Grammatical alternatives and pragmatic development

Alan Bale\textsuperscript{1,2} and David Barner\textsuperscript{3}

\textsuperscript{1}Department of Linguistics, Concordia University, \textsuperscript{2}Department of Linguistics, McGill University

\textsuperscript{3}Department of Psychology, University of California, San Diego

Contact Information:

Alan Bale
Department of Linguistics
Concordia University,
H663, 1455 de Maisonneuve Blvd. W.
Montreal, Quebec (H3G 1M8)
alancbale@gmail.com

David Barner
Department of Psychology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0109
barner@ucsd.edu
Acknowledgements

We would like to thank Jessica Sullivan for her thoughtful and valuable comments on previous versions of this paper. We would also like to thank Francesca Foppolo and Danielle Panizza for their comments and suggestions during the review process. This work would not have been possible without a standard research grant to Alan Bale from Social Sciences and Humanities Research Council of Canada (SSHRC) (“Language, Number and Mathematics”) and a grant to David Barner from the James S. McDonnell Foundation.
1. Introduction

Language acquisition offers a unique window into linguistic competence. As children acquire language, different aspects of competence, like knowledge of syntax, semantics, and pragmatics, emerge at different moments, allowing them to be logically dissociated. As a result, developmental data can be used to decide between competing linguistic models that posit different structures, but nonetheless make similar predictions regarding mature competence. It is this premise that has guided a recent surge in the study of pragmatic development. If children are delayed in their ability to make pragmatic inferences then researchers might not only distinguish semantic and pragmatic sources of meaning, but they might also be able to decompose pragmatic reasoning into its component parts.

In this paper, we discuss evidence from language acquisition to address the nature of pragmatic inference, and the role of grammatical alternatives in scalar implicature. We argue that children’s selective difficulty with scalar implicature supports the idea that such inferences rely critically on grammatically constrained scales (as originally proposed by Horn, 1972). Furthermore, such evidence suggests that relations between scale mates must be acquired gradually over development, and do not emerge automatically from the acquisition of their syntactic and semantic features.

As argued in the sections below, various forms of experimental evidence indicate that young children (even as young as two) have sophisticated pragmatic reasoning abilities. Yet children of this same age are unable to compute simple quantity implicatures for sentences that involve quantifiers like *some*. For example, unlike adults, children do not derive the implicature in (1b) from the utterance in (1a).

(1) a. John ate some of the cookies.

b. John didn’t eat all of the cookies.
This type of failure has been found even in the late stages of acquisition (between the ages of 5 and 9; see Huang, Snedeker, & Spelke, in press; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Noveck, 2001; Musolino, 2004; Papafragou & Musolino, 2003; Smith, 1980). For example, in a study by Papafragou and Musolino (2003), 5-year-old children were shown a scene including three horses, in which all three animals successfully jumped over a log. After viewing the scene, a puppet commented, “Some of the horses jumped the log.” When children were asked whether the puppet ‘answered well’ (a so-called Felicity Judgment Task) they replied “yes”, unlike adults. Similar failures have been reported for a range of other scalar contrasts, including might vs. must (Noveck, 2001), a vs. some (Barner et al., 2009), and or vs. and (Chierchia, Crain, Guasti, Gualmini, & Meroni, L., 2001; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001).

Despite these failures, children have no difficulty deriving strengthened meanings for sentences that contain numeral modifiers like two, three, etc. Even young children are able to compute the inference in (2b) from the utterance in (2a).

(2) a. John ate two of the cookies.

       b. John didn’t eat three of the cookies.

According to most previous accounts in the psycholinguistic literature, children’s failure with inferences like (1) suggests that they cannot compute implicatures (see Braine & Roumaine, 1981; Noveck, 2001; Chierchia et al., 2001; Musolino, 2006; Musolino and Lidz, 2006; Papafragou, 2006; Papafragou and Musolino, 2003; Pouscoulous et al., 2007; Hurewitz et al., 2006, among others). Stemming from this conclusion, others have concluded that the strengthening of numerals, like in (2), must not be pragmatically mediated, and thus that numerals must have lexically strengthened, exact, meanings (see Huang and Snedeker, 2009; Huang, Spelke, and
In contrast, our analysis of the literature suggests that young children are often quite competent at deriving quantity implicatures as long as they have acquired the appropriate scale. By our account, children’s difficulties are much more local. Not only are their failures specific to quantity implicatures, but they are restricted to certain types of scales. To compute quantity implicatures, we argue, children must not only know the meanings of words like *some* and *all*, but they must learn that such words are related to one another as members of scales - i.e., as scalar alternatives. These scales, we suggest, are constrained by grammar, but nonetheless require the acquisition of associations between scalar items. In other words, by our account, the syntax and semantics of scales are not sufficient to guarantee that words will be accessed *qua* scale mates when processing sentences.

These points are supported by data from language acquisition, which not only allow us to adjudicate between competing models of how scalar inferences are computed in adults, but also address a broader set of issues regarding the role of Gricean inference in word learning, quantifier interpretation, and the acquisition of number words like *one*, *two*, and *three*.

Our discussion proceeds as follows. First, in Section 2 we review the linguistic arguments in favor of scales as grammatical objects, discussing in particular why quantity implicatures cannot be given a more general Gricean treatment independent of grammatical scales. In Section 3, we describe several sources of evidence which suggest that children are competent pragmatic reasoners, and that their failure to compute scalar implicatures is due to an inability to access certain scalar alternatives. Evidence comes from word learning, number word development, and children’s
ability to interpret utterances that depend on contextual scales. As discussed in Section 4, these data rule out the possibility that scales emerge automatically from the semantics of their members, and suggest that scales are not only grammatical in nature, but that they must become associated with one another gradually over acquisition.

2. Scales as Grammatical Objects

To derive the inference in (1b), *John did not eat all of the cookies*, from the utterance in (1a), *John ate some of the cookies*, linguists since Grice (1975) have posited that utterances are contrasted with a set of alternative statements that the speaker could have made, but chose not to. This section outlines the basic steps in deriving inferences from potential alternatives and the role of scales within this process. Although we describe our proposal in the context of a standard Gricean approach to implicature, our analysis and its conclusions are also consistent with grammatical approaches to implicature, such as those discussed by Landman (2000), Chierchia (2004), Fox (2006) and Chierchia, Fox and Spector (2010).

According to the standard Gricean analysis, upon hearing an utterance like the one in (1a), a listener assumes that the speaker is being cooperative (Principle of Cooperation), and thus that the utterance is the most informative one that could have been made (Maxim of Quantity). Consequently, they assume that the speaker believes that a sentence like the one in (3) could not be true.

(3) John ate all of the cookies.

The sentence in (3) asymmetrically entails the sentence in (1a) (i.e., eating all of the cookies necessarily involves eating some of them but not vice versa), and thus by most (neo-)Gricean accounts of informativeness, the sentence in (3) is more informative. To maintain the assumption that (1a) is the most informative statement
that could be made, the listener must assume that there is some reason that (3) could not have been uttered, namely, that it is false (since uttering 3 in a situation in which it is false would violate the Maxim of Quality). In assuming this, the listener derives a meaning from (1a) that is equivalent to the sentence in (4).

(4) John ate some, but not all, of the cookies.

In previous work (Barner & Bachrach, 2010; Barner, Brooks, & Bale, 2011), we have characterized the set of steps involved in this inference as in (5):

(5) I. Compute basic meaning of an utterance U containing a scalar item.

II. Generate a set of alternatives \{p1, p2, \ldots, pn\} to U, called U_{alt}. These are all the propositions that can be generated by replacing the scalar item with its scalar alternatives.

III. Restrict the alternatives in U_{alt} by removing any alternative that is entailed by the literal meaning of the original utterance U. Call this restricted set U^{*}.

IV. Strengthen the basic meaning of U (containing the scalar item) with the negation of all of the members of U^{*}.

First, the listener computes the basic, literal, meaning of the expression (Step I). Second, they generate the set of alternative sentences that might have been uttered (by substitution of scalar alternatives; Step II). Third, they restrict these alternatives by removing those that are less informative (Step III). Finally, they “strengthen” the interpretation of the sentence by negating the remaining alternatives – i.e., by assuming that the speaker believes that the alternatives are false (Step IV).

---

1 As noted by Sauerland (2004), the Gricean account must assume a stronger version of Quality than originally discussed in Grice (1975). Grice’s algorithm easily derives that the speaker is ignorant about the truth and falsity of the stronger alternative but does not derive the inference that the speaker believes it to be false. The listener must assume that the speaker is opinionated about such alternatives in order to get the latter inference. We will ignore such subtleties in this paper since they are orthogonal to our main thesis.
The basic algorithm in (5) makes reference to scalar items and scalar alternatives. However, in his original discussion of Quantity, Grice made no mention of an explicit method for generating such alternatives. Hence, any sentence not entailed by the original utterance was a potentially viable candidate. As noted in the subsequent work of Horn (1972; 1989) and Gazdar (1979), without a method for restricting scales, the Gricean algorithm generates many unattested inferences. Thus, Horn introduced the notion of a scale to limit the number of alternatives and hence restrict the power of the Gricean algorithm.

To see why such limits are needed, suppose a speaker utters the sentence in (1a), *John ate some of the cookies*. In such a case, each sentence in (6) is more informative than (1a).

(6) a. John ate some of the chocolate cookies.
   b. John ate three of the cookies.
   c. John ate some of the small cookies.

If these sentences were permitted as alternatives, then it would follow from the algorithm in (5) that (1a) should imply the falsity of (6a-c). In other words, the listener would take (1a) to communicate a message similar to the sentence in (7).

(7) John ate some of the cookies, but not any chocolate ones or small ones and not more than two.

In fact, the implicature would likely be much stronger still. Consider the scenario where even two or three additional alternatives are included, such as those in (8).

(8) a. John ate some of the peanut butter cookies.
   b. John ate two of the cookies.
   c. John ate some of the big cookies.
With all of these more informative alternatives and other similar sentences, the message communicated by (1a) would be as in (9).²

(9) John ate some of the cookies, but not any specific kind of cookie, nor any specific size, and not more than one.

To avoid such unattested inferences, Horn (1972, 1989; see also Gazdar, 1979) refined the Gricean account by hypothesizing that certain utterances contain special phrases and/or lexical items called scalar items. Such items are associated with a restricted set of possible substitutions that limit the number of alternatives (see Matsumoto, 1995, and Gazdar, 1979, for a discussion).

More precisely, assume that there exist two functions: one called SCALE which maps syntactic objects to a set of possible substitutions, and another called ALT which maps sentences to a set of possible alternative sentences. The nature of the link between syntactic objects and substitution classes is a matter of much debate, however for certain phrases the link is well established empirically. For example, by most accounts the output of \(\text{SCALE}(\text{some of the cookies})\) is the set \{\text{some of the cookies}, all of the cookies\}.³ Given this substitution class, we can characterize Horn’s account of the set of alternatives for a sentence like (1a) with the definition in (10).

---

² The situation becomes even worse if John ate some but not all of the cookies is considered as an alternative. The strengthened meaning of the utterance would imply that John ate all of the cookies, and would thus contradict the implicature that John did not eat all of the cookies. To avoid such contradictions, hearers would need to avoid computing strong implicatures altogether and would thus be limited to weak, ignorance implicatures such as ‘the speaker is unsure whether the sentence John ate all of the cookies is true or false’. However, speakers hardly ever limit themselves to such weak conclusions.

³ Note that we will follow Horn (1972, 1989) in this paper by referring to the substitution class as scales. However, unlike Horn (1972, 1989) we do not assume that the substitution class is ordered. Rather, ordering occurs with respect to the sentential alternatives derived from the substitution class, and thus such orderings depend on the context of the substitution rather than on a predetermined rank in a scale.
Where $[\ ]_{SI}$ indicates a scalar item, $\text{ALT}(\text{John ate } [\text{some of the cookies}]_{SI}) = \{p: \text{for some } x \text{ in } \text{SCALE}([\text{some of the cookies}]). p = [\text{John ate } x]\}$

Critical to this definition is the limited set of substitutions allowed for the scalar item \textit{some of the cookies}, and in particular the impossibility of phrases such as \textit{some big cookies}, \textit{two of the cookies}, \textit{some peanut butter cookies}, etc. Since these phrases are not scale members, it follows that the sentences in (6) and (8) do not figure into the derivation of quantity inferences with respect to (1a).

Horn (1972, 1989) not only introduced the concept of scales but also sought to define constraints on which syntactic items could be associated with other syntactic items in terms of substitution classes; i.e., the grammatical constraints on forming scales (see also Gazdar, 1977, 1979). He proposed scalar restrictions based on grammatical category (all scalar items must be of like grammatical category; see also Atlas and Levinson, 1981) as well as semantic properties such as monotonicity (all scalar items must have the same monotonic properties, c.f. Horn, 1989).

Within this grammatical tradition, there are at least two views of how scales might emerge in acquisition. First, it is possible that scales are automatically generated when the meanings and syntactic properties of the component parts of the scales are acquired. In other words, for any syntactic item $X$, $Y$ is a member of the scale (i.e., substitution class) associated with $X$ if and only if it has the same syntactic structure as $X$ and the same monotonic properties as $X$ (i.e, $\text{SCALE}(X) = \{Y: Y \text{ has the same category as } X \text{ and } [Y] \text{ has the same monotonic properties as } [X]\}$). Such an account would predict that as soon as the child acquires the function $\text{SCALE}$, they should also acquire the ability to generate any scale, \textit{modulo} the availability of the required lexical
meanings. Furthermore, if these constraints were innately specified, then the set of possible scales would be innately constrained.

Alternatively, acquiring adult-like knowledge of scales may require additional learning, beyond the syntax and semantics of scalar items. For example, to be rapidly accessed for the purposes of computing online pragmatic inferences, scalar items may need to be explicitly linked in acquisition, rather than being generated from a constraint-based algorithm. This is not to say that any two words might act as scalar alternatives, or that the constraints proposed by Horn (1989) are not involved in the formation of scales (by our account they are). Instead, these criteria might be necessary prerequisites for the creation of scales, but not sufficient; acquisition may involving learning, scale by scale, which substitution sets are associated with which scalar items, such that they are automatically activated whenever any of its alternatives are. As a consequence, knowledge of scalar alternatives may emerge at different moments in acquisition for different scales, such that, early in acquisition, children compute quantity-based implicatures for some scalar items but not others.

3. Pragmatic Development

Whereas studies of scalar implicature have often reported failures in children’s pragmatic reasoning, evidence from other areas of developmental psychology have reached a rather opposite conclusion - that children are sophisticated pragmatic reasoners from an early age, and use Gricean principles to guide their earliest linguistic processing, from the acquisition of nouns and verbs to the interpretation of numerals. Such sophisticated abilities run contrary to the hypothesis that a general

---

4 It should be noted that Horn (1972) explicitly rejected the possibility that scales were generated. He thought that scales were, in some sense, arbitrarily set, although universal.
inability to calculate implicatures or general insensitivity to pragmatic cues is responsible for children’s difficulties with certain kinds of quantity implicatures.

3.1 Children and the Principle of Contrast

Studies of lexical and syntactic acquisition in young children demonstrate that even infants are aware of subtle contextual cues that indicate a speaker's intention or desire - cues such as eye-gaze, facial expression, body language, and intonation (see Baldwin, 1993; Bloom, 2000; Rapacholi & Gopnik, 1997; Tomasello, 1992, among others). Furthermore, word learning experiments have demonstrated that children as young as two years of age reason in a distinctly Gricean fashion when inferring the likely referent of a novel word.

For example, in her discussion of the Principle of Contrast, Clark (1987, 1988) argued that children restrict the meanings of words on the basis of the meanings of existing lexical items, such that no two words have identical meanings (see also Gathercole, 1989). The simplest example of this form of inference comes from studies of Mutual Exclusivity (e.g., Markman, 1989). In these experiments, children are typically shown two objects, one of which has a known label (e.g., a car) and another which is completely novel. The speaker utters a sentence containing a nonce term -- e.g., a sentence such as “Can you point to the dax?” or “Can you give me the dax?”.

Children as young as two consistently assume that the nonce term refers to the novel object rather than the familiar one.

According to Clark, this reasoning process is Gricean in nature. Upon hearing an utterance with a novel noun, children assume that the noun is being used to refer to one of the two objects set before him/her. One of the objects has a well-known conventional label whereas the other does not. Thus, children assume that if the speaker intended to refer to the known object, then they should have used its
conventional label (e.g., “Can you point to the car?”, “Can you give me the car?”). Since the speaker did not use such an utterance, the child assumes that the speaker is talking about the unknown object. As noted by Gathercole (1989, p.694), the child’s reasoning can be characterized by the following steps (adapted from Gathercole, 1989, to be consistent with the examples discussed above):

(11) (i) That person is using a word I don't know, dax, in reference to one of those two objects.

(ii) One of those objects is a car, the other is a novel object.

(iii) The word ‘car’ is usually used to refer to cars.

(iv) I assume that person is trying to communicate with me.

(v) That person must not have been referring to the car or she would have asked me for “the car”.

(vi) Therefore, she must want the other object, and dax must refer to that other object.

This reasoning process involves four key properties. First, intentional ascription: children’s inferences are couched in assumptions they make about the intent of the speaker to refer to a particular thing in the context. Second, communicative cooperation: the child assumes that the speaker is being cooperative in their effort to communicate, and thus should use conventional terms for known entities when possible. Third, awareness of alternative utterances: children’s inference that a novel word does not refer to a familiar entity relies on their ability to generate alternative utterances that the speaker might have said, namely, that entity’s conventional label. Finally, negation of alternative meanings: if an existing label is accessed for the familiar entity, children can infer that the novel label must not apply to the familiar entity, but must instead refer to the novel one. Since the speaker did not use a
potential alternative, it is assumed that the speaker does not want the request communicated by the alternative to be fulfilled.

As discussed in Barner and Bachrach (2010) as well as Barner, Brooks and Bale (2011), not only is this type of reasoning highly complex, the steps involved are similar to the computation of quantity inferences, as discussed by Grice. Let's consider the steps again, drawing some parallels.

In the case of contrast, the child first assigns a basic meaning to the novel utterance. This is similar to STEP I in (5), the quantity implicature algorithm. The only difference is that the basic meanings for quantity inferences are assumed to be known, whereas word-learning scenarios by definition involve unknown meanings. Since it is clear from context that the unknown word refers to one of two objects, the child computes two possible meanings for the sentence: one according to which the novel word refers to one object, and another in which it refers to the other object. Next, the child considers alternative propositions. This is similar to STEP II in (5). In the calculation of quantity inferences, the alternatives are determined by a scale. In the case of Contrast, alternatives are determined by knowledge of object labels. After determining the set of possible alternatives, the child assumes that such alternatives are not appropriate in the current situation (or the speaker would have uttered them). This is similar to Step (IV) in (5). In fact, the only major difference between the derivation of quantity implicatures and the reasoning involved in Contrast is that Contrast does not involve eliminating weaker alternatives – i.e., Step III in (5). Otherwise, the algorithms are almost identical.

Interestingly, in addition to explaining how children interpret novel nouns and adjectives, Clark’s Principle of Contrast has been used to explain how children might initially interpret unknown number words. For example, in her 1992 study, Wynn
categorized children according to the numeral meanings they knew - e.g., one-knowers, two-knowers, etc. She observed that children who knew only one exact meaning – i.e., one-knowers – were able to use this knowledge to restrict how they interpreted other, yet to be learned numbers. For example, when she presented a one-knower with two sets – one containing one individual, and the other containing five – children pointed systematically to the larger set when asked to point at the set with five (see also Condry & Spelke, 2008). To explain how children knew that five cannot refer to sets of one despite not knowing its meaning, Wynn (1992) argued that they respect Clark’s (1988) Principle of Contrast: “Since all the children knew that the word ‘one’ refers to a single item, then if they knew that, for example, the word ‘five’ refers to a numerosity, they should infer that it does not refer to a single item since they already have a word for the numerosity one.” (p. 229) Further, Wynn showed that such inferences were limited to the numeral scale; when one-knowers were shown the same two sets and asked, “Can you show me the blicket balloons?”, where blicket could be interpreted as a non-numerical attributive adjective, they showed no preference for the set of five.

Wynn’s conclusion, and the conclusion of others after her (Condry & Spelke, 2008), was that children assume: (1) that numerals contrast in meaning, (2) that they contrast only with other numerals, and (3) that the meanings of known numerals restrict the possible denotations of unknown numerals. Such reasoning mirrors the steps involved in quantity implicatures even more closely than the examples with novel nouns (i.e., sentences containing unknown numerals have a restricted set of possible alternatives, determined by substitutions using the numeral scale).

3.2 Relevance Implicatures
Another example of children computing inferences involved in scalar implicature comes from experiments investigating evasive answers to questions, such as the answer to (12a) in (12b).

(12)  

a. Does Bill like orange juice?  
b. He likes apple juice.

The response in (12b) implies that Bill does not like apple juice, especially in contexts where it is given that the speaker knows Bill well (i.e., they are not ignorant of his feelings about orange juice). This inference can be derived from standard Gricean reasoning as follows: The listener assumes that the speaker is being cooperative in his response and thus is being relevant (Maxim of Relation). However, the literal meaning of his response does not answer the question and therefore is not completely relevant. Furthermore, there is a salient alternative, namely *Bill does like orange juice*, which would have been a more relevant response. To maintain cooperation, the listener assumes that there is some reason that the more relevant answer could not be stated. For example, the listener could assume that the speaker does not believe that the more relevant answer is true (Maxim of Quality). Thus, the message communicated by the speaker is equivalent to the sentence in (13).^5^  

(13)  

Bill likes apple juice but does not like orange juice.

The pragmatic reasoning involved in computing this kind of inference is very similar to the type of reasoning one finds with quantity implicatures. Both types involve computing the basic literal meaning of the utterance – Step I in (5). Both also involve alternative propositions – Step II in (5) – and both require that the literal

---

^5^ Again, Sauerland’s (2004) discussion of the opinionated speaker is relevant to this derivation. It is possible that the hearer will derive the weaker conclusion that the speaker does not know whether or not Bill likes orange juice. This depends on how strongly the hearer believes that the speaker knows Bill’s preferences. For simplicity we do not discuss this possibility here, although computing such inferences involves the same steps as the stronger inference, minus the negation of alternatives.
meaning be strengthened by negating potential alternatives. There are only two minor differences: (1) evasive inferences do not require any restrictions on the set of alternatives and (2) evasive alternatives are evaluated in terms of relevance rather than informativeness.

Experimental evidence demonstrates that children as young as 4 years old are able to compute implicatures based on evasive answers. For example, in a study by Sullivan, Davidson and Barner (2011), children and adult controls were shown three animals: a bunny, a puppy, and a chicken. In one scenario in the study, one character asked the question *Does Bunny play the guitar?* A second character then responded, e.g., *Puppy and Chicken play the guitar*, thus evading mention of Bunny. The children and adult controls were then asked by the experimenter whether Bunny played guitar or piano. Children as young as 4-years-old were more likely to judge that Bunny played piano than that he played guitar, a tendency which increased in older children.

In fact, across all age groups in this study children’s ability to calculate implicatures from evasive answers did not differ from their ability to calculate entailment relations. For example, although 3-year-olds, the youngest age-group tested, did not respond significantly above chance with respect to evasive answers, their behavior also did not differ significantly from their ability to calculate entailment relations, suggesting that they found it difficult to compute any kind of inference in this task, and not just relevance implicatures.

### 3.3 Scalar reasoning with Contextual Scales

Since Hirschberg (1985), understanding the relationship between traditional Horn scales and contextual scales has been an important focus of pragmatic research. Horn Scales typically contain functional lexemes such as *and, or, some, all* as well as the numerals (*one, two, three*, etc.). Such scales are accessible in almost any context and
hence the implicatures derived from them do not require any special contextual construction or activation. In contrast, contextual scales often require world knowledge (i.e., contextual information), such as explicitly learning the order of military ranks, the steps required to complete a task, or even the itinerary of a vacation. For example, consider the sentence in (14).

(14) John made it to Boston.

In most contexts, (14) does not imply that John did not make it to Montreal. However, if we learn that John had planned a business trip in which he would first go to Toronto, then Detroit, then Boston and finally Montreal, and we are aware that the speaker knows of this itinerary, then (14) would imply that John did not make it to Montreal.

According to (1985), knowledge of John’s itinerary establishes a contextual scale, where Boston is associated with the substitution class \{Boston, Toronto, Detroit, Montreal\}. Furthermore, such knowledge establishes a contextual entailment between sentences like (14) and the ones in (15).

(15) a. John made it to Toronto.
    b. John made it to Detroit.
    c. John made it to Montreal.

If (14) is true given John’s itinerary, then (15a) and (15b) would also be true, but not vice versa. If (15c) were true given John’s itinerary, then (14) would also be true, but not vice versa. Thus, relative to the contextual information, (14) is more informative than (15a) and (15b) but (15c) is more informative than (14). Hence,

---

6 Sometimes the implicatures derived from Horn Scales are called Generalized Quantity Implicatures, although such a label is controversial. Grice (1975) questioned whether any implicature is truly general.
following the algorithm in (5), (14) should imply that the speaker believes that (15c) is false.

Note that we have followed Hirschberg (1985) in assuming that there are only two differences between traditional scalar implicatures and contextually-mediated ones: (i) contextually determined scales, and (ii) relativization of informativeness to background information. Otherwise, the algorithm in (5) remains unchanged.  

Two previous studies that have investigated children’s understanding of contextual scales have reported no difference between these scales and Horn Scales. First, in a study by Papafragou and Tantalou (2004), 4- to 6-year-old Greek children watched a scene in which a character was assigned a task, such as wrapping two gifts: a toy parrot and a doll. The instructions outlining the task established either a contextual scale minimally consisting of definite noun phrases and their conjunctions (\{the parrot, the doll, the parrot and doll\}) or a Horn scale (\{some of the gifts, all of the gifts\}). After receiving the instructions, the character went into a dollhouse (out of view of the children) to do the chore. Upon returning, the experimenter asked the character if they had completed the task: e.g., “Did you wrap the gifts?” The character responded with a partial answer such as “I wrapped the parrot” (contextual scale condition) or “I wrapped some of the gifts” (Horn scale condition). Children were asked to reward the character if they thought that the task was adequately completed.

According to Papafragou and Tantalou, children refused to reward the character 90% of the time for contextual scales, and 77.5% of the time for Horn scales (no 7 Hirschberg (1985) argues convincingly that implicatures derived from Horn scales and contextually determined substitution classes should both be classified as Scalar Implicatures. However, whether contextual substitution classes and Horn scales should be treated as a single phenomenon is beyond the scope of this paper. Our main point in treating them as a single phenomenon is to highlight their similarities for the purpose of understand the role of alternatives in pragmatic competence.
statistics comparing these cases were conducted). Based on this, Papafragou and Tantalou concluded that children successfully computed implicatures in each case, and thus interpreted the responses as indicating only partial completion of the task (i.e., “I wrapped the parrot” implies the falsity of “I wrapped the parrot and doll”; “I wrapped some of the gifts” implies the falsity of “I wrapped all the gifts”). Consistent with this, most of children’s justifications of their responses mentioned the character not completing the task (e.g., not wrapping “the doll”, or not wrapping “all the gifts”).

However, as shown in an experiment described by Sullivan, Davidson, and Barner (2011), such behaviors may not reflect implicatures after all, since the task used strongly favors a focus on contrast, rather than on implicature per se. In a replication of Papafragou and Tantalou (2004), Sullivan et al. tested children using contextual and Horn scales, and as in the original study they found that children rejected weak statements approximately 63% of the time. However, Sullivan et al. also tested cases where the character uttered a statement that was stronger than expected, and thus which entailed the weaker, expected utterance. For example, when an animal was asked to paint “some of the stars” or “the star and the circle” the character said that they had painted “all of the stars” or “the star, the circle, and the square”. In these cases, children were just as likely to say that the character had failed to satisfy the request as when it uttered a weaker-than-expected statement, rejecting the stronger statements 55% of the time. This suggests that children were sensitive to whether the animal’s statement contrasted with the request, rather than using scalar implicature to reason about what the character did or did not do. Critically, of the children who rejected the weaker-than-expected utterances, 85% also rejected the stronger statements. Clearly, if children do not compute entailment relations in such a task (but favor contrast), then it cannot be concluded that the task demonstrates competence
with implicature (since understanding entailment relations is a precondition for computing implicatures).

In a different study, Katsos and Bishop (2011) also described experiments in which children performed similarly for both contextual scales and Horn Scales. In two separate experiments, children watched a scene where an animated caveman witnesses another character performing an action, such as a monkey eating two pieces of fruit (a biscuit and an orange), or a giraffe eating all of a given set of pears. The experiment began with a preamble that pointed out relevant objects in the scene, such as the items of food that could be eaten, thus plausibly establishing a contextual scale consisting of definite noun phrases and their conjunctions. The caveman was asked a direct question about what the character did: e.g. “What did the monkey/giraffe eat?”. The caveman responded with an under-informative answer either using a scalar item (e.g., “some” in “The giraffe ate some of the pears”) or using an item that would be more consistent with contextual scales (e.g., “the biscuit” in “The monkey ate the biscuit,” where the contextual scale would minimally be a superset of \{the orange, the biscuit, the orange and biscuit\}).

The children evaluated the caveman’s statement in two separate tasks. In one task children were asked whether the caveman had answered “right”. In another task, they were asked to reward the caveman with respect to how good his answer was, with either a small, medium, or large reward. In both tasks, children treated scalar and contextual items alike. In the binary “right” vs. “wrong” task, most children accepted under-informative statements for both types of scale. In the reward task, children generally gave a medium sized reward for under-informative statements, regardless of scale type. Each case offers only weak evidence, however, for children’s underlying reasoning processes. In the first case, all utterances were logically “right” and thus
asking small children whether they were “right” or “wrong” is not likely a good test of felicity to begin with. In the second case, many states of affairs would predict children giving a medium-sized award, many of which do not involve implicature, including uncertainty as to the nature of the question, or as to whether the answer was correct. Finally, in both types of tasks, semantic factors other than implicature likely affected children’s relative uncertainty about whether or not to reject statements – i.e., the fact that WH questions typically require an exhaustive response. Several previous studies have shown that children have significant difficulties with exhaustivity, suggesting that this problem alone could account for all of the Katsos and Bishop results (e.g., Caponigro, Pearl, Brooks, & Barner, 2011; Roeper, Schulz, Pearson, & Reckling, 2007). Children who are simply not sure whether a non-exhaustive response to a question is correct may offer intermediate rewards, and similarly may fail to outright reject statements as false.

Subsequent studies provide more unambiguous evidence that children treat contextual scales differently from Horn scales. In a recent study Barner, Brooks and Bale (2011) explored the role of contextual alternatives by assessing children’s ability to interpret sentences that differed only with respect to the type of scale involved. To remove pragmatic inference as a variable from their study (and thus focus on the role of scales), Barner et al. tested children with sentences that included the focus element *only*. The semantics of *only* guarantees that sentences will have a strengthened meaning according to an algorithm that mirrors the steps involved in scalar implicature. For example, computing the meaning of a sentence like (16a) or (16b) involves denying a set of alternatives.

(16)  
   a. Only some of the animals are sleeping.  
   b. Only the dog and the cow are sleeping.
For (16a), the denial of the alternative in (17a) is logically entailed, rather than being merely implied. For instance, although scalar implicatures are defeasible, as shown by the lack of contradiction in (17b), equivalent sentences involving only are not, as shown by the contradiction in (17c). Critically, the deductive inference that is triggered by only, like the equivalent scalar implicature, requires knowing that all of the animals is a scale mate of some of the animals.

(17)  a. All of the animals are sleeping.

b. Some of the animals are sleeping. In fact, all of them are.

c. Only some of the animals are sleeping. In fact, all of them are.

In contrast, the potential alternatives for (16b) are determined by context. For example, in a context that contains three characters – e.g., a dog, a cow and a cat – the sentence in (16b) would entail the falsity of the sentence in (18a) but not the one in (18b). If an additional character were introduced in the original context – e.g., a chicken – then (16b) would entail the falsity of both (18a) and (18b).

(18)  a. The cat is sleeping.

b. The chicken is sleeping.

In Barner, Brooks and Bale (2011), 4-year-old children were either shown a picture where two out of three characters were doing a specific type of activity like sleeping, or where three out of three were doing the activity. The children were asked a question (utilizing only) about whether two of the characters were involved in the activity (e.g., Are only the dog and the cow sleeping?).

---

8 The relevant scale for [the dog and the cow] would be {[the dog], [the cow], [the cat], [the dog and cow], [the dog and cat], [the cat and cow], [the dog, cow and cat]} – i.e., the set of all individual denoting terms and group denoting terms involving the relevant characters in the given context. Denial of the non-weaker alternatives ‘The cat is sleeping’ and ‘The dog, cow and cat are sleeping’ would both result in the entailment that the cat is not sleeping.
On trials where the picture showed only two of the characters doing the activity (e.g., the dog and the cow sleeping but not the cat), the 4-year-olds answered “yes” to the question 95.8% of the time. When the picture showed all three characters doing the activity (e.g., the dog, cow and cat sleeping), the 4-year-old children answered “yes” only 14% of the time. However, when children were shown these identical scenes but were presented with questions that contained the scalar item *some* they showed no equivalent difference between utterances. When asked “Are only some of the animals sleeping” most children said “yes” even when all three of the animals were sleeping.
Because these results were obtained in a paradigm that avoided pragmatic inference altogether, children’s failure with the *some/all* scale cannot be attributed to a tendency to prefer literal interpretations (e.g., as proposed by Noveck, 2001 in his account of implicature failures) or a more general difficulty reasoning pragmatically. Instead, the results suggest that children’s access to scalar alternatives differs for Horn scales and contextual scales, and that this difference alone may explain children’s frequent failures with scalar implicature.

Very early success with contextual scales has also been shown for quantity implicatures that do not involve grammatical triggers like “only”. For example, in an experiment by Stiller, Goodman and Frank (2011) 3 to 5-year-old children were shown three objects, such as three happy faces. One of the objects had no unique identifying properties, one had an identifying property that it shared with the third object, while the third object had two identifying properties: the one it shared with the second object and a second unique property. For example, in the trials with happy faces, one of them was unadorned, another had glasses, and the third had glasses and a top-hat. With these three objects on display, a puppet made a statement to the children highlighting the property that was shared by two of the objects (e.g., *My friend has glasses*). Children were then asked to identify the object that the puppet was talking about.

The picture sets up a contextually determined scale (e.g., \{glasses, a hat, a hat and glasses\}). Thus, a competent pragmatic reasoner should interpret the puppet’s utterance relative to such scales (e.g., *My friend has glasses* will have the alternatives *My friend has a hat* and *My friend has a hat and glasses*). Relative to the context, *My friend has a hat* is more informative than the utterance *My friend has glasses* (since only one character has a hat, *My friend has a hat* uniquely identifies one of the
characters whereas the literal meaning of the utterance does not). Furthermore, My friend has a hat and glasses is more informative than My friend has glasses (the former entails the latter). Thus, according to the algorithm in (5), My friend has glasses implies the falsity of both alternatives. The denial of either sentence implies that the friend does not have a hat.

To test whether children computed this inference, Stiller et al. asked them to identify the puppet’s friend across different trial types. Even in the youngest age-group, between the ages of three and four, children were better than chance at picking out the object that had only one identifying property, such as the happy face with glasses but without a top-hat. In fact, the difference between 3-year-olds and 4-year-olds was not significant.

In summary, there is some evidence that children reason about utterances containing contextual scales more readily than they do about utterances containing Horn scales, and that difficulties accessing Horn scales may impair children’s ability to compute implicatures. These results suggest that children’s problems with other types of implicature are not likely due to the complexity of the algorithm outlined in (5).

3.4 Children and Numeral Scales

Not only are children able to compute pragmatic inferences rooted in contrast, and contextual alternatives, but, as alluded to in our discussion of number word acquisition, they also show evidence of using pragmatic inference to enrich their interpretation of numerals. In this section, we review evidence, originally presented in Barner and Bachrach (2010), that children initially assign non-exact meanings to numerals in acquisition and strengthen these meanings pragmatically to derive exact interpretations. An exact meaning for a numeral $N$ becomes available to children only
once they have acquired the next numeral in the number sequence (i.e., \(N+1\)). Such evidence supports the hypothesis that exact numeral meanings, even for young children, rely on the availability of relevant alternatives, as described in the algorithm in (5).

To understand this point, it is first helpful to review the stages by which children acquire numeral meanings, and to reconsider the assumptions typically made about these stages. As mentioned in Section 3.1, Wynn (1992) identified children’s “knower-level” for numerals based on their behavior on the Give-a-Number task, in which they were asked to give different numbers of objects in a titration paradigm (e.g., if they answered correctly for \(two\), they were tested with \(three\); if they answered incorrectly for \(two\), they were tested with \(one\) instead). If children consistently gave only one object when asked for \(one\) but did not give one for larger numbers, they were labeled as \(one\)-knowers. If they consistently gave two when asked for two, but not for larger numbers, they were labelled as \(two\)-knowers. This same system was used to classify children as \(three\)-knowers. Children who could enumerate larger sets were classified as “Cardinal Principle knowers” on the assumption that these children understood how to count, and could use counting to identify the cardinality of large sets.

Note that this system of classification implicitly assumes that numerals have exact lexical meanings, and that being a one-knower requires having an exact meaning. By definition, this classification system could never identify a child as a one-knower if numerals had weak lexical meanings, and exact meanings were derived pragmatically. This is because if exactness were derived via scalar competition, as outlined in (5), then so-called one-knowers would actually need to know both the meanings of \(one\) and \(two\), such that \(one\) could be strengthened via the negation of its stronger scale.
mate two. In other words, the children Wynn labeled as one-knowers might in fact be described as two-knowers, if exactness is derived pragmatically, rather than specified lexically for each individual lexical item.

Whether children acquire exact lexical meanings for numerals or derive exactness via scalar implicature is an open empirical question. In many cases, the two proposals make overlapping predictions regarding children’s behavior on counting tasks. However, there are nonetheless important empirical differences. Most notably, only the scalar account predicts that children who treat N as exact (Wynn’s N-knowers) must also know the meaning of its successor N+1. The semantic hypothesis makes the opposite prediction – N knowers by definition do not yet know N+1, and should not require knowledge of N+1 to interpret N as exact. Therefore, to decide between the scalar and semantic accounts we can ask whether there is evidence that N-knowers know something about the meaning of N+1 despite failing to assign it an exact interpretation. That is, we can ask whether N-knowers have a weak, non-exact meaning for N+1.

The relevant data can be gathered from two experimental paradigms – Wynn’s Give-a-Number task, and the “What’s on this Card” task. First, consider children’s behavior on the Give-a-Number task. It has sometimes been informally observed by researchers that N-knowers frequently give correct amounts for N+1 (e.g., non-knowers frequently give one when asked for one). However, based on Wynn’s criteria, discussed earlier, these children cannot be called N+1 knowers because they also give N+1 frequently for higher numbers. The rationale for not calling these children N+1 knowers is that their correct behavior cannot be distinguished from a response bias to give N+1 (e.g., by chance). This conclusion would perhaps be convincing if such cases were rare, or were randomly distributed across unknown
numerals. However, the behavior is highly frequent and indicative of specific, non-exact knowledge of N+1.

Consider three data sets: Japanese and English data contributed by our own labs (from Barner, Libenson, Cheung, & Takasaki, 2009; Barner et al., 2009), Japanese, Russian, and English data from Sarnecka and colleagues (Sarnecka et al., 2007) and English data from LeCorre et al. (2006). For each data set, Barner and Bachrach (2010) analyzed how often children who were non-knowers gave one object when asked for one, how often one-knowers gave two for two, and how often two-knowers gave three for three. Thus, how often did N-knowers, who only treat N as exact, give N+1 for N+1.

Across the studies, N-knowers gave a correct amount for N+1 47% of the time. The lowest reported level of correct responses was 23% (by two-knowers in LeCorre et al., and non-knowers in Barner et al.’s English-speaking children), while the highest level was 75% (by Sarnecka et al.’s English-speaking children). Overall, children’s level of responding differed significantly from chance, consistent with the hypothesis that they have some meaning for N+1.

By definition, we know that many of these children also gave N+1 for numerals greater than N+1, and that this is what led them to be categorized as N-knowers, rather than N+1 knowers. However, children’s behavior for N+1 cannot be attributed to a simple response bias or default response. For example, we might suppose, by the exactness hypothesis, that children have some form of default response of giving N+1 objects for all numerals that they do not know (i.e., for N+1, N+2, etc.). However, this type of response cannot explain the data. Among children who had data for N+1 and higher numbers (i.e., the data collected by Sarnecka et al., 2007), N-knowers gave N+1 objects significantly more often when asked for N+1 than when asked for N+2 or
higher numbers. So, for example, non-knowers, by Wynn’s terminology, were more likely to give one object for *one* than for higher numerals. This provides compelling evidence that children labeled as N-knowers, who only treat N as exact, actually treated N+1 as distinct from higher numerals. This fact is clearly inconsistent with the hypothesis that responses beyond N are random, or purely attributable to a default response or response bias. Only a theory that posits that children know something about the meaning of N+1 could predict such behavior.

Additional evidence comes from Wynn’s “What’s on this card?” task. In this task children are shown large flashcards that depict different numbers of common objects such as bears, shoes, or balls, and are asked simply, “What’s on this card?” If children fail to provide a numeral, they are prompted to provide more information. For example, the child might be asked, “How many?” and if they respond with only a numeral they might be asked, “So, what’s on this card?”, with the intention of eliciting a complete noun phrase (e.g., “three bears”).

Barner and Bachrach (2010) analyzed two data sets (LeCorre et al., 2006; Barner, Lui, & Zapf, 2011) and found that when children were asked to name sets beyond their number knower level (as determined by Give-a-Number) their performance was far from random. For example, when non-knowers were asked to name sets of one object, they responded correctly 57% of the time, with the word *one*. Similarly, one-knowers labeled sets of two with *two* 58% of the time, and two-knowers labeled sets of three as *three* 58% of the time. Overall, N-knowers gave correct responses 58% of the time for N+1 and only 31% of the time for N+2. Also, they said N+1 for sets of N+1 significantly more often than they said N+1 for sets of N+2 (58% vs. 37%). Thus, N-knowers were more likely to give correct responses for N+1 than for N+2, and were more likely to say N+1 for N+1 than for other numerals.
To salvage the claim that these children actually don’t know N+1, we might again try to explain some correct usages via a response bias. Several studies have noted that, before becoming two-knowers, many children say *two* for all sets greater than one. This has led to suggestions that *two* may act as an early marker of plurality for children, before they acquire its precise meaning (Carey, 2004; Clark & Nikitina, 2009; Sarnecka et al., 2007; although see Barner, Lui, & Zapf, 2011). This proposal might explain why one-knowers correctly say *two* for sets of two between 70% and 80% of the time. However, it cannot explain why non-knowers say *one* for sets of one (more than 50% of the time) but not for sets of two (less than 5% of the time), or why two-knowers say *three* for sets of three (but not four), and yet do not say *four* for sets of four. In the Barner and Bachrach analysis, when all *two* responses were removed from the dataset, the pattern of results actually becomes even clearer than before. When *two* responses are not considered, N-knowers say N+1 for N+1 52% of the time, but 16% of the time for N+2. Similarly, N-knowers only say N+2 for N+2 about 22% of the time, on average.

In summary, data from children acquiring three different languages, collected by three independent labs, find evidence that N-knowers frequently provide correct responses when asked to give or name sets of N+1. As argued above, such data cannot be naturally explained by existing accounts that posit exact numeral meanings since such accounts are built on the assumption that N-knowers do not know the meaning of N+1. The data are consistent with the hypothesis defended by Barner and Bachrach (2010) that children’s initial numeral meanings are non-exact. Exact meanings are derived through scalar reasoning with non-exact numeral interpretations.

4. Conclusion: *Children and Horn Scales*
In many respects, children are competent pragmatic reasoners. Even by two-years of age they compare alternative utterances, reason about speaker intentions, and pay attention to subtle contextual cues when learning new words (see section 3.1). Around the same age, they are able to make inferences that mirror quantity implicatures when interpreting numerals (see section 3.2) and less than a year later, they are proficient at deriving quantity implicatures using contextually determined scales (section 3.3). However, despite their competence with respect to these types of inferences, they still have difficulty computing basic quantity implicatures from Horn Scales like some vs. all. Such findings raise the important question: what causes the observed differences between children’s ability to strengthen utterances that contain lexical and contextual scales?

Several studies shed some light on this difference. Experiments on quantifier meaning and logical connectives have revealed that children know the meanings of and, or, might, must, some and all well before they are able compute quantity implicatures (Paris, 1973, Brain and Roumaine, 1981, 1983; Noveck, 2001; Barner, Chow et al., 2009; Barner, Libenson et al., 2009). Thus, their inability to compute such inferences does not reflect a lack of semantic knowledge. Furthermore, children between the ages of 3 and 6 can compare sentences in terms of informativeness, even with respect to problematic scalar items such as and and or. An example of this comes from an experiment by Chierchia et al. (2001). Although children do not consistently reject the use of or in situations where and would be more appropriate (i.e., they do not compute an exclusive-or interpretation; see Chierchia et al. 2001; Paris, 1973; Brain and Roumaine, 1981), they do consistently choose statements with and over those with or when the situation is logically compatible with either. For instance, in the third experiment in Chierchia et al., children between 3- and 4-years-old (3;2 to
6;0, mean 4;8) were asked to judge which of two puppets made a more appropriate statement (a Felicity Judgment Task). The puppets uttered two different sentences that were identical except for the use of and vs. or (e.g., in one trial, one puppet said “Every farmer cleaned a horse and a rabbit” while the other said “Every farmer cleaned a horse or a rabbit”). On 93.3% of the trials, children chose the puppet that used and over the one that used or. Thus, children had no difficulty comparing sentences in terms of their informativeness (see also Katsos and Bishop, 2011, discussed in section 3.3).

If children are proficient at calculating quantity implicatures with contextual scales, and if they have no difficulty with the semantics of scalar items nor with comparing sentences in terms of informativeness, then what else could explain children’s deficiency with quantity implicatures? One possibility is the nature of logical scales themselves. Unlike contextual scales, logical scales are accessible independent of context. Thus, such scales require either grammatical generation or memorization to be accessible.

Given children’s success at computing quantity implicatures for numerals (section 3.4), it seems unlikely that an inability to generate logical scales is at the root of children’s difficulties. Clearly numerals are not contextually determined. Thus, if logical scales were grammatically generated, children should have this ability by the age of two. Yet, they are unable to reliably calculate implicatures for some, might and or until the age of nine.

A more likely explanation for children’s failure is the hypothesis that logical scales require explicitly learning substitution classes. In other words, it is not sufficient to know that some of the cookies and all of the cookies belong to the same syntactic category and that their meanings have the same monotonic properties (c.f.
Horn, 1989), nor is it sufficient to know that all and some are both (logical) quantifiers. The child must also know that \{some of the cookies, all of the cookies\} forms a substitution class, or at least minimally that \{all, some\} forms such a class. In other words, the fact that these quantifiers are alternatives to one another might have to be learned during the course of development. If this is the case, then children’s difficulties might be due to the fact that they have not yet learned this association – an extra piece of information beyond the syntactic and semantic properties of the two words.

Such a hypothesis can account for the differences between contextual scales and numerals, on the one hand, and quantifier scales, sentential connectives and modal operators on the other. Since contextual scales are established by context, memorization is not necessary. Information readily available in the context of utterance, or the co-text (i.e., the other utterances in the discourse), will be enough to establish the relevant scale. Although numerals are not context dependent, they are recited as a group in a sequence from a very early age. In fact, studies of early number word acquisition find that children begin to explicitly memorize a count list before they learn any numeral meanings, reciting it like the alphabet (Fuson, 1988). In contrast, no child is taught to recite quantifiers in a song, nor for that matter sentential connectives (and vs. or) or modal operators (might vs. must).

Thus, when interpreting utterances containing two, children might be highly sensitive to the fact that two and three are members of the same substitution class due to the fact that both words belong to the explicitly memorized count list. However, when interpreting utterances containing some, children may be unaware that all is a relevant alternative, since some and all have not yet been assigned to the same substitution class. For the same reason that children (or adults) do not treat two as an
alternative to *some*, they also might not treat *all* as an alternative. For them, these two expressions may differ in informativeness, but are simply not accessible alternatives when interpreting sentences (for a similar proposal, see Foppolo, 2007; Foppolo and Guasti, 2005).

This does not mean, however, that *all* is never an alternative to *some*. Since children are proficient at contextual scales, the two quantifiers, in principle, could be assigned to the same substitution class given the right circumstances. In fact, this might explain why children become much better at calculating quantity implicatures with respect to *some* vs. *all* when the context is enriched (see Guasti et al. 2005; Papafragou and Tantalou, 2004). However, much more research needs to be done to determine which contextual cues support the creation of such scales.

One issue that remains an open empirical question is how quantifiers, sentential connectives and modal operators finally get assigned to substitution classes that constitute logical scales. It is possible that, in the absence of brute memorization as with numerals, children group semantically related lexical items as scale mates by a gradual association of syntactically replaceable alternatives (see Katzir, 2007). Scale members may also be associated as children hear them explicitly contrasted in conversation -- e.g., “Give me some of the cookies but not all of them”. In each case, some form of gradual learning would be required, on top of acquiring the semantics of individual lexical items. Currently, no empirical studies address how scales might be constructed, and even in adults, the psychological status of scales is poorly understood -- i.e., which words form scales and which do not. Future studies should explore the nature of scales both in adults and children, to better understand how scales are constructed in acquisition.
References


Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The


Huang, Y., Spelke, E., & Snedeker, J. (under review). What exactly do number words mean?


