What exactly do numbers mean?

Yi Ting Huang, Jesse Snedeker & Elizabeth Spelke

Harvard University

Address correspondence to: Yi Ting Huang
Laboratory for Developmental Studies
Shannon Hall
25 Francis Avenue
Harvard University
Cambridge, MA 02138, USA
huang@wjh.harvard.edu

Abstract

The semantics of number words is a topic of much dispute in linguistics and of great relevance to understanding both the nature and development of number concepts and the division of labor between meaning and interpretation. Neo-Gricean theories posit that numbers, like scalar terms such as *some*, have lower-bounded semantics (*two* means "at least 2 and possibly more") and receive exact interpretations via the pragmatic rule of scalar implicature. Other theorists propose that numbers have exact semantics (*two* means "exactly two") and that apparent lower-bounded interpretations are achieved by contextual restriction or reference to subsets of the array. Prior findings that both children and adults strongly favor exact interpretations do not resolve this debate, because it is not clear whether this preference reflects an exact semantics or the calculation of a scalar implicature. We tested adults and 2- and 3-year-old children, with limited knowledge of number words, in a novel paradigm that teases apart semantic and pragmatic aspects of interpretation (the covered box task). Our findings establish that when scalar implicatures are cancelled in the critical trials of this task, both adults and children consistently give exact interpretations for number words. These results provide the first unambiguous evidence that number words have an exact semantics.

Keywords: number words; scalar terms; semantics; pragmatics; language acquisition

1. Introduction

Where does language end and communication begin? This relationship between semantic and pragmatic interpretation has been a perennial puzzle both in psychology and linguistics. General agreement about the existence of these levels of representation contrasts with controversy over their boundaries. Developmentally, this question arises because learning a language not only involves identifying words and establishing their structural relations, but also making pragmatic inferences that determine the speaker's intended meaning. How do children solve the simultaneous problem of knowing which aspects of the utterance are critical to lexical semantics and which ones are particular to the use of language?

This current paper examines these questions by exploring the controversial test case of number words. Linguists have long noted that number words appear to have two distinct interpretations (Horn, 1972 & 1989; Gadzar, 1979; Levinson, 1983). Although numbers are often interpreted as specifying the exact numerosity of sets, on occasion they can be used in a context where the total quantity of items is greater (lower-bounded or "at-least" interpretations). For example, *two* seems to mean *exactly two* in sentence (1), whereas in (2), speaker B seems to be interpreting *two* as something like *at least two and possibly more*.

(1) A bicycle has two wheels, while a tricycle has three.

(2) A: Do you have two children?

B: Yes, in fact I have three. (Horn, 1989, pg. 251)

Similarly in (3) we would not be surprised to find out that David actually had three or more chairs in his office.

(3) Bonnie: I need to borrow two chairs. Do you know where I could get them?

David: Sure, I've got two chairs.

The fact that number words can be interpreted in both of these ways creates a challenge for accounts of number word semantics. Many theorists have suggested that utterances like (2) and (3) reveal the lower-bound, lexical semantics of numbers and that exact interpretations only arise through pragmatic inferences (Horn, 1972 & 1989; Gadzar, 1979; Levinson 1983). Other theorists contend that number words have an exact semantics with lower-bounded interpretations arising through semantic composition or pragmatic interpretation (Koenig, 1991; Breheny, 2004).[1] Both types of theories argue that the mapping from lexical semantics to ultimate interpretation is complex, suggesting that we cannot trust our pre-theoretical intuitions to guide us to the underlying meaning of these terms. This opacity creates a potential complication for number word acquisition. The child, like the linguist, only has access to the final output of interpretation. She must work backwards to establish which aspects of meaning are assigned at the semantic level and which aspects are assigned at the pragmatic level.

We attempt to tease apart these two theories of number word semantics by examining how number words are interpreted, both in adults and in young children who have just recently learned them. In the remainder of the introduction, we will flesh out these two theories, examine the reasons why data from young children might be particularly revealing, and briefly look at other recent studies on children's interpretation of number words (see Musolino, 2004 for an earlier discussion of these issues). Finally, we describe a series of experiments designed to distinguish these two accounts.

1.1. Two means AT-LEAST-TWO: the Neo-Gricean proposal

The issue of number word semantics came to prominence when Horn (1972) argued that the interpretation of numbers closely paralleled the interpretation of *scalar terms*: sets of words

---

[1] A third possibility is that number words have underspecified meanings which are supplemented by pragmatic interpretation (Carston, 1985). We defer consideration of this theory to the General Discussion.

that can be arranged in an ordinal relationship with respect to the strength of the information they convey. For example, *some* is part of a scale that includes the stronger term *all*, and *warm* is situated within a scale that also includes *hot*. Scalar terms are typically interpreted as having both an upper and lower bound, giving rise to an interval reading which parallels the exact reading for number words. Thus a sentence like (4) will generally be taken to imply that Henry ate some, but not all, of the ice cream.

(4) Henry: I ate some of the ice cream.

However, in certain contexts scalar terms, like number words, also take on lower-bounded interpretations. Thus in (5), Karl asserts that Leif ate both *some* and *all* of the lutefisk (an infamous Norwegian dish made of fish soaked in lye), indicating that *some* in this context has a meaning which does not exclude the stronger term *all.*

(5) Eva: Did anyone eat some of the lutefisk?

Karl: Yeah, Leif ate some. In fact, he ate all of it.

Neo-Gricean theorists have discussed phenomena like these under the rubric of scalar implicature (Horn, 1972; Gadzar, 1979). Following Grice (1967/1975) they argued that weak scalars like *some* do not have a lexically-encoded upper bound and therefore are semantically compatible with stronger scalar terms like *all*. Scalars can receive an upper bound interpretation, as in (4) via a process of pragmatic inference. This inference is motivated by the listener's implicit expectation that the speaker will make his contribution to the conversation "as informative as is required" (the Maxim of Quantity). For example, if Henry had polished off the ice cream, (6) would be a more informative utterance than (4).

(6) Henry: I ate all of the ice cream.

However, since the speaker did not use this stronger statement, the listener can infer that he does not believe it to be true. Although scalar implicatures are robust across many contexts, they are by definition not a part of the truth conditional content of the sentence and can be cancelled, resulting in overt lower-bounded utterances like in (5). Further evidence for this division of labor between semantics and pragmatics comes from studies of online sentence processing which document temporal delays in the calculation of the upper bound through scalar implicature (Bott & Noveck, 2004; Noveck & Posada, 2003; Huang & Snedeker, 2005; Storto, Tannenhaus, Carlson, 2005).

The Neo-Gricean account of number semantics capitalizes on these parallels, arguing that numbers are simply another set of scalar terms (Horn, 1972 & 1989; Gadzar, 1979; Levinson, 2000; Krifka, 1999; Winter, 2001; Kratzer, 2003).[2] Like other scalars, they possess lower-bounded semantics (*two* means AT LEAST TWO) but receive an upper bound via scalar implicature. Implicatures are calculated in most situations and listeners access the exact interpretation of the utterance. But when implicatures are cancelled, the true meaning of the number word is visible, yielding the marked lower-bounded interpretation, as in (2) or (3).

1.2. Two means two: Theories of exact semantics

While the Neo-Gricean account neatly captures the parallels between cardinal numbers and scalar quantifiers, it flies in the face of the pre-theoretical intuition that numbers have exact meanings. Shades of this robust belief color even the theoretical literature. For unequivocal scalar terms like *some*, the exact interpretation is often glossed as "some but not all" implying that the core meaning of *some* includes cases where *all* also applies. In contrast the exact interpretation of a number like *two*, is typically glossed as "exactly two," suggesting that the

[2] Many others presuppose the Neo-Gricean symmetry between number and scalar interpretations at the level of compositional semantics without explicitly adhering to this account of lexical semantics (see e.g., Chierchia, 2000; Fox & Hackl, 2004; van Rooy & Shulz, 2004; Geurts & Nouwen, 2005).

upper bound is not an additional restriction, but merely part of the precise meaning of a cardinal number. Several theorists have pursued this intuition, arguing that numbers, unlike scalar quantifiers, have exact lexical semantics which delimits both their upper and lower boundaries (Saddock 1984; Koenig, 1991; Horn, 1992; Scharten, 1997; Breheny, 2004; Bultinck, 2005). The challenge for these accounts is to explain how exact meanings can give rise to sentences that appear to have lower-bounded interpretations.

There are two possible solutions to this problem. Koenig (1991) argues that it is the compositional semantics of distributive count phrases which opens the door for apparent lower-bounded interpretations. He points out that utterances with cardinal phrases, such as (7) can be interpreted in two ways.

(7) Two boys carried a box.

Under the collective interpretation (8a), "two boys" picks out a set of boys with two members who, as a group, carry out the action. Koenig argues that collective readings necessarily lead to exact interpretations (8b). This inflexibility is inconsistent with the Neo-Gricean view since upward-bounding scalar implicatures, like all implicatures, are by definition defeasible.

(8a) Two boys together carried a box.

(8b) *Two boys together carried a box. In fact, three boys together did so.

In contrast in the distributive interpretation (9a), the number word is used to assert that two such individuals exist, rather than to enumerate a set. While we may infer that this is the total number of individuals meeting these requirements, the semantics is mute on this point, making this inference defeasible in (9b).

(9a) Two boys each carried a box.

(9b) Two boys each carried a box. In fact, three boys each did so.

Thus distributive readings give rise to apparent lower-bounded interpretations despite the fact that number words have exact meanings.

A second possibility is that apparent lower-bounded interpretations arise via pragmatics. Breheny (2004) suggested that numbers have an exact lexical semantics and refer to the exact numerosity of particular sets. However, within a given utterance pragmatic factors play a role in determining which set is being referenced. Since larger sets necessarily contain smaller sets, this pragmatic flexibility creates the opportunity for lower-bounded interpretations. In most contexts, count phrases will typically refer to the maximal set because it is relevant to know the total number of items, e.g. (1) and (7). However, in certain circumstances it is more pertinent to the goals of the conversation to ascertain whether a subset of a particular numerosity exists. For example in (3), Bonnie is more interested in whether David has *any* set of two chairs than she is in learning the total number of chairs in his possession. Under these circumstances the number still has an exact interpretation but gives rise to apparent lower-bounded interpretations because the quantified phrase refers to a subset rather than the maximal set.

To summarize, both Neo-Gricean and Exact Semantics accounts provide prima facie adequate explanations for the dominance of exact interpretations of number words and the occasional appearance of lower-bounded interpretations. Furthermore, both accounts do so by making use of the distinction between semantic representations and pragmatic processes (see Figure 1). A Neo-Gricean account does so by stating that, like other scalar quantifiers, number words have lower-bounded semantics (*two* means AT LEAST TWO) and receive exact interpretations via the pragmatic inference of scalar implicature. Alternatively, an Exact Semantics account states that number words have exact meanings (*two* means EXACTLY TWO), which are evident when they refer to collective sets (Koenig, 1991) or the maximal set (Brehney, 2004), while apparent lower-bounded interpretations are attributed to distributive readings and pragmatic factors that lead to set decomposition.

1.3. Why children's interpretation of numbers might be particularly informative

There are several reasons to think that we might gain a better understanding of the semantics of number words by looking at how they are interpreted by young children (see Musolino, 2004 for discussion). First, children are notoriously poor at calculating scalar implicatures (Paris, 1973; Smith, 1980; Braine & Rumain, 1981; Noveck, 2001; Papafragou & Musolino, 2003; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001; Musolino & Lidz, in press). In many cases, children's pragmatic failures make the lower-bounded semantics of scalar terms visible in their overt judgments. For example, Noveck (2001) found that seven- to ten-year-olds, unlike adults, often interpreted weaker modal statements (e.g. x *might* be y) as being compatible with stronger statements (e.g. x *must* be y). Thus if numbers are upper-bounded only via scalar implicature, as the Neo-Gricean theory contends, we might expect that children would accept lower-bounded interpretations, even in contexts where adults prefer the exact interpretations (see Papafragou & Musolino, 2003).

Second, by studying number words as they are being acquired we can examine how children's interpretation of one number word is affected by learning the next. While many toddlers can produce a range of small number words in the context of the counting routine, there is considerable evidence that, these words are only mapped onto quantities slowly and sequentially during the preschool years (Wynn, 1990 & 1992; Fuson, 1988; Briars & Seigler, 1984; Le Corre et al., in press; Sarnecka, 2003; Mix et al., 2002). Wynn documented this progression using the "Give-a-number" task in which children were asked to give the experimenter different numbers of objects from a larger set. In her study, the youngest children began with a consistent and reliable interpretation of *one*; when asked for "one fish," they would hand the experimenter exactly one but failed to show any consistent interpretation of any larger numbers. By two-and-a-half years old, most of the children were "two-knowers:" they gave the experimenter exactly two fish when asked for *two*, but continued to grab a handful when asked for larger quantities. Several months later, these children began responding consistently to *three* ("three-knowers") and by around four-years of age, many had mastered "four" ("four-knowers") along with the ability to apply their counting routine to enumerate even larger sets. [3]

This prolonged period of limited competence could have profound effects on children's interpretation of known numbers. Scalar implicature critically depends on our knowledge of what the speaker might have said. According to the Neo-Griceans, we interpret *two* as exactly two because we know that a cooperative speaker could have said "three" if the situation had warranted it. But it's not clear that a "two-knower" has learned enough about the meaning of *three* to support such an inference. In the absence of a stronger term to drive the implicature, a principled Neo-Gricean should predict that the underlying lower-bounded semantics of the term would guide its use. Levinson (2000) developed just such an argument in reference to languages that only possess a small and finite set of numbers (Pica et al., 2004; Gordon, 2004).

---

[3] The alert reader might wonder whether the Wynn studies resolve the issue of the nature of early number semantics. At first glance Wynn seems to find evidence for lower-bounded semantics in the stage before each number is mapped to a specific quantity (for the two-knower *three* is interpreted as MORE THAN TWO) and exact semantics after the mapping is performed. Two further considerations lead us to reject this analysis. First, selection of an exact numerical match is consistent with both an exact interpretation and a lower-bounded interpretation since an ultimate preference for an exact interpretation is predicted by both Neo-Gricean and Exact Semantics proposals. Thus the stable response patterns that children acquire cannot tell us about the underlying semantics of the words. Second, in the period before the word is mapped to its correct meaning, its interpretation is identical to all larger numbers on the child's count list. This feature is not easily reconciled with a lower-bounded semantics. One would have to assume that the child allows several words to be mapped to the same numerical concept, that she spontaneously abandons most of these mappings as soon as she maps the lowest number of that group to an exact numerosity, and that she immediately remaps the others to the next lower-bounded quantity. It seems more parsimonious to assume that the child has identified these as quantity terms on the basis of their distribution, recognizes that they contrast with known numbers, but has not made a commitment to their precise meaning (Bloom & Wynn, 1997; Bloom, 1994).

> The scalar prediction is clear in these cases: we have a finite scale <'three', 'two', 'one'>, where 'one' or 'two' will implicate certis paribus an upper bound; but because there is no stronger item 'four', the cardinal 'three' should lack this clear upper bounding by GCI <Generlized Conventional Implicature>. (p. 90)

He noted that consistent with the logic of scalar implicature, speakers of these languages often use their largest number word to denote even greater quantities.

Finally, by studying young children we can look at the interpretation of number words prior to contact with formal mathematics. Levinson (2000) suggests that while number words in natural languages have a lower-bounded semantics, educated adults (and presumably school-aged children) have also acquired exact meanings for the mathematical numerals through formal education (p.89). Since numerals and number words are homophonous, it is difficult for mathematically literate informants to separate them when making judgments. However, since two and three-year-old children have had little direct exposure to formal mathematics, they are unlikely to be influenced by this putative lexical ambiguity.

These three considerations suggest that Exact and Neo-Gricean theories of number word semantics may best be distinguished by examining their interpretations in young children. Similar reasoning has prompted several recent experiments.

1.4. Prior studies of number words interpretation in children and adults

This issue was first explored by Papafragou and Musolino (2003) who tested both five-year-olds and adults using a pragmatic judgment task. Consistent with previous research, they found that children, but not adults, were content to accept weak scalar predicates (*started*) and quantifiers (*some*) in situations where the stronger scalar term applied (i.e. *finished* or *all*). In contrast, children treated count phrases in an adult-like manner, refusing to accept statements like "Two of the horses jumped over the fence," in a context in which they saw exactly three horses jump. They concluded that while children often have difficulty computing implicatures, they are able to achieve an exact interpretation for number words. This conclusion receives further support from Hurewitz et al. (in press) who asked three- and four-year-olds to find pictures where "The alligator took two of the cookies." Children, like adults, selected only the picture in which the character had exactly two of the four cookies, rejecting the one in which he had all four. However, these same children, unlike adults, happily selected both pictures when asked a parallel question about *some*.

Both of these studies suggest that number word interpretation does not follow the same developmental trajectory as the interpretation of true scalar quantifiers like *some*. However, these studies do not provide clear evidence for the underlying semantics of number words. As we saw earlier, both the Exact and the Neo-Gricean theories are able to account for the existence of lower-bounded and exact interpretations and both theories predict that the exact interpretation will frequently be favored. Thus knowing that children typically entertain exact interpretations of numbers does not, by itself, tell us how they reach this interpretation.

In fact, this pattern of discrepancies across items (numbers vs. unequivocal scalars) and age groups (adults vs. children) is compatible with two quite different accounts. First, these findings could be taken as evidence that number words have an exact semantics, on the assumption that children's performance with the true scalars in these experiments is indicative of a profound and global difficulty in calculating scalar implicatures (see e.g., Musolino, 2004). This interpretation is called into question by the evidence that preschoolers do not completely lack the capacity to calculate such inferences. Preschool children succeed at calculating scalar implicatures when they are given instructions and training emphasizing pragmatic interpretation over literal truth (Papafragou & Musolino, 2003) and when experimental tasks more closely approximate the role of implicatures in communicative interactions (Papafragou & Tantalou, 2004). If implicature is variable in childhood rather than absent, then the discrepancy between true scalars and numbers could reflect differences in the pragmatic processing of these items rather than differences in their meanings. In that case, children's number words may have a lower-bound semantics from the start and simply learn to calculate upward-bounding implicatures for *two* earlier than they learn to calculate them for *some* or *begin*. This precocity could be fueled by several factors: the frequency with which particular scalar implicatures are suspended, differences across scalar terms in the degree to which contexts unambiguously support upper-bounded interpretations, parental feedback about the correct use of number words (see General Discussion), and the role of the counting routine in allowing children to generate the alternative descriptions (Papafragou & Musolino, 2003).

1.5. What would it take to tell the difference?

To determine the lexical semantics of number words, we need to test participants in contexts in which the Exact and the Neo-Gricean theories make distinct predictions. The following experiments were undertaken to do just that.

Theories of exact lexical semantics predict that lower-bounded interpretations can be accessed whenever numbers are interpreted distributively or pragmatic considerations allow the set to be decomposed into subsets (Koenig, 1991; Breheny, 2004). To limit these theories to a single prediction, we used utterances and displays that favored a collective reading and clearly defined the set under consideration: the number words always occurred in a definite description (e.g., "the card with two fish") to minimize the possibility of a distributive interpretation, sets were delineated by clear physical boundaries to decrease the likelihood of set decomposition, and participants were asked to provide the entire set ("Give me the card/box with two fish").

In contrast, Neo-Gricean theories predict that the lower-bounded interpretation will be preferred only when scalar implicature are canceled. Prior studies (Papafragou & Musolino, 2003; Hurewitz et al., in press) have employed contexts where implicatures were readily available, as evidenced by adults' upper-bounded interpretation of *some.* This leaves open the possibility that exact responses for numbers were driven by a lower-bounded semantics. To ensure that the Neo-Gricean theory would predict a lower-bounded interpretation, we sought to cancel scalar implicatures in two ways.

First, we designed tasks in which even adults would be forced to suspend scalar implicatures. We reasoned that if implicatures are defeasible, they should be abandoned when they conflict with other information about the speaker's intentions or the context. In our tasks we do this by using the critical term as part of a definite description ("give me the card with two fish") in a situation with a salient lower-bounded target (e.g. three fish) and an upper-bounded target (e.g. one fish) but no exact match. The use of a definite description in a command presupposes that there is a unique referent available which can be acted upon. In this study the set of possible referents is clearly limited to the cards that are currently on display. If the critical term has lower-bounded semantics this presupposition can be satisfied by canceling the scalar implicature and accepting the lower-bounded target. This requires the participant to accept a violation of the Maxim of Quantity but it allows her to avoid the more drastic conclusion that the speaker has abandoned to cooperative principle by requesting something which is clearly unavailable (Grice, 1975). In this respect the current study is quite different from previous experiments on children's interpretation of number words, which have used contexts that support scalar implicature in adults (Papafragou & Musolino, 2003; Hurewitz, et al., in press).

Second, we tested two- and three-year-old children, a population in which scalar implicature appears to be weak or even absent. To further stack the deck against the use of implicature, we tested each child on the largest number word that they had a stable mapping for (as determined by their performance on Wynn's "Give-a-Number" task). Children, who lack knowledge of the word for next largest quantity, should fail to make strong predictions about how a cooperative speaker would use this label. Under these circumstances, the Neo-Gricean theory would predict that participants should allow number words to describe sets larger than their exact value while the Exact Semantics theory would predict that they should not.

In Experiment 1, we test these predictions in both children and adults using a standard forced-choice task. In Experiment 2 we introduce the covered-box task, a novel paradigm involving a mystery box whose contents are unknown. We use this task to test participants' interpretation of an unambiguous scalar quantifier (*some*) and find that both children and adults cancel scalar implicatures in this task when there is a salient lower-bounded match but no visible exact match. In Experiment 3, we use the covered-box task to test the interpretation of numbers, demonstrating that both adults and children stick firmly to exact interpretations of "two," even in the context in which they fail to calculate scalar implicatures for "some."

2. Experiment 1

2.1. Methods

2.1.1. Subjects

Ten native English-speaking undergraduates from Harvard University and 30 English-speaking children between the ages of 2;6 and 3;9 (mean 3;3) participated in this experiment. Child participants for all three experiments were recruited from a database of children who had previously participated in studies at the Laboratory for Developmental Studies at Harvard University.

2.1.2. Procedure and materials

Wynn's (1990) "Give-a-Number" task was used as a pretest to determine the children's knowledge of number words and assign them to experimental groups (adults were not given the pre-test). The children were presented with several small plastic fish and were simply asked to put different quantities into a basket ("the pond"). This procedure has been shown to be a stable and reliable measure of children's maximum number to word mapping during acquisition (Wynn, 1992; Le Corre et al., in press), correlating with performance on other procedures that require children to label existing sets and to choose between number pairs depicted cards. Children were classified as two-knowers if they gave one fish when asked "one," two in response to "two," and an arbitrary larger number in response to all other requests. Three-knowers performed like two-knowers except that they also gave three fish when asked for "three." Four-knowers were children who performed the task correctly for all numbers between one and four and many also demonstrated use of the counting routine to enumerate larger numbers up to six. The first ten children that we identified in each of these groups participated in this study. Three children who failed to achieve the minimal criterion for knowing "two" were excluded from the study.[4] The mean ages of the groups were 3;2 for two-knowers, 3;4 for three-knowers, and 3;5

[4] An anonymous reviewer pointed out that by using the Give-a-Number task as a pretest, we ran the risk of systematically excluding from the study any children who had a lower-bounded semantics for numbers. We chose to use this pretest for two reasons. First, the prior work demonstrates that slightly older children prefer exact matches for numbers when they are available (see e.g., Hurewitz et al., in press). The question at hand was whether children would accept a larger set when no exact match was present. Second, we wanted to avoid testing children on numbers that they had not yet mapped to a meaning (see footnote 3). To assess whether our procedure might be eliminating children with a lower-bounded semantics, we analyzed the quantities that children gave us when they failed to provide an exact match. Lower-bounded semantics would be manifest in two ways. First, children should never give fewer items than the exact value of the number requested. We found that 70% of the children who were classified as one- , two- or three-knowers did this at least once (e.g. two-knowers gave 3 fish when asked for *five*). Second if children have a lower-bounded semantics then on average the amount that they give for larger numbers

for four-knowers. We also elicited children's knowledge of the count list and found that all participants were able to count up to ten.

Both adults and children participated in the test phase. Different sets of test trials were used for each group to ensure that children were not tested on number words that they did not know. In particular, two-knowers were tested only on "two," three-knowers were tested on both "two" and "three," and four-knowers and adults were tested on "two", "three" and "four." Table 1 lists the comparisons that were given to each group. On every trial, participants were shown two cards and asked to select one of them. The cards depicted fish and sheep to ensure that the presence of the plural morpheme did not provide an additional cue to number.

The first block of trials consisted of eight critical trials and four filler items.[5] For the critical trials neither of the cards matched the exact value of the number being requested. For example, two-knowers were shown a card with one fish and a card with five and told *"Give me the card with two fish"*. One card was always larger than the exact value of the number requested, while the other was smaller than the exact value. If number words have lower-bounded semantics, we would expect participants to select the card with the larger quantity in the absence of an exact implicature match. If number words have an exact semantics then both of these choices should be unacceptable and we would expect participants to either explicitly reject both the larger quantity and the smaller quantity or randomly select between them.

---

should be greater than for smaller numbers (since the lower bound would also be higher). But we found no differences in the average quantity given for unknown numbers. Thus on average one-knowers gave the same amount for all numbers greater than two, two-knowers gave the same amount for all numbers greater than three, and three-knowers gave the same for *four* and *five* (all p's > .1). These results are consistent with previous findings demonstrating that children's failure to produce an exact response in the give-a-number tasks reflects ignorance about the meaning of that number word, rather than an initial Neo-Gricean interpretation (Wynn, 1990 & 1992).

[5] In the filler trials participants were shown a card with ten items and a card with one item and they were asked to select the card with one. All participants responded by selecting the card with just one fish, allowing us to rule out the possibility that they failed to understand the task or had a strong bias to always choose the larger of the two cards.

The critical trials were followed by six control trials which ensured that the participants could correctly pick out the exact match when it was present (as predicted by both the Exact Semantics and Neo-Gricean theories). Half the participants were given trials where an exact match was pitted against a smaller number (e.g., two-knowers were shown cards with two fish and one fish and asked for *two*) and the other half of the participants were given trials where an exact match was pitted against a larger number (e.g., for two-knowers, two vs. three or five).

2.2. Results

Participants in all four groups reliably selected the exact match during the control trials, regardless of whether it was pitted against a larger or smaller quantity (M = 96% for two-knowers, M = 92%, for three-knowers, M = 95% for four-knowers, and M = 100% and adults). Thus both the children and the adults clearly knew the number words that they were tested on and were able to perform the task. In contrast, performance on the critical trials varied across the four groups of participants (Figure 2). Participants rarely selected the smaller quantity (M = 1% for two-knowers, M = 0% for three-knowers, M = 6% for four-knowers, and M = 0% for adults), indicating that they did not adopt a strategy of randomly picking a card in the absence of an exact match. Instead they either selected greater quantity or simply rejected both choices. Adults rejected both cards on 91% of the trials, something that they never did when the exact match was present ($W = 55$, $Z = 2.78$, $p < .005$).[6] Thus they rigidly maintained an exact interpretation of the number words, even when no exact match was present.

Figure 2 suggests that the children's performance on the critical trials varied systematically with their knowledge of number words. Four-knowers, like adults, generally rejected both cards when no exact match was present (M = 74%), while two-knowers and three-

[6] Adult responses in all studies tended to be categorical. To cope with the lack of variability we used nonparametric tests wherever possible. Wilcoxon signed-rank tests were used for all within-participant pairwise comparisons and Wilcoxon ranked-sums tests were used for between-participants pairwise comparisons.

knowers typically selected the greater quantity (M = 96% and M = 85%, respectively). We explored this further by analyzing the proportion of rejections among children in a subjects ANOVA with level of number knowledge as a between subjects variable and trial type (critical vs. control trials) as a within subjects variable. There were significant main effects of both trial type ($F(1, 27) = 41.36$, $p < .001$) and number knowledge ($F(2, 27) = 23.33$, $p < .001$), as well as a reliable interaction in the proportion of rejections ($F(2, 27) = 23.79$, $p < .001$). All of these effects appear to be driven by the high number of rejections on critical trials by four-knowers.

When we directly compare the performance of the children and adults on the critical trials we find a significant difference in the proportion of rejections among the four groups, $F(3, 36) = 29.39$, $p < .001$. Post hoc analysis using Tukey's HSD revealed that four-knowers and adults as a group rejected both card choices significantly more often than two- and three-knowers ($p < .005$). There were no differences between the four-knowers and the adults or between the two- and three-knowers ($p > .10$).

What could account for this pattern of performance? We considered the possibility that children's ability to reject both card choices could be related to the degree to which they tested their most recently acquired number. Specifically, we predicted that three-knowers would be more likely to reject on trials asking for *two* (a previously acquired number word) than *three* (a recently acquired number word) and similarly four-knowers would be more likely to reject for *two* and *three* as compared to *four*. We explored this by separating out the two kinds of trials and analyzing them in using a two-way ANOVA with knowledge level as a between participants variable and recency of acquisition as a within participant variable. There was a significant main effect of recency, where lower numbers which were more established were more likely to be rejected than recently acquired ones ($F(1, 18) = 9.70, p < .01$). Unsurprisingly there was also a main effect of knowledge level ($F(1, 18) = 12.93, p < .01$). [7]

2.3. Discussion

Both adults and children overwhelmingly favored an exact interpretation when a quantity match was present. However, in the absence of an exact match, there was a developmental trend where the child's knowledge of numbers was positively related to their tendency to demonstrate a rigid exact interpretation of the number word. Specifically, we found that two- and three-knowers frequently selected the greater quantity while four-knowers and adults were more likely to reject both card choices.

The Neo-Gricean theory seems to provide a natural explanation for the choice of the greater quantity in the less proficient children and, perhaps, for the observed developmental progression. Taken at face value, the response pattern of the two and three-knowers suggests that they assign a lower-bounded semantics to number words. Under this interpretation there are two alternate explanations for the observed difference between these children and the four-knowers or adults. First, the change in behavior could reflect a difference in the lexical semantics of numbers. Perhaps acquiring additional number words gives the children the necessary information to replace or supplement a lower-bounded semantics with an exact

[7] Performance on numerical tasks is often sensitive to the ratio of the numerosity of the target and foil, with smaller ratios being more difficult to discriminate. In our task the relevant metric is the ratio between the number requested and the number depicted on the lower-bounded card. Anonymous reviewers have suggested that the effect of recency might be attributable to a difference in ratio since larger numbers stand in a smaller ratio to their nearest neighbors than smaller number. While ratios were not tightly controlled in this study, it was varied in some conditions allowing us to explore this possibility (see Table 1). We find no evidence that differences in ratio account for this effect. First, while recency of acquisition was confounded with ratio for four-knowers, it was not confounded for three-knowers (the average ratio for both two and three trials was 1:1.5), yet the effect of recency persisted in this group ($p < .05$). Second,, both two-knowers and three-knowers were tested on their most recently learned number at two different ratios (1:2.5 and 1:1.5 for two and 1: 1.67 and 1:1.33 for three). In neither case did we find a significant difference between the close and far ratios ($p > .10$). In fact three-knowers performed more poorly when tested on a newly acquitted number with a big ratio (1 vs. 5 ask for three) than when tested on previously acquired number with a smaller ratio (1 vs. 3 ask for two; $p < .05$). The absence of a ratio effect is not surprising: the numbers that were being tested were quite small, all the ratios were fairly large and the stimuli were made available for as long as the participant wished.

semantics. Alternately, the shift in behavior could reflect differences in the participants' ability to apply scalar implicature in this task. While we believed that implicatures would be cancelled by the absence of an exact match, it is possible that the adults and four-knowers calculated them anyway. If true, this would suggest that the ability to calculate these implicatures is tightly linked to the child's knowledge of number words. This offers a satisfying explanation for the finding that three- and four-knowers were more likely to accept the greater quantity when they were tested on their most recently acquired number: within the Neo-Gricean framework, implicatures should only be calculated when there is a stronger term on the scale that is well-established and mutually-known. Three-knowers, by definition, did not have a stable label for quantities greater than three, and four-knowers may have had difficulty linking larger number words to quantities in a task which they were not explicitly encouraged to count (see e.g., Le Corre & Carey, in prep).

In contrast, an Exact Semantics provides the most parsimonious explanation for why both the four-knowers and the adults were willing to violate the implicit requirements of the task and reject both of the cards on critical trials. But the performance of the two- and three-knowers is hard to reconcile with this theory. While it is entirely plausible that the children simply found it too daunting to challenge the task by rejecting both cards, this would fail to explain why they consistently selected the greater quantity on the critical trials. It also fails to account for the tight relationship between number knowledge and task performance.

One possibility is that two- and three-knowers had an exact semantics for number words but were able to access a lower-bounded reading in this context, by interpreting the utterance as referring to a subset of the array on the card (Breheny, 2004). While we had hoped that the use of definite descriptions would eliminate set decomposition, the use of smaller number words in the presence of a larger array may have had the opposite effect. Perhaps when asked for "the card with two fish," two-knowers reconstrued the card with three fish as card containing a subset of two fish. Two-knowers may be more vulnerable to this interpretation precisely because they lack a stable label for the maximal set of items on these greater quantity cards. This conjecture is supported by a more detailed analysis of the children's behavior during the task. On the critical trials, where no exact match was present, many two-knowers (seven out of ten) explicitly pointed to two of the fish when selecting the card containing 3 or 5 fish suggesting that they interpreted *two* as an exact subset within this larger array. These children did not point to specific fish during the control trials, when an exact match was present, suggesting that this gesture reflected the need to break a larger set down into subsets. Thus while the results of Experiment 1 are consistent with the Exact Semantics proposal, they are somewhat ambiguous and the children's gestures suggest that their performance may have been distorted by the task demands.

To resolve these issues we designed a new task that would allow participants to demonstrate an exact semantics without requiring them to step outside the task and explicitly reject the options that are presented. We also further discouraged set-decomposition by enclosing the sets in boxes to emphasize that they were bounded. In Experiment 2, we validate this task by using it to test children's and adults' interpretation of the scalar quantifier "some". We find that on the critical trials, when the exact match is not visible, both children and adults suspend or fail to calculate the implicature, accepting instances of all as examples of "some". In Experiment 3 we return to the question of number semantics, using our new task to demonstrate that adults and children derive an exact reading for number words, even in a context in which implicatures are suspended.

3. Experiment 2

Our goal in designing a new task was to create a context in which scalar implicatures would be cancelled while simultaneously providing participants with a clear way to indicate an exact interpretation of the critical word. When an exact match is provided adults perform scalar implicatures (see e.g., Hurewitz, et al., in press). When an exact match is absent there is a strong task demand to select one of the options provided, even if it does not match the description (see Experiment 1). In our new task we chose to have it both ways, providing a decoy which could be interpreted as an exact match, or not, as the participant saw fit. This decoy was simply a covered box with unknown contents. Again a definite description was used as part of a command and a salient lower-bounded target was provided. If the word in question has lower-bounded semantics then the presupposition of the definite description would be unambiguously satisfied by the visible lower-bounded match. Since this referent was accessible to both parties (in common ground) while the contents of the covered box were not, we expected that scalar implicatures would be cancelled in this task. Canceling the implicature would bend the maxim of quantity but it would avoid an egregious violation of the maxim of manner (Grice, 1975).[8]

To validate our procedure, we used a long time parade case for scalar implicature, the quantifier *some* (Grice, 1975). There is widespread agreement that the lexical meaning of *some* is compatible with *all* but that listeners typically restrict its interpretation via a pragmatic inference.[9] This analysis in supported by studies of adult processing and interpretation (Noveck & Posada, 2003; Huang & Snedeker, 2005) and children's propensity to give *some* an overt lower-bounded interpretation (Smith, 1980; Noveck, 2001; Papafragou & Musolino, 2003;

---

[8] Our predictions for this task could also be couched in the framework of relevance theory (Sperber & Wilson, 1986). Imagining an exact match in the covered box presumably requires additional cognitive effort while the visible lower-bounded option is readily available. Thus participants may relax their usual criterion for relevance and accept the lower-bounded reading because they fail to construct the other analysis in time.

[9] There is, however, considerable disagreement about the nature of this pragmatic inference and its relationship with compositional semantics (Chierchia, et al., 2001; Levinson, 2000; Katsos, et al., 2005; Recanati; 2003; Carston; 1998; Sperber & Wilson, 1986).

Hurewitz et al., in press). On the critical trials we presented participants with a box where Cookie Monster had none of the cookies, another box where he had all of the cookies and a covered box, and we asked them to "*Give me the box where Cookie Monster has some of the cookies*." If we are correct, and the covered box task succeeds in canceling the scalar implicature, then both the children and the adults should select the box where he has all the cookies.

3.1. Methods

3.1.1. Subjects

Thirty native English-speaking undergraduates from Harvard University and 10 English-speaking children between the ages of 2;6 and 3;5 (mean 2;9) participated in this experiment.

4.1.2. Procedure

The study was composed of three parts. During the Pretest phase, we used a modified version of the "Give-a-number" task to elicit children's knowledge of *some* and *all*. Children were asked to "*Put some (all) of the fish*" in a pond. All children who were tested demonstrated knowledge of these scalar terms by putting at least one fish in the basket when asked for *some* and by putting the entire quantity when asked for *all*. .

During the Familiarization phase, we introduced participants to the covered box task. On each trial they were presented with two open boxes containing toy animals and a third covered box and were asked to give the experimenter the box that contained a particular animal. The target animal was in one of the open boxes on two of the familiarization trials and hidden inside of the covered box on the remaining two. This sequence of four trials was repeated twice. The first time participants were given feedback after each choice and were allowed to open the covered box in their search for the target animal. The second time through, they were told not to open the covered box and they were not given any feedback.[10] All participants selected correct boxes when the target animal was visible and selected the covered box when it was not and were therefore included in this experiment.

During the Test phase, participants were presented with boxes containing pictures in which characters like Cookie Monster and Big Bird possessed cookies (see Figure 3). They were asked to "*Give me the box where Cookie Monster has some of the cookies*" in the following contexts:

1. CONTROL TRIALS: NONE vs. SOME. On three trials participants were presented with: one upper-bounded match (a box where Cookie Monster had none of the cookies), one implicature match (where Cookie Monster had some but not all of the cookies) and a third covered box. These corresponded to the "Exact versus Less" control trials in Experiment 1. If children and adults know the meaning of "some" they should consistently pick the implicature match.
2. CONTROL TRIALS: ALL vs. SOME. On three trials participants were presented with: one lower-bounded match (all of the cookies), one implicature match (some but not all of the cookies) and a third covered box. These corresponded to the "Exact versus More" trials in Experiment 1. These trials were similar to past studies examining the interpretation of "some" by adults and children (e.g., Smith, 1980; Hurewitz et al., in press). Thus we expected that our findings would closely parallel these studies; adults would calculate the implicature and select the implicature match, while children would fail to do so, choosing either the lower-bounded match or implicature match at random.
3. CRITICAL TRIALS: NONE vs. ALL. On three trials participants were presented with: one upper-bounded option (none of the cookies), one lower-bounded option (all of the

[10] The covered box was surreptitiously replaced with an empty box at this point.

cookies) and a third covered box. These corresponded to the critical trials from Experiment 1. These trials test our hypothesis that scalar implicatures will be cancelled when a definite description is used in the absence of a clear implicature match but in the presence of a set which matches the lower-bounded semantics of the term. We predicted that both children and adults would select this lower-bounded option. If they instead selected the covered box, it would suggest that they had calculated the implicature, rejected the lower-bounded match and inferred that an implicature match must be present in the mystery box.

While children received all trial types in randomized order, we adopted a between-subjects design for adults to prevent them from inferring the contents of the covered box by drawing comparisons across trials. Thus each adult saw only one of the three types of items and ten adults were assigned to each condition.

3.2. Results

As Figure 4 indicates, the response pattern varied across the three types of trials and across the two age groups. Both children and adults in the NONE vs. SOME control trials overwhelmingly selected the box containing SOME of the cookies (M = 93% and M = 100%) and there was no difference between the two groups (W = 95, Z = .76, $p > .4$). Thus the children clearly understood the task and recognized that *some* is incompatible with *none*. In ALL vs. SOME control trials, however, adults overwhelming favored the box with SOME of the cookies (M = 90%) while children were equally disposed to select either the box containing SOME or ALL of the cookies (M = 50% and M = 40% respectively). This resulted in a reliable difference in the proportion of SOME selections between the two groups (W = 69.5, Z = 2.68, $p < .01$). In fact the children did not have a reliable preference for the implicature match ($t(9) = .60$, $p > .5$), unlike the adults (t(9)=5.61, p<.001). These results are consistent with previous studies demonstrating that adults calculate scalar implicatures when an implicature match is present while young children often do not.

Finally, on the critical NONE vs. ALL trials, both the children and the adults strongly favored the box containing ALL of the cookies (M = 83% and M = 87% respectively, W = 96.5, Z = .64, p > .5). Very few selected the covered box (M = 7% and M = 13% respectively, W = 100, Z = .73 p > .5). Thus in the absence of a visible implicature match, even adults accepted a lower-bounded match (the box with ALL), rather than questioning the request (as they did in the Critical Trials of Experiment 1) or inferring that the covered box must contain an implicature match. We conclude that our contextual manipulation succeeding in canceling the scalar implicature.

3.3. Discussion

Experiment 2 had two primary findings. First, like other researchers we found that children failed to calculate scalar implicatures even when the context supported it (Hurewitz et al., in press; Papafragou & Musolino, 2003). Specifically, in the ALL vs. SOME trials, children split their choices between the box with SOME and ALL of the cookies, demonstrating a failure to use scalar implicature to restrict the reference of the definite description. Adults, on the other hand, overwhelmingly favored the implicature match. However, we also found that on the critical trials, when no clear implicature match was provided, both children and adults overwhelmingly selected the lower-bounded quantity (i.e. ALL of the cookies). This supports our conjecture that scalar implicatures are cancelled when the critical terms are used as part of a definite description in the presence of a salient and accessible lower-bounded alternative and the absence of a clear implicature match.

We interpreted our data in light of the prior literature on the interpretation of "some" and the development of scalar implicature, but it is possible that the children's response pattern could reflect much simpler strategies that were unique to this experiment. We explored three such alternatives by conducting a control study in which children were asked to select the box where Cookie Monster had "all of the cookies." The experiment was identical in all other respects to Experiment 2 and ten children between 2;6 and 3;5 participated. First, we considered the possibility that the children failed to understand that each box was to be evaluated in isolation and instead interpreted the quantifier with respect to the contents of all three boxes. This would lead them to see every intended instance of all as an exact match for some, and thus explain their willingness to accept implicature violations. In the all control study such a construal would make the experimenter's request impossible to honor, since on no trial did a single character possess all the visible cookies. However, the children's performance was quite systematic suggesting that they were often able to identify a single correct answer. For example, in the ALL vs. NONE trials, the children consistently selected the card where Cookie Monster had all the cookies and Big Bird had none, despite the fact that the Big Bird on the other card also had some cookies (M = 93%). Second, we examined the possibility that the children were simply ignoring the quantifier and picking a card in which the target character was associated with cookies. If this were the case, we would expect the children perform the same way in the control study as they had in Experiment 2. But they did not. On the ALL vs. SOME trials, children who were asked for "all" systematically selected the card in which Cookie Monster had all the cookies (93%), while children who were asked for "some" split their responses between the some and all cards ($t(18) = 3.29$, $p < .01$). Finally, this experiment allowed us to explore whether the children in Experiment 2 may have failed to select the covered box because they were unable to make inferences about its contents. If so we would expect that they would fail to select it even if the semantic requirements of the quantifier were incompatible with the visible sets. But in the SOME vs. NONE trials, the covered box was the modal response for children who were asked for "all" (M=52%), suggesting that many of them were able to make the inference that target must be in that box. The remainder of the children experienced more confusion, dividing their responses between the box where Big Bird had all of the cookies and Cookie Monster had none (18.5%) or the box where Cookie Monster had some of them (29.5%).[11]

In sum, the children's pattern of performance for "all" demonstrates that they are able to use semantic information about the quantifier to reliably select the correct box when it is visible, and sometimes even to infer its presence when it is absent. The results of Experiment 2, in conjunction with this control study, validate the critical trials of the covered-box task as a means of teasing apart semantic restrictions from pragmatic implicatures.

4. Experiment 3

As Experiment 2 demonstrates, our new task has the critical properties that are needed to distinguish between the Neo-Gricean and Exact theories of number semantics. Because scalar implicatures are cancelled in the critical trials and the sets under consideration are clearly defined, the alternate theories should map on to distinct response patterns, even in pragmatically sophisticated adults. In Experiment 3 we focused on just two of the four groups that were tested in Experiment 1: 2 and 3 year old children who are two-knowers and adults. These two populations are at opposite ends of the developmental continuum and showed the most categorical differences in performance in Experiment 1.

[11] We did not test adults in this control condition because their response patterns in Experiment 2 ruled out the first two alternate hypotheses: by systematically choosing the implicature match in the ALL vs. SOME trials the adults demonstrated that they were using the quantifier and evaluating it with respect to a single box. As we shall see, the adults' responses in Experiment 3 rule out the possibility that they are unable to imagine alternative in the covered box if their construal of the utterance demands it.

During the critical test trials, participants were shown a box with one fish, another box with three fish*,* and a covered box with an opaque cover, and were asked to "*Give me the box with two fish*." If numbers have an Exact semantics then participants should select the covered box, inferring from the definite description that a set of exactly two fish must be there. If, on the other hand, numbers have a lower-bounded semantics, then participants should select the box with three fish to satisfy the description, just as they selected ALL as an instance of the scalar quantifier *some* in the critical trials of Experiment 2.

4.1. Methods

4.1.1. Subjects

Thirty native English-speaking undergraduates from Harvard University and 10 English-speaking children between the ages of 2;6 and 3;5 (mean 3;0) participated in this experiment. Children were pre-screened with the give-a-number task (Wynn, 1990) and only those who met the criteria for being a two-knower were asked to participate in this study. Four children were excluded because they did not fit the two-knower criterion.

4.1.2. Procedure

This study was composed of two parts. In the Familiarization phase, participants were introduced to the box task using the procedure from Experiment 2. Only participants who selected correct boxes when the target animal was visible and selected the covered box when it was not were included in this experiment. One child was excluded for failing to meet this criterion. During the Test phase, participants were asked to *"Give me the box with two fish"* in the following contexts (see Figure 5):

1. CONTROL TRIALS: EXACT VS. LESS. During these three trials, participants were given a box containing two fish (exact match), a box containing one fish (upper-bounded option)

and a covered box. This condition corresponded to the "Some versus None" control trials in Experiment 2. Both the Neo-Gricean theory and Exact theory would predict that children and adults should consistently pick the exact match since it falls within the meaning of the term and the other visible option does not.

2. <u>CONTROL TRIALS: EXACT VS. MORE.</u> During these three trials, participants were given box containing two fish (exact match), a box containing either three or five fish (lower-bounded option) and a covered box. This condition corresponded to the "Some versus All" control trials in Experiment 2 and closely parallels the task used by Hurewitz et al. (in press). The Exact theory would predict that all participants should consistently select the exact match (since it is the only choice matching the meaning of the term). The Neo-Gricean theory makes the same prediction in any population that can calculate a scalar implicature. The presence of both an implicature match and a lower-bounded match creates a context in which the contrast set (two vs. three) is highlighted and the implicature is readily calculated (see Experiment 2).
3. <u>CRITICAL TRIALS: LESS VS. MORE</u>. During these four critical trials the exact match was absent. Participants saw a box containing one fish (the upper-bounded option); a box containing either three or five fish (the lower-bounded option) and a third covered box. In Experiment 2 we found that when an exact match was absent, both children and adults failed to use scalar implicature. Thus in this context the Exact and Neo-Gricean theories make distinct predictions. The Exact theory predicts that the participants will reject the visible choices, since they are not compatible with the meaning of expression, and will infer that the correct set is in the covered box. In contrast the Neo-Gricean theory predicts

that both adults and children will select the lower-bounded option since it is compatible with the semantics of the number.

Children received all three trial types in randomized order. However, for the adults, we adopted a between-subjects design with the rationale that the presence of exact matches on earlier trials might bias the adults towards an exact interpretation on the critical trials.

4.2. Results

We conducted our analyses on the proportion of covered box choices. In the critical trials, we interpret these choices as directly corresponding to an exact interpretation. Since the response patterns for the two types of control trials were statistically identical, we collapsed them. The control trials allow us to rule out the possibility that participants select the covered box based on its novelty or an alternate construal of the task. As Figure 6 illustrates, participants overwhelming preferred the visible exact match for both types of control trials ($M = 95\%$ and $M = 100\%$ for children and adults respectively). But on the critical trials, when there was no exact match, participants consistently selected the covered box ($M = 95\%$ and $M = 100\%$ for children and adults respectively). This resulted in a reliable effect of trial type; both children and the adults were far more likely to select the covered box on the critical trials ($W = 55$, $Z = 2.78$, $p < .005$; $W = 55$, $Z = 2.78$, $p < .001$, respectively). There were no reliable differences between children and adults for either the control trials ($W = 95$, $Z = .76$, $p > .2$) or the critical trials, ($W = 95$, $Z = .76$, $p > .2$). Thus even at this early stage of acquisition, the children's interpretation of number words closely mirrors that of adults.

A comparison of Figures 4 and 5 suggests that participants responses for "two" were quite different than their responses for "some". This was confirmed in a 2 x 2 ANOVA examining the proportion of covered box choices by adults and children for the critical trials of

Experiments 2 and 3 (those trials in which the exact match was absent). While we found no main effect of age ($p > .10$) and no significant interaction between age and experiment ($p > .10$), we did find that subjects selected the covered box significantly more in Experiment 3 than Experiment 2, $F(1, 36) = 192.24$, $p < .001$. These results strongly suggest that semantic interpretations of number words differ from that of true scalar terms.

4.3. Discussion

The results from Experiment 3 provide a stark contrast to participants' performance with the scalar quantifier *some*. Here we found that participants overwhelmingly selected the exact match when it was visible and selected the covered box when it was not. This pattern of performance clearly supports the Exact Semantic proposal and is inconsistent with a Neo-Gricean account. Since the task was to find a box that satisfied the definite description ("the box with two fish"), a theory of lower-bounded semantics predicts that participants could have selected the larger quantity as a satisfying alternative without ever needing to consider the covered box. Instead participants rejected the visible lower-bounded option and inferred that the covered box must have two fish in it. However, when there was a visible exact quantity match, they had no difficulty ignoring the covered box and selecting this item.

These results are problematic for the Neo-Gricean account. On the Neo-Gricean theory, the rejection of the larger or lower-bounded target can only be explained as the effect of an upward-bounding scalar implicature. There are two features of this study which make such an explanation unlikely. First, the children that we tested in Experiment 3 gave exact interpretations of *two* even though they seemed to have little knowledge of the meaning of the word *three*. On the Neo-Gricean account, scalar implicature is motivated by mutual knowledge of the terms on the scale and their relative informational strength. Second, on the critical trials participants were tested in a context in which scalar implicatures were cancelled. In Experiment 2 we demonstrated that when a scalar term is used as part of a definite description in the presence of a salient lower-bounded match both children and adults will fail to calculate the implicature and accept the lower-bounded match (all). When we presented number words in a parallel context, our participants consistently rejected the larger set, suggesting that the upper-bound was a semantic requirement rather than a pragmatic preference.

5. General Discussion

This study sought to provide evidence to differentiate between semantic and pragmatic contributions in the interpretation of number words. Consistent with the Exact Semantics account, we found that children interpreted words like *two* as referring to an exact quantity at the earliest stage of acquisition. Like adults, they were able to reject salient lower-bounded targets and used the number word to infer the presence of an exact match elsewhere in the array (i.e. in the covered box). This contrasts with performance with a true scalar term like *some*. Here, we found that both children and adults readily selected ALL as an example of *some* when there was no other visible alternative.

These results are consistent with prior studies that find a division between children's interpretation of number words and their interpretation of scalar quantifiers (Papafragou & Musolino, 2003; Heurwitz et al., in press). In addition, they resolve crucial questions that were unanswered in the prior work. In the earlier studies participants were tested in contexts in which scalar implicatures were robust (Papafragou & Musolino, 2003; Hurewitz et al, in press), Thus the adults' preference for exact interpretations was compatible with both the Neo-Gricean and Exact theories. Similarly the children's preference for exact interpretations of numbers and lower-bounded interpretations of other scalars could reflect differences in lexical semantics or the precocious acquisition of pragmatic implicatures with number words (see Papafragou & Musolino, 2003). In contrast, we tested children and adults in a context in which scalar implicatures were cancelled (the critical trials of Experiments 2 & 3), thus we can rule out the possibility that their consistent use of the exact interpretation was achieved by combining a lower-bounded semantics with an upward-bounding implicature. In sum our findings are problematic for the Neo-Gricean account and strongly suggest that both children and adults assign exact meanings to number words.

Methodologically, the covered box task used in Experiment 2 and 3 may be useful for mapping the semantic boundaries of a variety of words and phrases. Typical selection tasks create a strong demand on participants to select at least one of the choices presented. This can be problematic in situations where the status of atypical category members or pragmatically infelicitous interpretations is under investigation. If more typical exemplars are included in the set of choices, then participants may be reluctant to choose less typical exemplars. This pattern could reflect the true extension of the term, or it could indicate that the participant has redefined their task as choosing the best exemplar rather than all possible exemplars. This problem may be magnified in young children who face limitations of memory and attention and thus may fail the select the exhaustive set. Alternately, if only marginal exemplars are included, participants may feel compelled to choose one of them and select an item outside of the true extension, redefining their task as finding the alternative that is "most like" the description. For example, in Experiment 1, two-knowers systematically chose the lower-bounded option when no exact match was available, even though their actions (pointing to a subset of the objects) and their performance in the subsequent experiment indicated that they assigned an exact semantics to the number word. Experiments 2 and 3 suggest that when participants are given a definite description and allowed to select an option that they cannot see, they will chose to do so only when none of the visible options match the meaning of the description. Thus the covered box task allows experimenters to test the extension of a description without making any a priori assumptions about the status of atypical exemplars (for an additional application see, Li & Huang, 2005).

The remainder of this discussion focuses on two additional issues that are central to understanding the semantics and acquisition of number words. First, we explore how our data bear upon a third theory of number word interpretation, Carston's Underspecified Theory (1985). Second, we examine how the existence of alternate interpretations for number words might affect the acquisition of their exact meanings.

5.1 Could the Semantics of Number Words be Underspecified?

We have considered two possible hypotheses about the role of semantics and pragmatics in number word interpretation: numbers could have lower-bounded meanings, with pragmatic supplementation leading to exact interpretation, or they could have exact meanings, with pragmatic factors giving rise to apparent lower-bounded readings.[12] Carston (1985 & 1998) offers a third alternative: number words could have an underspecified meaning which is supplemented by a process of general pragmatic inference and enrichment. Working within the framework of Relevance Theory (Sperber & Wilson, 1986), she suggests that the meanings of number words are "neutral among the three interpretations, AT LEAST N, AT MOST N, EXACTLY N,

[12] We do not specifically address cases where number words are used imprecisely to approximate quantities of an indeterminate set (e.g., A farmer says, "There are 500 chickens" when there are actually 502). Dehaene & Mehler (1992) suggest that speakers' use "round numbers" in the presence of large quantities since their mental representations are less accurate. If so, semantic imprecision in this case is one of a more general phenomenon occurring when classification principles do not permit sharp divisions between categories, e.g. discrimination of neighboring colors (Chierchia & McConnell-Ginet, 1990). Others have suggested that the use of numbers to approximate quantities reflect sensitivity to the varying degrees of precision that are relevant in specific situations (Lasersohn, 1999; van der Henst, Carles, & Sperber, 2002).

so that they don't have any one interpretation until they are placed in a particular sentential context" (Carston, 1998, pg. 15). Thus the meaning of *two* could be represented as (10), where the variable X represents a function that is selected by pragmatic inference and is necessary to complete the interpretation of the number word.

(10) [X [TWO]]

We did not directly address this alternative in our studies for two reasons. First theoretically, our goal was to examine the semantic representations of number words while the underspecified theory conversely describes a process of pragmatic enrichment. In (9), the lexical meaning of the word *two* is defined as an unspecified function on the concept TWO but little is said about the content of this embedded concept. Second methodologically, we attempted to disentangle alternate theories of number word semantics by controlling the factors that would make the exact and lower-bounded readings permissible under each theory. The underspecified theory claims that both interpretations arise from a common process where listeners fill in the function with the operator that is most relevant in that context. In the absence of an independent metric of relevance, it was not clear how these predictions could be distinguished from those of other accounts.

Nevertheless, some features of this data are problematic for the underspecified account. In Experiment 1, lower-bounded interpretations were salient and clearly relevant to the task. Yet the adults and the more numerically sophisticated children stepped outside the task and rejected the experimenter's question. Given the obvious cognitive and communicative effort involved, it's hard to imagine how the underspecified theory could capture this result without granting privilege to the exact reading. In addition, in Experiments 2 and 3 we found substantially different patterns of interpretation for *two* and *some*, despite the strong similarities between the communicative goals and contexts in the two situations. While Carston (1998) suggests that the saturation and enrichment processes involved in the interpretation of number words are distinct from the conversational implicatures involved in interpreting scalars (the former affect truth-values, the later do not), both processes are argued to be part of monostratal pragmatic processing system, governed by the single principle of relevance (Carston, 1998 & 2004). If the absence of an "exact" match makes the lower-bounded interpretation sufficiently relevant for interpretations of *some*, then why is it not sufficient to bias selection of the at-least function for *two*?

At first glance the Underspecified theory appears to have one advantage over its rivals: it readily accounts for situations in which numbers adopt an upper-bounded (at-most) interpretation. For example, in a statement like (11), the inference that a person on this diet could also eat one candy bar has led many to argue that in cases like these *two* is interpreted as "at most two".

(11) You can eat two candy bars every day and still loose weight.

Both the Neo-Gricean and Exact accounts lack the theoretical resources to derive this interpretation. On the Neo-Gricean account the lower boundary is lexically-encoded and therefore inviolable. Breheney (2004) & Koenig's (1991) accounts can explain how an Exact Semantics can be compatible with larger sets, but their machinery would not work in reverse. Thus on these accounts, *two* in (11) cannot refer to smaller quantities because maximal sets of one candy bar could never include a subset of two.

The need to account for at-least and at-most readings has been amplified by recent research demonstrating that adults and even five-year-old children readily derive lower- and upper-bounded interpretations of number words (Musolino, 2004).[13] Both children and adults readily assented to sentences like (12) when the Troll missed just one hoop (upper-bounded or at-most readings) and sentences like (13) when Goofy had a total of four cookies (lower-bounded readings).

(12) Goofy said the Troll could miss two hoops and still win the coin. Does the Troll win the coin?

(13) Let's see if Goofy can help the Troll. The Troll needs two cookies. Does Goofy have two cookies?

Since utterances like (13) are readily accounted for by the Exact theories (see section 1.2), we will focus our attention on utterances like (11) and (12). Note that these non-exact interpretations, unlike the subset readings like (13), do not refer to actual events. Instead they make claims about the status of possible events, thus there is no real-world referent for the cardinal number phrase, and no actual set whose numerosity we can examine. Listeners must evaluate the interpretation of these utterances by introducing situations and examining how people apply the rule or condition set forth in the utterance to those situations. But it is not clear that these inferences provide direct information about how the quantified noun phrase itself is interpreted. This can be illustrated by looking at parallel cases that do not involve numbers.

In a sense, all of the statements in (14) have an upper-bounded reading; in each case, the Troll would be declared the victor even if he missed none of the hoops. Similarly the sentences in (15), have lower-bounded readings (the Troll would be the victor even if he got all the hoops around the post).

(14) a. The troll can miss the red and blue hoops and still win

[13] In Musolino (2004) these studies are not presented as evidence of underspecified semantics, but rather as evidence that children have access to the same range of interpretations of numbers as adults. In fact Musolino concludes that numbers are likely to have an exact semantics.

b. The Troll can miss all of the hoops and still win.

c. The Troll can quit the game right now and still win.

(15) a. The Troll must put the red hoop around the post to win.

b. The Troll must put a hoop around the post to win.

c. The Troll must put the large hoop around the post to win.

But these examples do not lead us to conclude that "red" has an underspecified meaning, which can be cashed out as RED, BLUE, RED AND BLUE, or NOTHING or that "the large hoop" sometimes means ALL OF THE HOOPS. Instead it seems more reasonable to conclude that in all these cases, predicates continue to have their standard interpretations but also establish a specific boundary in the context of modal statements. Similarly in (11) and (12), the number word *two* identifies a point on a scalar continuum that would constitute success (i.e. "winning the game" or "loosing weight") while the modal verb (*can* or *need*) in conjunction with the context guides listeners in their inferences of how to divide up related events on either side of this boundary. Like others (Harnish, 1976; Koenig, 1991), Scharten (1997) argues that "[M]odal predicates like *need, must, necessary, requirements* and *criteria* indicate that the amount specified constitutes a minimum requirement . . . Numerals which are in the scope of a modal predicate can thus be seen to receive an interpretation which cannot be accounted for without taking the semantics of the modal predicate into account" (pg. 60).[14] If this analysis is correct, it suggests that the young

[14] Note that we are not disputing the systematicity of these inferences. Nor are we arguing that these phenomena fall outside the purview of pragmatics or semantics. We are merely suggesting that in all of these cases cardinal number phrases continue to refer to sets of an exact numerosity. Interestingly, Carston (1998) argues that the upper-bounded (at-most) interpretation is a core phenomenon for understanding numbers, but concludes that exact semantics is a viable contender. She suggests that the semantics of modal operators might weaken the exact semantics of cardinal numbers. In contrast we are arguing that the meaning of the cardinal number stays the same in all cases. However the modal operator (in conjunction with the context) enriches interpretation, contributing the additional content that distinguishes the at-least and at-most readings. While context is clearly playing a role in interpretation, we remain agnostic about whether this additional content is contributed in compositional semantics, as an explicature, or as an implicature. For example "You can eat three green beans" might have a lower-bounded reading if food is scarce but

children in the Musolino (2004) study were able to make complex and flexible inferences about how to apply this boundary to novel circumstances. While this may seem surprising in light of previous research supporting children's poor performance on pragmatic tasks, recent work by Papafragou & Tantalou, (2004) suggests that five-year-olds are quite adept at making ad hoc inferences of this kind. For example, when given a statement like "*I ate the cheese,*" children were able to correctly infer that the character had not eaten the rest of the sandwich.

5.2 How can two interpretations lead to the acquisition of a single meaning?

In many ways, the child's task in acquiring word meaning is similar to the linguist's task of determining semantics. Like the linguist, the child only has access the final output of interpretation and needs to work backwards to establish which aspects of meaning are assigned at the semantic level and which aspects are assigned at the pragmatic level. The presence of two interpretations could potentially create as much confusion for the child as it has for theorists. If the evidence present in the linguistic input is consistent with both the Neo-Gricean and Exact Semantics account, how can the child resolve whether lower-bounded interpretations should be attributed to set decomposition or to the cancellation of a scalar implicatures?

There are several ways in which this ambiguity might be resolved. First, independent of the input, one set of number concepts could be privileged in the mind of the learner. Thus numbers could have exact meanings because the lower-bounded interpretations are simply less salient, or not available, at the time when number words are learned. A second possibility is that the input provides strong support for the exact reading. If cases of unambiguous non-exact readings are vanishingly rare there would be little reason for children to consider lower-bounded hypotheses. While a true test of this hypothesis is beyond the scope of this article, an informal

---

an upper-bounded reading if the addressee is a young child negotiating at dinner. But context-sensitivity tells us only that a completely encapsulated bottom-up semantic analysis will be inadequate.

examination of the CHILDES database (MacWhinney, 2000) provides some initial support for this conjecture.[15] We limited ourselves to looking at contexts in which the exact numerosity of the set was explicitly mentioned or inferable based on world knowledge. In all of these cases, we found that the parents used cardinal numbers to refer to the exact numerosity of maximal sets like in (16) and (17).

(16) Mother: there are his ears

Adam: four

Mother: two

Mother: and two eyes

[Adam file 09, 2;7 years]

(17) Mother: one letter

Eve: one

Mother: another one over there?

Eve: yeah

Eve: one

Mother: you have two letters

Mother: one, two letters (mother counts letters)

[Eve file 04; 1;7]

Additionally, during the process of acquisition, the frequent recitation of the count sequence may serve to entrench the notion that number words refer to unique points on a continuum. Our

---

[15] A full test of this hypothesis would require a corpus of parental speech that was annotated with information about the exact numerosity of each set under discussion. Instead we searched a subset of the Adam, Eve, and Sarah transcripts for the word "two" (Brown, 1973). We eliminated all cases in which the number word was used as part of the count sequence, where it referred to an abstract entity ("two minutes") or where the numerosity of the set was not explicitly mentioned in context notes, established by counting, or readily inferable from world knowledge. We coded the first ten such uses of "two" in the input and by the child. In all 30 cases the adult used the number word to refer to the maximal set. Children made many errors which are discussed in the text.

perusal of CHILDES also suggested that children are also given explicit evidence that lower-bounded interpretations are generally unacceptable. As the examples in (18) and (19) suggest adults are quick to correct children when they use smaller numbers in the presence of larger quantities.

(18) Colin: how many crackers do you have?

Eve: have two cracker

Mother: how many crackers?

Mother: more than two.

[Eve file 05, 1;8 years]

(19) Adam: two garage

Mother: two in the garage?

Adam: yeah

Mother: you have three.

[Adam file 07, 2;6 years]

Thus we suggest that unambiguous input for exact readings, perhaps in conjunction with a bias for a simple mapping from interpretation to lexical meaning, may lead children to attribute exact interpretations to semantics rather than to scalar implicature.[16]

[16] Some readers may be surprised to see parental feedback suggested as a possible information source for language acquisition. There are two strong reasons for believing that feedback plays a minimal role in acquisition, neither of which applies to this particular case. First, there is ample evidence that feedback on the syntactic and phonological properties of utterances is patchy at best, undermining its value to the learner (see arguments in Marcus, 1993). In contrast our CHILDES analysis suggests that feedback on the incorrect usage of number words may be quite robust. This is not surprising; conversational partners presumably treat errors which affect meaning quite differently from those which do not. In their seminal study of parental feedback, Brown and Hanlon (1970) observed that while syntactic errors went unnoted, errors affecting the truth of the utterance were promptly corrected. Second, many theorists have noted that even systematic negative feedback would be insufficient to explain language acquisition without positing internal mechanisms for grouping the data appropriately and generating hypotheses. We agree. We are not arguing that feedback could replace these cognitive structures; we are merely suggesting that it could be used to decide between two internally generated hypotheses.

In conclusion, we have created a task in which scalar implicatures are cancelled, allowing us to disentangle semantic and pragmatic contributions to meaning. This task establishes a clear disassociation between the interpretation of number words and true scalar quantifiers, and it offers a potential alternative to judgment tasks for probing semantic interpretation. Our experiments using this task demonstrate that number words have exact semantics from the very earliest stage of acquisition. These findings help to resolve a longstanding controversy in linguistics, and they validate a key assumption underlying much of the current developmental work on number word learning.

Acknowledgments

We would like to thank Julien Musolino for introducing us to these issues and Susan Carey for her persistent questions which pushed us to tighten up both our arguments and our experiments. This work benefited from conversations with Anna Papafragou, Gennaro Chierchia, Steven Pinker, Jeff Lidz, the members of the Laboratory for Developmental Studies, MIT Hungry Language Club, and Harvard Cognition, Brain, and Behavior seminar. We are grateful to Sasha Yakhkind for her assistance in recruitment and testing.

References

Bloom, P. (1994). Generativity within language and other cognitive domains. *Cognition*, 51, 177-189.

Bloom, P. & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, 24, 511-533.

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.

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

Briars, D. & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, 20, 607-618.

Breheny, R. (2004). Some scalar implicatures aren't quantity implicatures—but some are. *Proceedings of the 9th annual meeting of the Gesellschaft fur Semantik*, (Sinn und Bedeutung VIIII), University of Nijmegen, Nov. 2004.

Brown, R. (1973). *A First Language*. Cambridge, MA: Harvard University Press.

Brown, R., Hanlon, C., 1970. Derivational complexity and order of acquisition in child speech. In: Hayes,J.R. (Ed.), *Cognition and the Development of Language.* Wiley, New York.

Bultinck, B. (2005). Why Paul Grice should have been a corpus linguist: An analysis of two. *Paper presented at the 9th International Pragmatics Conference* at Riva del Garda, Italy, July 2005.

Carston, R. (1985). A reanalysis of some "quantity implicatures." Unpublished ms., University College London.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In: R. Carston & S.

Uchida (eds.). *Relevance Theory: Applications and Implications*, 179-236. Amsterdam: John Benjamins.

Carston, R. (2004). Relevance theory and the saying/implicating distinction. In: L. Horn & G. Ward (eds.). *Handbook of Pragmatics*. Oxford: Blackwell.

Chierchia, G. & McConnell-Ginet, S. (1990). *Meaning and grammar: An introduction to semantics*. Camrbidge, Mass: MIT Press.

Chierchia, G. (2000). Scalar implicatures and polarity items. Paper presented at *NELS 31*, Georgetown University, Washington, D.C.

Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domingues, & A. Johansen, (Eds.), Proceedings of the 25th Boston University Conference on Language Development (pp. 157-168). Somerville, MA: Cascadilla Press.

Dehaene, S. & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1-29.

Fox, D. & Hackl, M. (2005). The Universal Density of Measurement. *Unpublished manuscript*. MIT, Cambridge, MA.

Fuson, K. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

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

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306, 499-503.

Geurts, B. & Nouwen, R. (2005). At least et al. The semantics of scalar modifiers. *Unpublished manuscript*. University of Nijmegen.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377-88.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and Semantics*, Vol. 3. (41-58). New York: Academic Press.

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

Harnish, R. (1976). Logical form and implicature. In T. Bever et al. (Eds.), *An Integrated Theory of Linguisitic Ability*, 313-391. New York: Harvester Press.

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (in press). Asymmetries in the acquisition of numbers and quantifiers. To appear in *Language Learning and Development.*

Horn, L. (1972). *On the semantic properties of the logical operators in English*. Doctoral dissertation, UCLA, Los Angeles, CA. Distributed by IULC, Indiana University, Bloomington, IN.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Huang, Y. & Snedeker, J. (2005). What exactly do numbers mean? Paper presented at the conference on *Experimental Pragmatics: Exploring the Cognitive Basis of Conversation.* April 2005. University of Cambridge. Cambridge, United Kingdom.

Katsos, N., Breheny, R., & Williams, J. (2005). Are scalar implicatures generated by the grammar? *Proceedings of the 28th GLOW in Geneva*.

Koenig, J. (1991). Scalar predicates and negation: punctual semantics and interval interpretations, *Chicago Linguistic Society* 27, Part 2: Parasession on negation, 140-155.

Kratzer, A. (2003). Scalar implicatures: Are there any? Paper presented at the workshop on *Polarity, Scalar Phenomena, and Implicatures*. June 2003. University of Milan-Bicocca.

Krifka, M. (1999). At least some Determiners aren't Determiners, in K. Turner (ed.), *The Semantics/Pragmatics Interface from different points of views*. Elsevier Science B.V.: 257-291.

Lasersohn, (1999). Pragmatic halos. *Language*, 75, 522-551.

Le Corre, M. & Carey, S. (in prep). How the counting principles are constructed out of core knowledge of number. *Unpublished manuscript*. Harvard University, Cambridge, MA.

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (in press). Revisiting the competence/performance debate in the acquisition of counting as a representation of the positive integers. To appear in *Cognitive Psychology*.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

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

Li, P. & Huang, H. B. (2005). A small fan and a small handful of fans: the acquisition of count-mass classifiers. *Proceedings of the 2005 biennial meeting of the Society for Research on Child Development*, Atlanta, GA, April 7-10.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, G.F. (1993). Negative evidence in language acquisition. *Cognition* 46, 53–85.

Mix, K. S., Huttenlocher, J., Levine, S. C. (2002). *Quantitative development in infancy and early childhood.* New York: Oxford University Press.

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition*, 93, 1-41.

Musolino, J. & Lidz, J. (in press). Why children aren't universally successful with quantification. To appear in *Linguistics*.

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

Noveck, I. A. & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85, 203-210.*

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.

Paris, S. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, 16, 278–291.

Pica. P. Lemer, C., Izard, V., & Dehaene, P. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306, 441-443.

Recanati, F. (2003). *Literal Meaning*. Cambridge: Cambridge University Press.

Saddock, J. (1984). Whither radical pragmatics? In D. Schiffrin (ed.), *Meaning, form and use in context*. Washington: Georgetown University Press, 139-149.

Sarnecka, B. (2003). *Language as lens: Plurality marking and numeral learning in English, Japanese, and Russian*. Doctoral dissertation, University of Michigan, Ann Arbor.

Scharten, R. (1997). *Exhaustive interpretation: A discourse-semantic account*. Doctoral dissertation, Katholieke Universiteit, Nijmegen.

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

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Storto, G., Tannenhaus, M, & Carlson, G. (2005). The calculation of scalar implicatures: Evidence from eye-tracking. *Paper presented at the annual meeting of the Linguistic Society of America.* Oakland, CA, January 2005.

van der Henst, J., Carles, L., & Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind and Language*, 457-466.

van Rooy. R. & Shulz, K. (2004). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Unpublished manuscript*. University of Amsterdam.

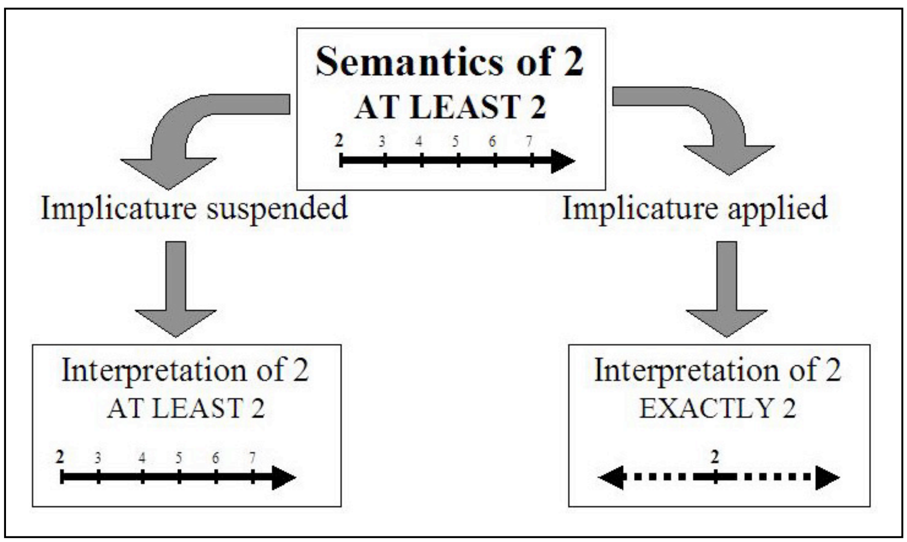
Winter, Y (2001). *Flexibility principles in boolean semantics*. Cambridge, Mass: MIT Press.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155-193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220-251.

Figure 1: The Neo-Gricean and the Exact Semantics account are able to explain both EXACT and AT LEAST interpretations of number words. (a) The Neo-Gricean account states that number words have lower-bounded semantics and receive exact interpretations via scalar implicatures. (b) The Exact Semantics account states that number words have exact meaning with apparent lower-bounded interpretations attributed to pragmatic set decomposition.

(a)



(b)

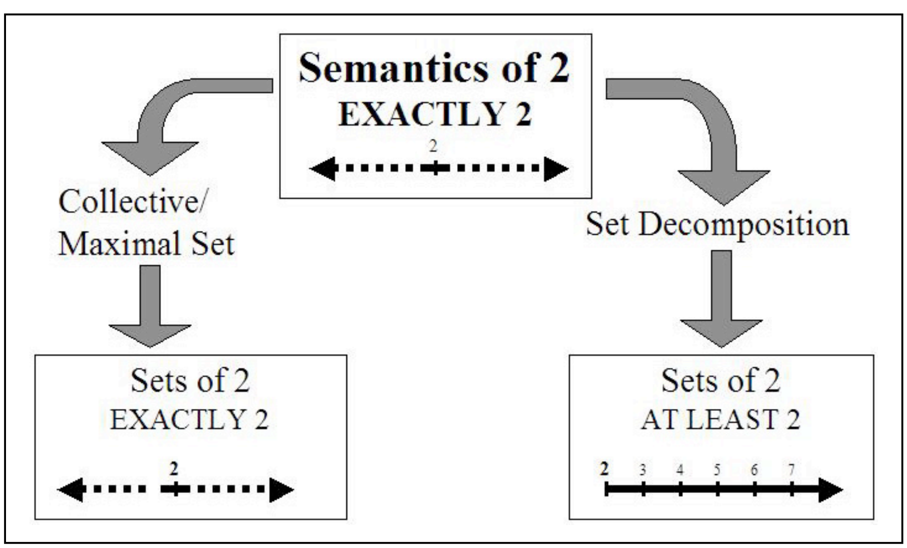


Figure 2: In Experiment 1, two- and three-knowers in critical trials selected cards with larger quantities and four-knowers and adults rejected both card choices.

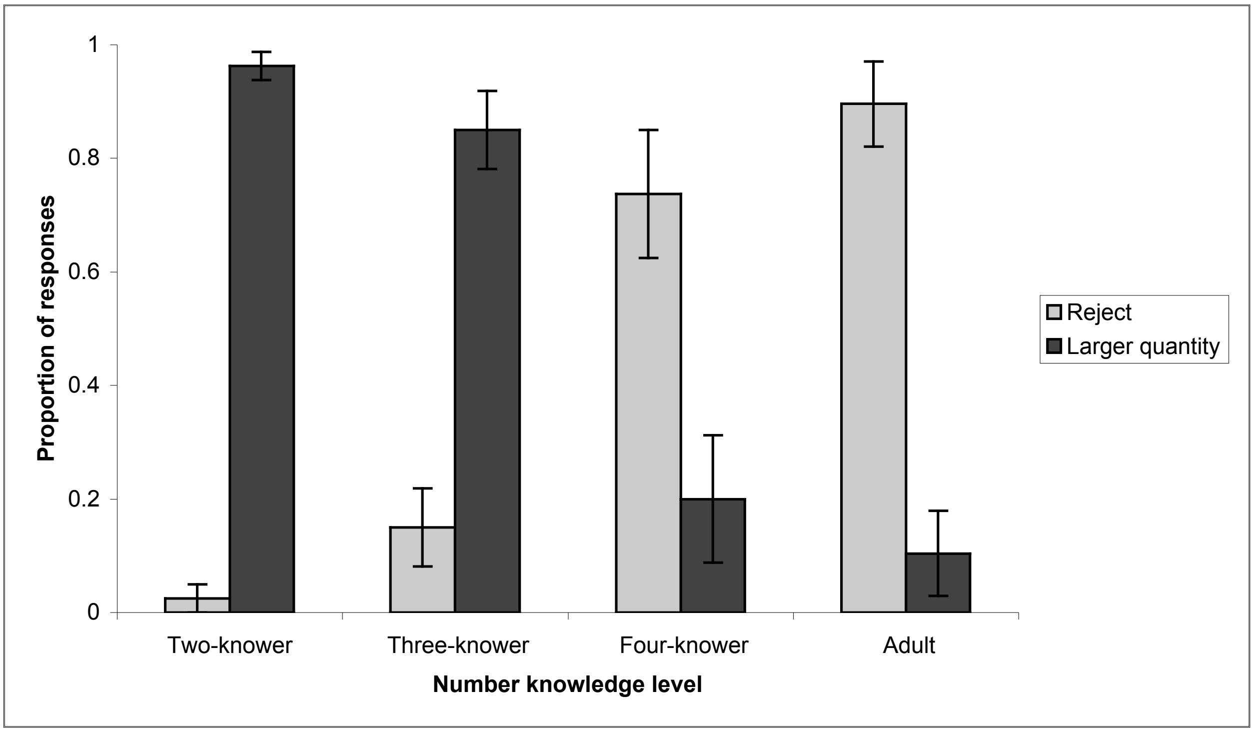
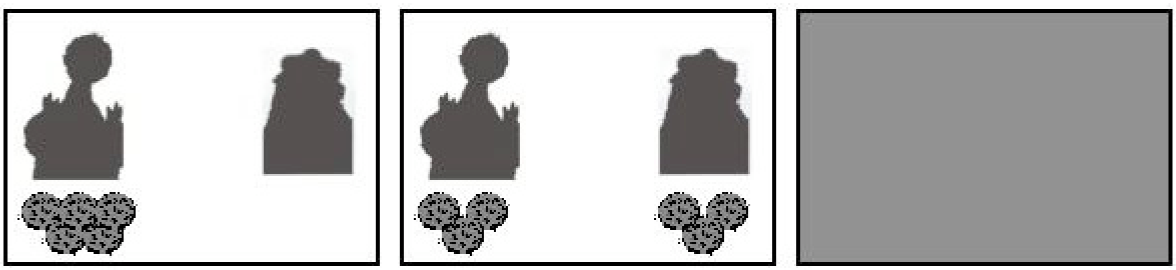
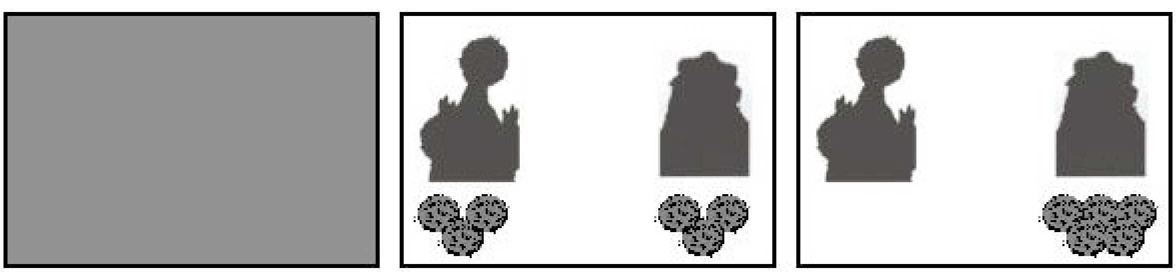


Figure 3: In Experiment 2, participants were presented with boxes where Cookie Monster (on the right) and Big Bird (on the left) possessed cookies and asked to "Give me the box where Cookie Monster has some of the cookies" in (a) NONE vs. SOME control trials, (b) ALL vs. SOME control trials, and (c) the critical NONE vs. ALL trials.

(a)



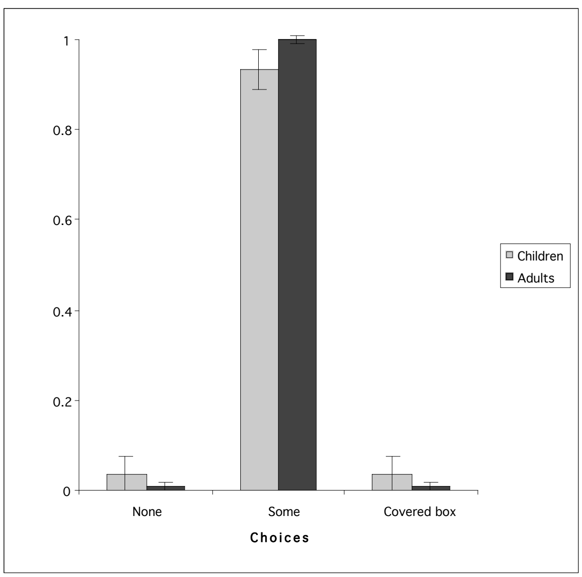
(b)



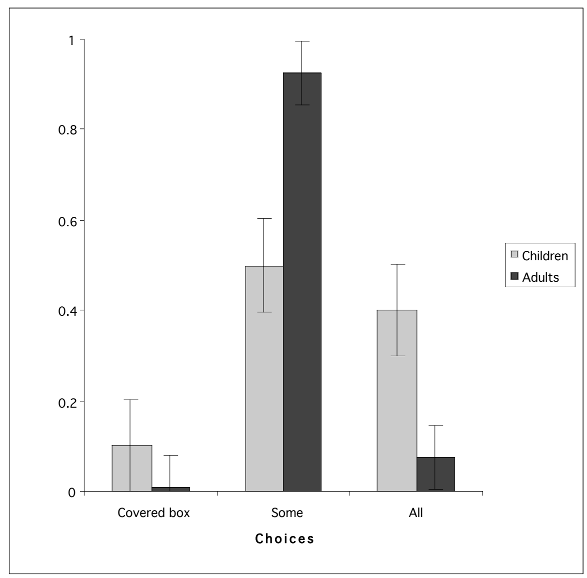
(c)

Figure 4: In Experiment 2, (a) In NONE vs. SOME control trials, participants selected the box with SOME of the cookies. (b) In ALL vs. SOME control trials, adults selected the box with SOME of the cookies while children were equally selected either the box with SOME or ALL of the cookies. (c) In the critical NONE vs. ALL trials, participants selected the box with ALL of the cookies.

(a)



(b).



(c)

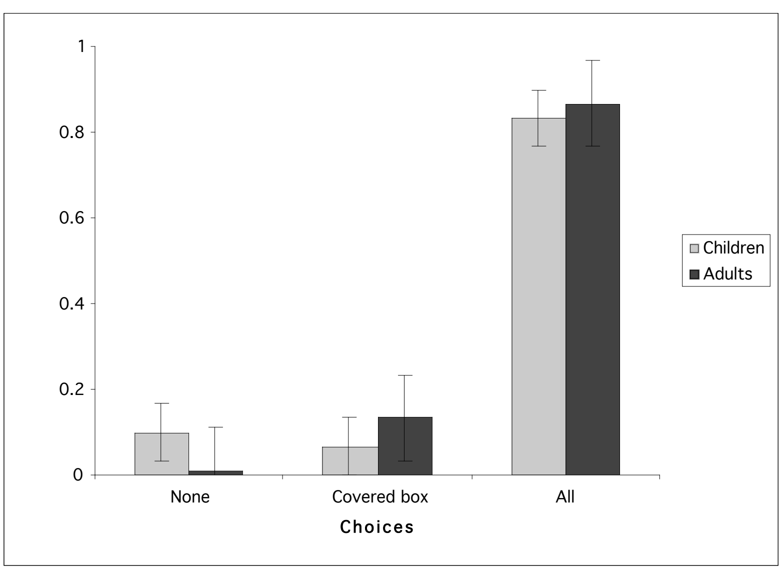
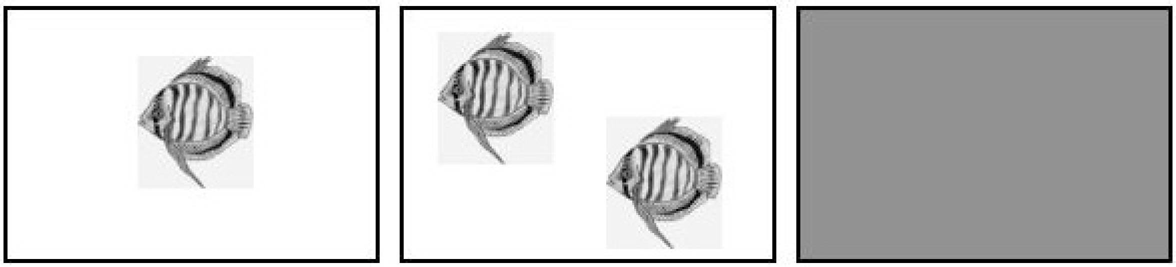
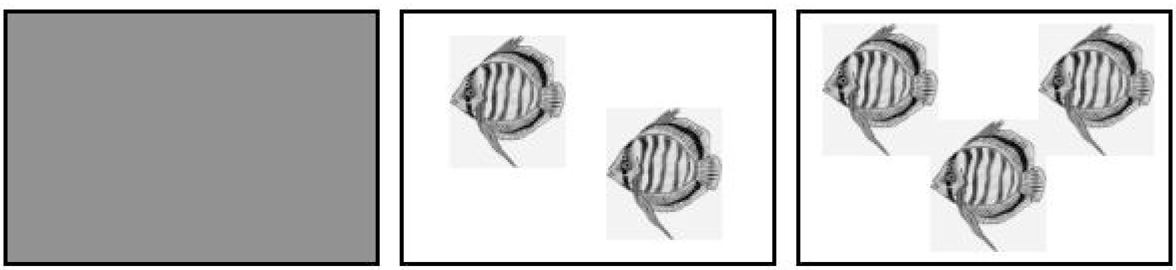


Figure 5: In Experiment 3, participants were asked to *"Give me the box with two fish"* for (a) EXACT VS. LESS control trials, (b) EXACT VS. MORE control trials, and (c) the critical MORE VS. LESS trials.

(a)



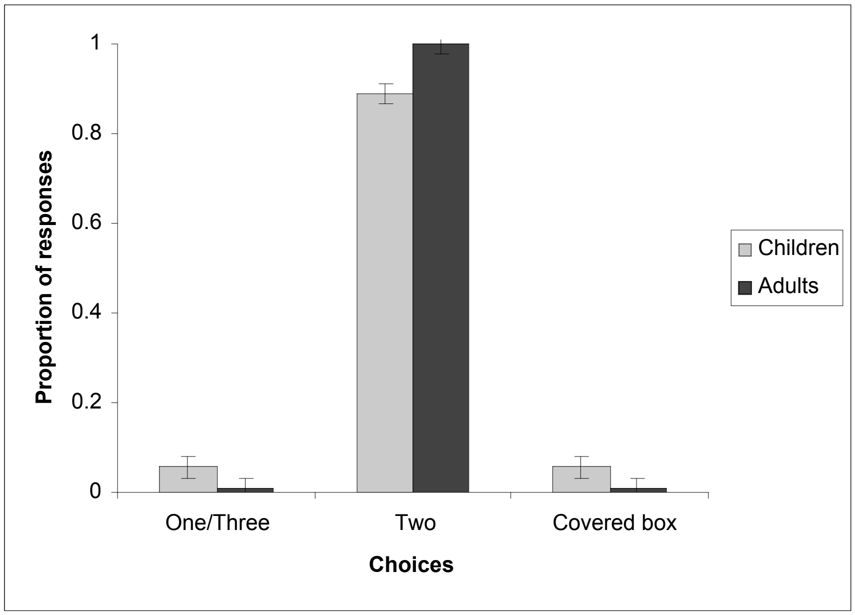
(b)



(c)

Figure 6: In Experiment 3, (a) in control trials, participants selected the exact match. (b) In the critical trials, participants selected the covered box.

(a)



(b)

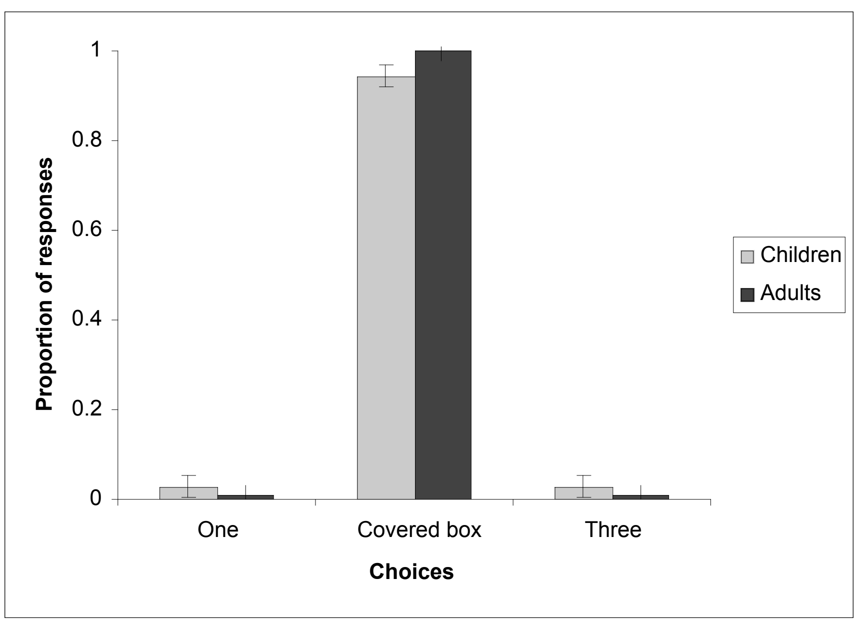


Table 1: In Experiment 1, participants were asked to select quantities in the presence or absence of an exact match. Trials were modified depending on participants' performance on the "Give-a-number-fish" task in the Pretest phase. Participants were asked to "Give me the card with X fish" where X is the number word noted in parentheses.

Knowledge Level	Critical Trials *(Less vs. More)*	Control Trials *(Exact vs. More)*	Control Trials *(Exact vs. Less)*
Two Knower	• ONE vs. THREE (2) • ONE vs. FIVE (2)	• TWO vs. THREE (2) • TWO vs. FIVE (2)	• TWO vs. ONE (2)
Three Knower	• ONE vs. THREE (2) • ONE vs. FIVE (3) • TWO vs. FOUR (3)	• TWO vs. THREE (2) • TWO vs. FIVE (2) • THREE vs. FOUR (3) • THREE vs. SIX (3)	• TWO vs. ONE (2) • THREE vs. ONE (3) • THREE vs. TWO (3)
Four Knower	• ONE vs. THREE (2) • TWO vs. FOUR (3) • THREE vs. FIVE (4)	• TWO vs. THREE (2) • TWO vs. FIVE (2) • THREE vs. FOUR (3) • FOUR vs. SIX (4)	• TWO vs. ONE (2) • THREE vs. ONE (3) • THREE vs. TWO (3) • FOUR vs. TWO (4)