lead children into SI generation unless that alternative could be viewed as relevant to the goal of discourse, despite being always present and in principle available for the children. In order for alternatives to be useful in SI generation, they need to be considered within the goals of the communicative exchange at hand (in our case, the goal of the experimental task), and they need to be seen as "relevant" alternatives to what is being said by the speaker. Our notion of relevance in this case can be made more concrete if we think of the conversational goal as a formalization of "relevance" regarding scalar inference: A stronger alternative to a weak scalar term is accessible and leads to SI generation if and only if the scalar set/subset relationship that the weak and strong alternatives enter into provides an answer that can be seen as furthering the goals of the conversation. In essence, the above formulation is what Grice (1967) would call "what is relevant" (see also Leech, 1983) and Relevance theorists would call "the expected cognitive effects" of the conversational exchange; see Sperber and Wilson (1986/1995) and Breheny, Ferguson and Katsos (2013) for related discussion.

This possibility might be linked to what has been described in the past as the "question under discussion" (QUD; Roberts, 1996; 2004). According to QUD accounts of pragmatics (Stalnaker, 1979; Roberts, 1996; 2004), discourse is based on conversational goals, foremost among which is an attempt by the communicative partners to discover the state of affairs that obtains with regard to their topic of conversation (see Hulsey, Hacquard, Fox, & Gualmini, 2004; Conroy, Lidz & Musolino, 2009 and references therein). In their attempt to do so, communicative partners posit and answer a series of explicit and implicit questions relevant to the aforementioned topic. An utterance is considered relevant[6] to the QUD if it provides an answer (full or partial) to it (Roberts, 1996; 2004; van Rooij & Schulz, 2004). The QUD can be seen as a formalization of the notion of relevance, and might have explanatory potential in general pragmatic inference. To appreciate this, we need to consider the conditions under which SIs typically arise: Intuitively, SIs are supposed to be derived only when the stronger scalar alternative is relevant within the contextual background (Carston, 1998; Levinson, 2000). Those intuitions are borne out by experimental studies demonstrating that adults are much more likely to generate SIs when the stronger scalar alternative is contextually relevant (Zondervan, 2010; Degen, 2013). Keeping this in mind, the somewhat vague notion of relevance can perhaps be made more explicit if we posit that the stronger scalar alternative is relevant only if it provides an answer to the QUD.

---

[6] The notion of relevance to the QUD used here is categorical (relevant vs. irrelevant). For a theoretical approach that uses a gradient notion of relevance, see Carnap (1950) and Russell (2012); for empirical studies with adults, applying such an approach to SI generation see Degen (2013).

For instance, in the *Quantity* condition of Experiment 2, the implicit QUD that is set up in the first block of the experiment through the *False-All*-trials concerns the quantity of blickets that possess an item:

(6) Implicit QUD: Do *all* of the blickets have a crayon?

"Some of the blickets have a crayon."

(SI: Some, but not all of the blickets have a crayon.)

This QUD remained stable throughout the trial battery and children were able to use it (and the stronger scalar alternative that it made relevant) to generate the SI. In the *Object* condition however, the QUD set up implicitly in the first block concerns the identity of objects that the blickets possess:

(7) Implicit QUD: What kind of item do the blickets have?

"Some of the blickets have a crayon."

(No SI).

Adults were able to shift to another QUD in the second block of trial battery (Do all of the blickets have a crayon?) and thus generate the SI. Children however were unable to accommodate this shift to a different QUD in the second block of the *Object* condition. As a consequence, they interpreted the statement "Some of the blickets have a crayon", as an answer to the original QUD they had in mind (What kind of item do the blickets have?). The stronger scalar alternative (*all*) was consequently not relevant when all blickets were shown to indeed have a crayon (and not a different item), and thus children were not able to generate the SI.

It is worth noting here that the use of the QUD as a formalization of relevance has some caveats: First of all, it is not currently clear whether the QUD and the relevance of the alternatives can be subsumed under the same cognitive mechanism.

Zondervan (2010) presents a detailed exploration of the effects of focus and exhaustivity (see also van Kuppevelt, 1996; van Rooij 2003; van Rooij & Shulz, 2004) on one hand, which seem to be directly linked to the QUD, and the relevance of scalar alternatives on the other as an independent factor. Zondervan's results confirm that a combination of both focus and relevance of the alternatives seems to increase SI generation in adults, however the findings of experiments attempting to tease apart the independent contribution of either of the two factors were less clear. In addition, Zondervan finds some potential evidence that exhaustivity as defined in van Rooij and Shulz (2004) and SI generation might not be reducible to a single cognitive mechanism; if this argument is confirmed, it casts further doubt onto QUD-based accounts of relevance for SIs. Second, as Russell (2012) points out, virtually all formal theories of SI generation, including the ones linked to the QUD (Gazdar, 1979; Groenendijk & Stokhof, 1984; Sauerland, 2004; van Rooij & Shulz, 2004), that attempt to formalize relevance, do not seem to be motivated by Gricean reasoning: rather they appear to be grammatically-motivated accounts that use formal objects (such as exhaustivity operators) to reason about logical propositions (and not utterances) in order to formally describe our intuitions about pragmatic inference (and SIs). More importantly perhaps, they have to rely on logical scales, and it is not clear how they can apply to ad-hoc scales.

5.3 Alternatives in Pragmatics and Lexical Learning

There seem to be intriguing parallels to our discussion regarding SI from the field of lexical learning. Young children have a well-developed ability to consider contrastive alternatives and this ability appears to be active from early on in language acquisition (cf. Papafragou & Tantalou, 2004). As we mentioned in the introduction,

2-year-olds have been shown to take advantage of the fact that an adult used a novel word (e.g., *dax*) rather than a known word (e.g., *car*) in an environment featuring a car and a novel, unlabeled object, to conclude that the novel word refers to the novel, unlabeled object (Carey, 1978; Halberda, 2006). It appears that, for young children, word meanings (and especially nouns used as labels) seem to be mutually exclusive (Clark, 1987, 1988; Markman, 1989). This assumption of mutual exclusivity or contrast does not seem to extend to terms that belong to different languages, or are used in different levels of description (*dog, poodle*; Au & Glusman, 1990; Diesendruck, 2005). A possible reason for this could be that these terms cannot be considered as appropriate alternatives for each other (Barner et al., 2011).

Such inferences in word learning have important differences from SIs. However, it is not implausible that (like SIs) they might require considering speaker goals and intentions within discourse and require accessing lexical alternatives to a term used. Based on these similarities to SIs, it has been suggested that they too might be Gricean in nature (Clark, 1990; Diesendruck & Markson, 2001; Gathercole, 1989; but see also Regier, 2003; Frank et al., 2009; de Marchena, et al., 2011, for a different perspective). However, recent research with ASD children (who are generally thought to have serious difficulties in most tasks involving pragmatic inference) shows that they seem to perform well with mutual exclusivity and word learning, indicating that perhaps contrast in these cases is not linked to pragmatic considerations (de Marchena, et al., 2011). Further research is necessary in order to provide convincing answers to the question whether lexical contrast and the evaluation of alternatives is based on the same underlying cognitive mechanism in word-learning and SI generation or not.

Progress on these topics has important ramifications for the linguistic and cognitive development of children.

5.4 SIs and Pragmatic Development

We would like to close our discussion by mentioning the broader implications of our findings for pragmatic development. Our results continue a strand of research showing that the mechanisms necessary for pragmatic inference in language are clearly in place for preschoolers (Chierchia et al., 2001; Gualmini et al., 2001, Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Feeney et al, 2004; Guasti et al., 2005; Papafragou, 2006; Pouscoulous et al., 2007; Bagassi et., al 2009; Katsos & Bishop, 2011; Barner et all, 2011; Foppolo et al., 2012). Therefore any attempt to explain the apparent pragmatic discontinuity between children and adult communicators in terms of an absence in children of the necessary pragmatic mechanisms should by now be abandoned.

Nevertheless, there still appear to be marked differences in the conditions under which children and adults are able to show their pragmatic sophistication. Our results strongly support the idea that those differences are due to performance factors linked directly to a communicator's ability to efficiently exploit contextual cues in order to interpret what is relevant in a communicative situation in any given time. Our studies isolate and manipulate the contribution of conversational relevance in such situations, by stripping away several layers of potential variability and drastically restricting the pragmatic inference task to a binary choice about what is relevant: The quantity of the blickets, or the type of the objects? This drastic restriction to minimally different sentences, differing only in what is crucial (and thus relevant) for judging them, arguably lessens the demands the pragmatic inferential mechanism imposes on

executive function systems that are of more limited capacity in children versus adults (Luciana & Nelson, 2002; De Luca, Wood, Anderson, Buchanan, … , Pantelis, 2003; De Luca & Leventer, 2008; Anderson, Jacobs & Anderson, 2008).

Pragmatic theory and the study of pragmatic development need to examine performance factors far more closely than in the past: How exactly do measures of executive function interact with pragmatic inference? Can controlling for these variables (through independent measures) account for the difference between the performance of children and adult communicators, or at least for part of them? Do children who perform better at pragmatic inference tasks, also perform better in language acquisition and are the language skills that drive the pragmatic performance or vice versa? Could there be a third unexplored factor that underlies both? Longitudinal studies and studies of special populations, as well as experimenting with working memory load and inhibitory control seem particularly promising directions for research.

In addition, the notion of conversational relevance (Grice, 1975) needs to be explored further and be provided with an operational definition that allows for experimental research. It has been proposed (Degen, 2013) that the QUD might provide such a formalization of relevance. However, other options are also open: A relevance theoretic approach (Sperber & Wilson, 1986/1995) where relevance is seen as an optimization process between the projected cognitive effects of pragmatic inference and the cognitive load it imposes on the computational system, is certainly possible if one manages the hard task of reducing what is meant by "cognitive effect" and "cognitive load" to step by step computations that can be shown to be psychologically real, measurable and able to be systematically manipulated. The

ultimate goal of such a research program would be an empirically testable and psychologically plausible theory of pragmatics that would have explanatory adequacy and would be able to predict the whole range of pragmatic behavior. Attempts to arrive to a computational model of pragmatic behavior through Bayesian probability inference are already meeting with some success (Frank & Goodman, 2012; Russell, 2012).

REFERENCES

Anderson, V., Jacobs, R., & Anderson, P. J. (Eds.). (2011). *Executive functions and the frontal lobes: A lifespan perspective*. Psychology Press.

Aslin, A. (2000). Why take the cog out of infant cognition? *Infancy, I(4)*, 463-470.

Bagassi, M., D'Addario, M., Macchi, L., & Sala, V. (2009). Children's acceptance of underinformative sentences: The case of some as a determiner. *Thinking and reasoning, 15(2)*, 211-235.

Bale, A. & Barner, D. (2013). Grammatical alternatives and pragmatic development. In Anamaria Falaus (Ed.), *Alternatives in Semantics, 'Studies in Pragmatics, Language and Cognition'*. (pp. 238-266). Palgrave Macmillan.

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

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

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8:6*, 492-499.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5)*, 1450.

Bloom, P. & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25-B31.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language, 51(3)*, 437-457.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66(1)*, 123-142.

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

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: toward a model of on-line access to conversational implicatures. *Cognition, 126(3)*, 423-440.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100(3)*, 434-463.

Brooks, N., Audet, J., & Barner, D. (2013). Pragmatic inference, not semantic competence, guides 3-year-olds' interpretation of unknown number words. *Developmental Psychology, 49(6)*, 1066-1075.

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics, 35*, 207–241.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*, 337-342.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller and M. Halle (eds.), *Linguistic Theory and Psychological Reality*, (pp.264-293). Cambridge, MA: MIT Press.

Carlson, S. & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72(4)*, 1032-1053.

Carlson, S., Moses, L., & Claxton, L. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299-319.

Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding "prior intentions" enables two–year–olds to imitatively learn a complex task. *Child development, 73(5)*, 1431-1441.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63(4)*, (pp.1-176). Wiley.

Carston, R. (1995). Quantity maxims and generalized implicature. *Lingua, 96*, 213-244.

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.

Chierchia, G., Fox, D., & Spector, B. (2009). Hurford's Constraint and the Theory of Scalar Implicatures. In: P. Egré & G. Magri (Eds.), *Presuppositions and Implicatures,* (pp. 47–62). Proceedings of the MIT-Paris Workshop. Cambridge, MA: MIT Working Papers in Linguistics.

Chierchia, Gennaro, (2004). Scalar implicatures, polarity phenomena and the syntax / pragmatics interface. In: Belletti, A. (Ed.), *Structures and Beyond*, (pp. 39-103). Oxford University Press, Oxford.

Chomsky, N. (1964) *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.

Clark, E., V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, E., V. (1988). On the logic of contrast. *Journal of Child Language, 15*, 317–335.

Clark, E., V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431.

Conroy, A., Lidz, J., and Musolino, J. (2009) 'The fleeting Isomorphism Effect', *Language Acquisition, 16(2)*, 106-117.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental psychology, 34(1)*, 28.

Crain, S., & McKee, C. (1985). *The acquisition of structural restrictions on anaphora.* Proceedings of NELS, 15, 94-110.

De Luca, C. R., & Leventer, R. J. (2008). Developmental trajectories of executive functions across the lifespan. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 23-56). New York: Psychology Press.

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the CANTAB. I: development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology, 25(2)*, 242-254.

de Marchena, A., Eigsti, I. M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition, 119(1)*, 96-113.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie), 54(2)*, 128-133.

de Villiers, P. A., de Villiers, J., Coles-White, D. J., & Carpenter, L. (2009). Acquisition of relevance implicatures in typically-developing children and children with autism. In J. Chandlee, M. Franchini, S. Lord, & G. M. Rheiner (Eds.), *Proceedings of the 33th Annual Boston University Conference on Language Development* (pp. 121–132). Somerville, MA: Cascadilla Press.

Degen, J. (2013) *Alternatives in Pragmatic Reasoning*. Doctoral Dissertation, Rochester, NY.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology, 41*, 451–463.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. (2004). The story of some: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology, 58:2*, 121-132.

Foppolo, F., Guasti, M.T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development, 8*, 365-394. Psychology Press.

Fox, D. (2007). Free Choice Disjunction and the Theory of Scalar Implicatures. In: U. Sauerland and P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics.* New York: Palgrave Macmillan.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science, 336*(6084), 998-998.

Frank, M., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20(5)*, 578–585.

Gathercole, V. (1989). Contrast: A semantic constraint? *Journal of Child Language, 16*, 685–702.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form.* New York: Academic Press.

Gelman, R. & Greeno, J.G. (1989). On the nature of competence: Principles for understanding domain. In L.B. Resnick (Ed.), *Knowing and learning: Essays in honor of Robert Glaser* (pp. 125-186), Hillsdale, NJ: Erlbaum Associates.

Geurts, B. (1998). Scalars. In P. Ludewig and B. Geurts (Eds.), *Lexicalische Semantik aus kognitiver Sicht*, (pp. 95-117). Tübingen: Narr.

Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.

Gopnick, A. & Astington, J.W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance reality distinction. *Child Development, 59(1)*, 26-37. Wiley.

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

Groenendijk, J. and Stokhof, M., 1984, *Studies in the semantics of questions and the pragmatics of answers*, Ph.D. Thesis, University of Amsterdam.

Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology, 46(6)*, 1710-1722.

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.

Haithe, M.M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development, 21(2)*, 167-179.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 53*, 310-344.

Harnish, R. (1976). Logical form and implicature. In T., G. Bever, J., J. Katz & T. Langdon (Eds.), *An Integrated Theory of Linguistic Ability*, (pp. 313–392). New York: Thomas Y. Crowell. Reprinted in S. Davis, ed. (1991), *Pragmatics: A Reader*, (pp. 316–364). Oxford: Oxford University Press.

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

Horn, L. R. (2006). *The border wars: A neo-Gricean perspective. Where Semantics Meets Pragmatics: The Michigan Papers*, (pp. 21-48). Oxford: Elsevier.

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

Horn, L., R. (1989). *A natural history of negation*. Chicago: University of Chicago Press. Repr. in 2001. Stanford: CSLI Publications.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology, 58(3)*, 376-415.

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

Hulsey, S., Hacquard, V., Fox, D., & Gualmini, A. (2004). The question answer requirement and scope assignment. *In Plato's problems: Papers on language acquisition. MIT Working Papers in Linguistics 48*. Cambridge, MA: MITWPL.

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development, 2(2)*, 77-96.

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

Katsos, N., & Cummins, C. (2010). Pragmatics: from theory to experiment and back again. *Language and Linguistics Compass, 4(5)*, 282-295.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy, 30*, 669–690.

Kempson, R. (1988): Grammar and conversational principles. In Newmeyer, F. (Ed.): *Linguistics: The Cambridge Survey*, (Vol. II. pp. 139-163). Cambridge, Eng.: Cambridge University Press.

Kidd, C., White, K.S., & Aslin, R.N. (2009). Children's use of disfluencies for pragmatic inference in lexical development. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, 1494-1499.

Kidd, E. (2004). Grammars, parsers, and language acquisition. *Journal of Child Language, 31(2)*, 480-483.

Leech, G. (1983). *Principles of Pragmatics*. London: Longman.

Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Liszkowski, U., Carpenter, M, & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*, 732-739.

Liszkowski, U., Schafer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science, 20(5)*, 654-660.

Loukusa, S., Leinonen, E., & Ryder, N. (2007). Development of pragmatic language comprehension in Finnish speaking children. *First Language, 27*, 279–296.

Loukusa, S., Ryder, N., & Leinonen, E. (2008). Answering questions and explaining answers: A study of Finnish-speaking children. *Journal of Psycholinguistic Research 37(3)*, 219-241.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4-to 12-year-old children. *Developmental neuropsychology, 22(3)*, 595-624.

Markman, E. M., & Seibert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology, 8(4)*, 561-577.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157.

Markman, E., M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

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

Noveck, I., A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of 'scalar inferences'. In N. Burton-Roberts (Ed.) *Advances in Pragmatics*. Basingstoke: Palgrave.

Onishi, K.H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258.

Ozturk, O., & Papafragou, A. (to appear). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language, Learning and Development*.

Papafragou, A. & Skordos, D. (in press). Scalar Implicature. In J. Lidz, W. Snyder and J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics*. Oxford: Oxford University Press.

Papafragou, A. (1998). The acquisition of modality: Implications for theories of semantic representation. *Mind & Language, 13(3)*, 370-399. Blackwell Publishers Ltd.

Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language, 33*, 721-757.

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

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

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep. *Science, 308*(5719), 214-216.

Pouscoulous, N., Noveck, I., Politzer, G., Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition, 14*, 347-375.

Regier, T. (2003). Emergent constraints on word-learning: A computational perspective. *Trends in Cognitive Science, 7(6)*, 263–268.

Roberts, C. (1996). Information structure: towards an integrated theory of formal pragmatics. *OSU Working Papers in Linguistics 49, Papers in Semantics*, (pp. 91—136).

Roberts, C. (2004). Information structure in discourse. *Semantics and Pragmatics*, 5, 1 – 69.

Russell, B. (2012). *Probabilistic Reasoning and the Computation of Scalar Implicatures*. Unpublished doctoral dissertation, Brown University.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367–91.

Shatz, M. (1983). Communication. In P. H. Mussen (Series Ed.) & J. Flavell & E. Markman (Vol. eds.), *Handbook of Child Psychology*: Vol. 3. Cognitive development (4th ed.), (pp. 841-889). New York: Wiley.

Shatz, M., & O'Reilly, A. W. (1990). Conversational or communicative skill? A reassessment of two-year-olds' behaviour in miscommunication episodes. *Journal of Child Language, 17(1)*, 131-146.

Sirois, S., & Jackson, I. (2007). Social cognition in infancy: A critical review of research on higher order abilities. *European Journal of Developmental Psychology, 4(1)*, 46-64.

Smith, C., L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology, 30*, 191–205.

Spector, B. (2007). Aspects of the pragmatics of plural morphology: On higher-order implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics*, (pp. 71–120). New York: Palgrave Macmillan.

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

Stalnaker, R. (1972). Pragmatics. In G. Harman and D. Davidson (Eds.): *Semantics of Natural Language* (pp. 380-397). Dordrecht: Reidel.

Stalnaker, R. 1979. Assertion. In: Cole, P. (Ed.), *Syntax and Semantics 9*, (pp. 315-332). Academic Press, New York.

Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicature in adults and children. *In Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, Boston, MA.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18(7)*, 580-586.

Tavano, E., & Kaiser, E. (2010). *Processing scalar implicature: What can individual differences tell us?* University of Pennsylvania Working Papers in Linguistics: Vol. 16: Iss. 1, Article 24.

Trueswell, J. C., Sekerina, I., Hill, N.M., & Logrip, M.L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

van Kuppevelt, J. (1996). Inferring from topics. *Linguistics and Philosophy, 19(4)*, 393–443.

van Rooij, R.(2003). Questioning to Resolve Decision Problems, *Linguistics and Philosophy 26*, 727–763.

van Rooij, R. & K. Schulz (2004). Exhaustive interpretation of complex sentences. Journal of Logic, *Language and Information 13(4)*: 491-519.

Verbuk, A. (2012). Developmental evidence against the theoretical distinction between Horn and pragmatic scales. *Journal of Pragmatics, 44*, 1680-1700.

Verbuk, A., & Shultz, T. (2010). Acquisition of relevance implicatures: A case against a rationality-based account of conversational implicatures. *Journal of Pragmatics, 42*, 2297–2313.

Vosniadou, S. (1986). Children and Metaphors. *Center for the Study of Reading: Technical Reports at University of Illinois at Urbana-Champaign, 370*, 1-64. National Institute of Education.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development, 72(3)*, 655-684.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Woodward, A., & Markman, E. (1998). Early word learning. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. Cognition perception and language*, (pp. 371–20). Hoboken, NJ: John Wiley & Sons, Inc.

Zondervan, A. (2010). *Scalar implicatures or focus: an experimental approach*. Unpublished doctoral dissertation, Universiteit Utrecht, Amsterdam.

Appendix

UNIVERSITY OF DELAWARE IRB APPROVAL FOR RESEARCH ON HUMAN SUBJECTS

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: March 2, 2012

TO: Dimitrios Skordos, MA
FROM: University of Delaware IRB

STUDY TITLE: [312739-1] Language Acquisition: Word learning and pragmatic inference

SUBMISSION TYPE: New Project

ACTION: APPROVED
APPROVAL DATE: March 2, 2012
EXPIRATION DATE: March 1, 2013
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of New Project materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Jody-Lynn Berg at (302) 831-1119 or jlberg@udel.edu. Please include your study title and reference number in all correspondence with this office.

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: October 10, 2012

TO: Dimitrios Skordos, MA
FROM: University of Delaware IRB

STUDY TITLE: [312739-2] Language Acquisition: Word learning and pragmatic inference

SUBMISSION TYPE: Amendment/Modification

ACTION: APPROVED
APPROVAL DATE: October 10, 2012
EXPIRATION DATE: March 1, 2013
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of Amendment/Modification materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Jody-Lynn Berg at (302) 831-1119 or jlberg@udel.edu. Please include your study title and reference number in all correspondence with this office.

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: November 2, 2012

TO: Dimitrios Skordos, MA
FROM: University of Delaware IRB

STUDY TITLE: [312739-3] Language Acquisition: Word learning and pragmatic inference

SUBMISSION TYPE: Amendment/Modification

ACTION: APPROVED
APPROVAL DATE: November 2, 2012
EXPIRATION DATE: March 1, 2013
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of Amendment/Modification materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Jody-Lynn Berg at (302) 831-1119 or jlberg@udel.edu. Please include your study title and reference number in all correspondence with this office.

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: February 20, 2013

TO: Dimitrios Skordos, MA
FROM: University of Delaware IRB

STUDY TITLE: [312739-4] Language Acquisition: Word learning and pragmatic inference

SUBMISSION TYPE: Continuing Review/Progress Report

ACTION: APPROVED
APPROVAL DATE: February 20, 2013
EXPIRATION DATE: March 1, 2014
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Jody-Lynn Berg at (302) 831-1119 or jlberg@udel.edu. Please include your study title and reference number in all correspondence with this office.