Numerals denote degree quantifiers: Evidence from child language

Christopher Kennedy and Kristen Syrett

University of Chicago and Rutgers University

DRAFT: April 26. 2019

Abstract A large body of work in both the theoretical and experimental literature suggests that upper bound implications in simple sentences with bare numerals are entailments arising from the semantics of the numeral, rather than implicatures of the sort associated with other scalar terms. However, not all semantic analyses of numerals make the same predictions about upper bound implications in all contexts. In particular, in sentences in which numerals are embedded under existential root modals, only a semantic analysis of numerals as scope-taking degree quantifiers derives upper bound implications as entailments; other analyses must derive upper bound implications as implicatures. In this paper, we provide an argument for the degree quantifier analysis by demonstrating that young children interpret such sentences as imposing upper bounds at an age at which they do not reliably calculate scalar implicatures. We then consider an alternative account of the child language data in terms of exhaustification and numeral scale salience, and argue that here too, the experimental results favor the degree quantifier analysis.

1 Numerals and bounding implications

1.1 Two semantic accounts of upper bounds

It is well-known that numerals can be understood as imposing different bounding constraints on the quantities that they pick out in different contexts of use. For example, the numeral *three* in (1) is most naturally understood as providing both a lower and an upper bound on the number of hits that Mookie got on the last day of the season; i.e., (1) is taken to mean that Mookie got exactly three hits.

(1) Mookie got three hits on the last day of the season.

In (2), on the other hand, *three* is heard to provide only a lower bound on the number of hits Mookie has to get in order to win the batting title: he won't win with fewer than three; he will win with three or more.

(2) Mookie has to get three hits on the last day of the season in order to win the batting title.

The classic, neo-Gricean analysis of the upper bounded (or "two-sided") interpretation of (1) is due to Horn (1972, p. 33), who argues that sentences containing numerals "assert lower boundedness — *at least n* — and given tokens of utterances containing cardinal numbers may, depending on the context, implicate upper boundedness — *at most n* — so that the number may be interpreted as denoting an exact quantity." More specifically, on this view (1) is asymmetrically entailed by alternative sentences in which *three* is replaced by a numeral that introduces a higher value (*four, five*, etc.). The cooperative speaker's failure to use a stronger alternative, in apparent violation of the Maxim of Quantity, can be justified by the assumption that doing so would clash with the Maxim of Quality's injunction against saying that which the speaker believes to be false (or lacks evidence for), which in turn derives the upper bound as an implicature (Grice 1975).

On this analysis, the fact that (2) is not heard as imposing an upper bound on the number of hits Mookie has to get is actually expected, given the use of the universal modal *have to*. The same reasoning that derives the implicature that Mookie didn't get four (or more) hits from an utterance of (1) derives the implicature that Mookie doesn't have to get four (or more) hits to win the batting title from an utterance of (2), which is of course not inconsistent with him getting four (or more) hits. Unfortunately, as pointed out by Geurts (2006), this reasoning fails to account for examples just like (2) in which the numeral does appear to introduce an upper bound, such as (3).[1]

(3) Mookie has to get three hits on the last day of the season in order to finish with a batting average of precisely .345.

Since Mookie's batting average is monotonically related to the number of hits he gets, the combination of semantic content plus expected implicature here gives the wrong results: (3) is not understood to mean that Mookie doesn't have to get four

[1]Two comments. First, this kind of example can be explained by a "grammatical" analysis of scalar implicature, in which upper bounding implicatures are derived compositionally by introducing a silent, alternative-sensitive exhaustification operator in the syntax, as in Chierchia 2006; Fox 2007; Chierchia, Fox, and Spector 2012; Spector 2013; etc. (Though see Kennedy 2013 for discussion of other examples in which the grammatical theory fares no better than the neo-Gricean theory in predicting patterns of upper boundedness with numerals.) For the most part, the choice between a traditional neo-Gricean theory of implicature and a grammatical one is irrelevant for our purposes, since both analyses agree that upper-bounding inferences with numerals are derived by the implicature system, whatever that is, and our experiments focus on a population for whom this system, whatever it is, does not function as it does for typical speakers and hearers. That said, we consider a specific implementation of the grammatical view of implicature in detail in section 4.

Second, baseball enthusiasts will recognize that the math is a bit more complex than (3) lets on, since batting average also depends on the number of official at-bats. So (3) should really be heard as prefaced by an implicit *"Assuming he has n official at-bats..."* for some appropriate n.

or more hits to finish with an average of precisely .345, it is taken to mean that he must get exactly three hits.

Examples like (3), in which a logical operator appears to compose with a proposition that involves an upper bounded interpretation of a numeral, are one instance of a large (and growing) set of challenges to the neo-Gricean analysis of upper-bounding inferences of numerals that have appeared in the theoretical and experimental literature over the past thirty years (see e.g. Sadock 1984; Koenig 1991; Horn 1992; Scharten 1997; Carston 1998; Krifka 1998; Noveck 2001; Papafragou and Musolino 2003; Bultinck 2005; Geurts 2006; Breheny 2008; Huang, Spelke, and Snedeker 2013; Marty, Chemla, and Spector 2013; Kennedy 2013). Taken as a whole, this literature largely agrees that upper bounded meanings are semantic; there is, however, no consensus about how exactly upper bounded meanings are derived and how they are related to lower bounded meanings. This is due partly to the fact that some of the literature does not take a position on the semantic content of numerals, and partly to the fact that there are multiple ways of characterizing the meaning of numerals (e.g. as determiners vs. cardinality predicates vs. singular terms that compose with either a parameterized determiner or a parameterized cardinality predicate). Some of these analyses are truth-conditionally distinct and some are not (see Kennedy 2013 for discussion), and the data under consideration in the literature on numerals and implicature often do not decide between them (though see Geurts 2006 for one attempt to do so).

At a general level, however, we can draw a distinction between two kinds of semantic approaches to upper-bounding, and the relation beween upper- and lower bounded meanings. The first class of approaches, which we will refer to as LOCAL ANALYSES, includes several types of analyses that are distinct from each other in many respects, but share the assumption that bounding inferences are introduced through composition of the numeral and the constituent that introduces the objects that it counts (typically a noun). This class includes analyses in which numerals introduce upper bounded content exclusively (Koenig 1991; Breheny 2008; Ionin and Matushansky 2006); analyses in which numerals or some part of the larger nominal constructions in which they appear are ambiguous between upper bounded and lower bounded meanings (Geurts 2006; Nouwen 2010); and analyses in which numerals are underspecified for bounding entailments but are then subject to post-compositional, truth-conditional enrichment (Carston 1998). For the purposes of this paper, we will use an example of the ambiguity analysis as the representative of this class of approaches, but our broad conclusions extend to the other variants as well.

For example, in the version of the ambiguity analysis articulated in Nouwen 2010, numerals themselves unambiguously denote numbers, which are model-theoretically an instance of the semantic type of degrees. The counting relation between the number denoted by the numeral and the object(s) denoted by a noun is introduced by an unpronounced, parameterized cardinality determiner MANY (Hackl 2001), which is ambiguous between the "weak" version in (4a) and the "strong" version in (4b), where '#' is a function that returns the number of atoms that a (possibly plural) individual consists of and $\exists!x$ means "there is a unique x."

(4) a. $[\![\text{MANY}_w]\!] = \lambda n \lambda P \lambda Q.\exists x[P(x) \wedge \#(x) = n \wedge Q(x)]$

b. $[\![\text{MANY}_s]\!] = \lambda n \lambda P \lambda Q.\exists! x[P(x) \wedge \#(x) = n \wedge Q(x)]$

Composition of e.g. *three* with MANY_w and MANY_s gives the determiner denotations in (5a) and (5b), respectively.

(5) a. $[\![\text{three MANY}_w]\!] = \lambda P \lambda Q.\exists x[P(x) \wedge \#(x) = 3 \wedge Q(x)]$

b. $[\![\text{three MANY}_s]\!] = \lambda P \lambda Q.\exists! x[P(x) \wedge \#(x) = 3 \wedge Q(x)]$

(5a) introduces lower bounded truth conditions because it involves existential quantification over the individual argument provided by its restriction and scope terms:

if there is a group of size 3+n that satisfies P and Q, then there is a group of size 3 that satisfies P and Q (when P and Q are distributive; see Koenig 1991). (5b) introduces upper bounded truth conditions because of the addition of a uniqueness requirement: if there is a unique group of size 3 that satisfies P and Q, then there is no group of size 3+n that satisfies P and Q. The difference in meaning between (2) and (3), on this view, reflects different choices for MANY: (2) involves MANY_w and has the truth conditions in (6a), and (3) involves MANY_s and has the truth conditions in (6b).[2]

(6) a. $[\![$Mookie has to get three MANY_w hits$]\!]$ =
$\Box[\exists x[\textbf{hits}(x) \wedge \#(x) = 3 \wedge \textbf{got}(x)(\textbf{mookie})]]$

b. $[\![$Mookie has to get three MANY_s hits$]\!]$ =
$\Box[\exists! x[\textbf{hits}(x) \wedge \#(x) = 3 \wedge \textbf{got}(x)(\textbf{mookie})]]$

The second class of approaches is one in which bounding entailments at the sentential level arise not directly through the composition of an ambiguous or underspecified numerical constituent with the noun (or other sortal term), but rather through scopal interactions between the numeral and other expressions in the sentence; we refer to these as SCOPAL ANALYSES. On this view, numerals are neither determiners nor cardinality predicates nor singular terms (denoting numbers), but

[2]Two comments. First, in analyses in which numerals unambiguously introduce upper bounded semantic content, such as Breheny 2008, only (6b) is derived in the semantics, and the lower bounded interpretation that is seen in examples like (2) is derived pragmatically. But this analysis is similar to the ambiguity analysis in failing to derive upper bounded readings semantically in the crucial examples involving existential modals that we will introduce in the next section.

Second, throughout this paper, we ignore parses in which the entire nominal (*three hits* in this example) takes scope above the modal. Such readings may be available, but are irrelevant to our discussion because they would involve *de re* interpretations of the noun, and all of the examples we are interested in are quite clearly interpreted *de dicto*.

are rather quantificational expressions in their own right, specifically generalized quantifiers over degrees (Kennedy 2013, 2015; cf. Frege 1980 [1884], Scharten 1997, and von Stechow's (1984, p. 56) treatment of measure phrases). This analysis is similar to Nouwen's in that numerals saturate a degree position in the nominal projection and the individual variable introduced by the nominal is existentially bound, but there is no need to assume a distinction between "strong" and "weak" MANY, the existential truth conditions of the weak version or some equivalent cardinality predicate are sufficient (cf. Cresswell 1976; Krifka 1989). The numeral *three*, for example, has the denotation in (7): it is true of a property of degrees if the maximal degree that satisfies it is the number three.

(7) $[\![\text{three}]\!] = \lambda P.max\{n \mid P(n)\} = 3$

On this analysis, (1) has the logical form in (8a) in which the numeral takes scope (assuming a movement-based analysis of scope-taking for simplicity); its clausal argument has the denotation shown in (8b); composition of the numeral with its complement derives the upper bounded truth conditions in (8c).[3]

(8) a. three [Mookie got t MANY hits]

b. $[\![\text{three}]\!](\lambda n.\exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{mookie})])$

c. $max\{n \mid \exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{mookie})]\} = 3$

Crucially, since numerals are quantificational, they may interact with other operators, and it is this interaction that accounts for the difference in meaning between (2) and (3). The lower bounded interpretation in (2) arises when the numeral takes

[3] A one-sided reading can be derived by lowering the quantificational denotation of the numeral to a singular term denotation as a number (Kennedy 2015), which returns a meaning identical to Nouwen's MANY_w parse. Experimental results reported in Marty et al. 2013 support an analysis in which the upper bounded interpretation of a simple sentence like (1) is basic and the one-sided interpretation is derived.

scope over the universal modal *have to*, as in (9a), and the upper bounded interpretation in (3) arises when the numeral takes scope below the modal, as in (9b).[4] (Here we reconstruct the subject below the modal for simplicity.)

(9) a. ⟦three [has to [Mookie get t MANY hits]]⟧ =
$max\{n \mid \Box[\exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{mookie})]]\} = 3$

b. ⟦has to [three [Mookie get t MANY hits]]⟧ =
$\Box[max\{n \mid \exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{mookie})]\} = 3]$

1.2 Upper bounds under existential modals

Our goal in this paper is to examine a set of data that allows us to draw a distinction between these two kinds of approaches, and, we claim, argues in favor of the scopal analysis. The crucial facts involve sentences in which a numeral is embedded in the complement of an existential modal, such as (10).

(10) Mookie can make three errors on the last day of the season and still have the best fielding percentage in the league.

The first conjunct in (10) is most naturally understood as imposing an upper bound on the number of errors that Mookie can make, but unlike (1) and (3), it does not impose a lower bound: it allows for the possibility of Mookie making two, one or zero errors.

In the scopal analysis of numerals, the upper bound reading of (10) is derived compositionally by scoping the numeral over the modal, which returns the truth

[4] (9a) says that three is the maximum n such that in every world in the modal domain (worlds in which Mookie wins the batting title), there's a group of hits of size n that Mookie gets. There are groups of size three in worlds in which he gets more than three hits, but there are no groups of size three in worlds which he gets fewer than three hits. (9a) thus places a lower bound of three on the number of hits that we find in each world that satisfies the modal claim.

conditions shown (11).

(11) $[\![$three [can [Mookie make t MANY errors]]$]\!]$ =

$max\{n \mid \Diamond[\exists x[\textbf{errors}(x) \wedge \#(x) = n \wedge \textbf{make}(x)(\textbf{mookie})]]\} = 3$

(11) says that three is the maximum n such that there is a world in the relevant modal domain in which Mookie makes at least n errors, which rules out the possibility that he makes more than three errors, and so semantically imposes an upper bound. The numeral may also take scope below the modal, deriving the truth conditions in (12), which require merely that there is a world in the modal domain in which Mookie makes exactly three errors.

(12) $[\![$can [three [Mookie make t MANY erros]]$]\!]$ =

$\Diamond[max\{n \mid \exists x[\textbf{errors}(x) \wedge \#(x) = n \wedge \textbf{make}(x)(\textbf{mookie})]\} = 3]$

This meaning is quite weak, because it does not rule anything out, but it does seem to be available. (*"Mookie can make three errors; in fact he can make as many as he wants!"*)

On the local analysis, there are two possible parses of (10), shown in (13a-b).

(13) a. $[\![$Mookie can make three MANY$_w$ errors$]\!]$ =

$\Diamond[\exists x[\textbf{errors}(x) \wedge \#(x) = 3 \wedge \textbf{make}(x)(\textbf{mookie})]]$

b. $[\![$Mookie can make three MANY$_s$ errors$]\!]$ =

$\Diamond[\exists! x[\textbf{errors}(x) \wedge \#(x) = 3 \wedge \textbf{make}(x)(\textbf{mookie})]]$

The crucial difference between these meanings and those in (11a-b) is that neither (13a) nor (13b) entails an upper bound. (13b) is logically equivalent to (12), and has the same weak truth conditions. (13a) also has weak truth conditions, and only entails that making fewer than three errors is allowed; it does not entail that making a greater number of errors is not allowed. But (13a) can be strengthened to an upper bounded interpretation via reasoning involving the Maxim of Quantity: (13a) (but not (13b)) is asymmetrically entailed by alternative propositions of the same form but with higher values for the numeral. Using the same Quantity reasoning that the classic neo-Gricean analysis appeals to in the case of simple sentences like (1), we can generate the implicature that the speaker believes that Mookie cannot make four, five, etc. errors, which derives the upper bound.

Note that neither the scopal nor the local analysis derives the so-called "free choice" inferences of (10) — that Mookie cam make exactly two errors, that he can make exactly one, and that he can make zero — as an entailment, given a standard Kratzerian semantics for the modal. That is, none of the parses in (11), (12) and (13a-b) entail that there are modally accessible worlds in which Mookie makes only two errors, worlds in which he makes only one, and worlds in which he makes none, though all are consistent with the existence of such worlds. We return to this point below.

Summarizing, the key difference between the two approaches is that the scopal analysis derives an upper bounded interpretation for (10) semantically, by scoping the numeral over the modal, while the local analysis can only derive such a meaning pragmatically, based on a semantics involving the lower bounded meaning of the numeral. The two analyses therefore make different predictions about how sentences like (10) will be evaluated when Quantity implicature calculation is suppressed or otherwise inactive: the scopal analysis predicts that upper-bounding inferences will be retained, all other things being equal; the local analysis predicts that they will disappear. We can therefore distinguish between the two approaches by examining how sentences like (10) are understood by a population that has competence with quantification but has difficulty with Quantity reasoning. In the next section, we describe one such population.

2 Numerals and quantity implicatures in child language

A broad range of acquisition studies support the conclusion that young children systematically have difficulty computing upper-bounding implicatures in contexts in which adults virtually automatically generate such meanings (Barner, Brooks, and Bale 2010; Chierchia, Crain, Guasti, Gualmini, and Meroni 2001; Huang et al. 2013; Hurewitz, Papafragou, Gleitman, and Gelman 2006; Gualmini, Crain, Meroni, Chierchia, and Guasti 2001; Guasti, Chierchia, Crain, Foppolo, Gualmini, and Meroni 2005; Noveck 2001; Papafragou and Musolino 2003; Papafragou 2006; Smith 1980). Although children's ability to compute scalar implicatures can be improved when certain conditions are met, e.g., by asking them to assess conversational interactions rather than descriptions of events (Papafragou and Tantalou 2004), by using ad-hoc and non-lexical scales (Barner et al. 2010; Papafragou and Tantalou 2004; Stiller, Goodman, and Frank 2015), by introducing relevant stronger alternatives (Skordos and Papafragou 2016), or by training them on the use of conventional terms (Papafragou and Musolino 2003; Guasti et al. 2005), the general conclusion that they differ from adults in their capacity to automatically calculate upper-bounding implications for scalar terms is robust.

One notable exception to this generalization is the case of numerals. For example, Papafragou and Musolino (2003) found that Greek-speaking five-year-olds who were assigned to a condition in which they were asked to evaluate sentences with *dio* 'two' in contexts in which a lower-bound reading is true but a upper bounded reading is false rejected the sentences on average 65% of the time. (Specifically six of the 10 children in the condition rejected the sentences on three or four of the four trials.) In contrast, children accepted sentences with *arxizo* 'start' and *meriki* 'some' over 80% of the time in contexts in which a sentence with a stronger scalar alternative (the Greek equivalents of *finish* and *all*) would have held true, while adults routinely rejected such sentences in these contexts. This pattern is reminiscent of the findings from a statement evaluation task conducted by Noveck (2001) in French, a pattern that was replicated by Guasti et al. (2005) in Italian. Further studies have replicated this difference between numerals and other scalar terms in child language — with the former having upper bounded interpretations and the latter lacking them — using different kinds of methodologies (see e.g. Huang et al. 2013; Hurewitz et al. 2006), and this distinction is now viewed by many as a central argument against a pragmatic account of upper bounded interpretations of numerals and for a semantic one.[5]

It is not the case that children always assign upper bounded interpretations to sentences containing numerals, however. Musolino (2004) showed that four- and five-year-olds correctly assign lower- and upper bounded interpretations to sentences that adults systematically treat in the same way, namely sentences with universal and existential modals similar to (2)-(3) and (10). In Musolino's first experiment, children were told stories in which a character had to perform an action with multiple objects in order to be awarded a prize. The rules were formulated using the numeral *two*, and differed based on the condition. In the 'at least' condition, participants were given a sentence such as (14a); in the 'at most' condition, participants given a sentence such as (14b).

(14) a. *At least condition*

[5] Barner and Bachrach (2010) argue for a different interpretation of this pattern, in which upper-bounding implications with numerals are derived via the implicature system. The crucial difference between numerals and other scalar terms in child language, they argue, is that children have difficulty constructing scalar alternatives with the latter (and so fail to exclude the stronger ones) but not with the former. It turns out that Barner and Bachrach's proposals together with an exhaustification-based, grammatical theory of scalar implicature makes similar predictions to the scopal analysis of numerals for the crucial examples. We compare the two approaches in detail in Section 4, arguing that there are independent reasons to prefer the scopal analysis.

Goofy said that the Troll had to put two hoops on the pole in order to win the coin.

b. *At most condition*

Goofy said that the Troll could miss two hoops and still win the coin.

In each condition, the Troll made five attempts to put hoops on the pole, ending up with four hoops on the pole and one miss. A puppet watched the stories alongside the children, and asked whether the troll wins the coin. In the ‘at least’ condition, children said the prize should be awarded 35% of the time, and in the ‘at most’ condition, they said the prize should be awarded more than 80% of the time. Suspecting that the low acceptance rate in the ‘at least’ condition was due to independent factors involving the children’s expectations about games (why did the troll keep going after he had put enough hoops on the pole to win?), Musolino ran a second ‘at least’ condition that controlled for these factors and brought the acceptance rate up to 80%, though in these cases the universal modal was not part of the test sentence, but rather part of an explicit question under discussion (what quantity of objects does a character need?).

Musolino’s study provides the starting point for the experiments we conducted to compare the local and scopal analyses of numerals. Recall that the crucial difference between these two analyses has to do with the mechanism for generating an upper bounded interpretation of sentences involving existential modals, such as (14b): in the scopal analysis, this meaning can be derived semantically by assigning the numeral scope over the modal; in the local analysis, it can only be derived pragmatically as an upper bounding implicature. Given that children generally do not calculate upper bounding implicatures, the local analysis predicts that they should fail to assign upper bounded interpretations to sentences like (14b). The scopal analysis, on the other hand, predicts that they should have no trouble computing such a meaning, provided they have acquired whatever general principles are necessary for scope taking in the first place; that children have acquired such principles has been shown by e.g. Lidz and Musolino (2002); Musolino and Lidz (2006); Syrett and Lidz (2009). Unfortunately, Musolino's experiments do not provide a basis for adjudicating between the two analyses, because his 'at most' condition looked only at scenarios in which the actual value of objects described was below the putative upper bound; he did not examine children's judgments about sentences like (14b) in scenarios in which the upper bound was exceeded. In order to decide between the two analyses of numeral meaning, we conducted a set of experiments that were similar to Musolino's but looked at children's interpretations of sentences like (14a-b) in three different scenarios: one in which the quantity of relevant items was less than the quantity named by the numeral ($<n$), one in which it was equivalent ($=n$), and one in which it exceeded the quantity named by the numeral ($>n$). In the next section, we describe these experiments and show that the results confirm the predictions of the scopal analysis.

3 Experiments

3.1 Experiment 1

The purpose of Experiment 1 was to determine whether children can assign an upper bounded interpretation to sentences in which a numeral is embedded under an existential root modal, at an age at which they are known to have trouble calculating scalar implicatures, by asking whether they accept or reject such sentences as descriptions of scenarios in which the upper bound is surpassed. Rejection of the crucial sentences as descriptions of such scenarios would provide evidence in favor of the scopal analysis of upper bounded interpretations and against the local analysis, since only the former can derive upper bounded interpretations for such sentences semantically. Acceptance of the crucial sentences in these scenarios would not decide between the scopal and local analyses, since both allow for the possibility of interpretations that do not impose upper bounds. In the scopal analysis, such an interpretation is the result of assigning the numeral scope under the modal; in the local analysis, such an interpretation should be the only option, since upper bounds can only be derived pragmatically.

Sentences with existential modals were compared with sentences in which the numeral is embedded under a universal root modal, which should allow for both upper and lower bounded interpretations, on either analysis. In contrast, an analysis in which numerals are always assigned upper bounded interpretations pragmatically (such as the traditional neo-Gricean approach described in the quote from Horn 1972) predicts that such sentences should have only lower bounded interpretations.

PARTICIPANTS 32 children (19 boys, 13 girls; range: 4;0-5;8; Mean 4;9, Median: 4;10) and 32 adults participated in Experiment 1, 16 participants per condition. Children were recruited at area preschools. Adults were undergraduates who earned course credit in a linguistics course in exchange for their participation. All participants were native speakers of English. Data from two additional adults were excluded due to native speaker status.

MATERIALS AND PROCEDURE The experimental task was a variant of the Truth Value Judgment Task (Crain and Thornton 1998). An experimenter told the participant a series of stories using animated images presented on a computer screen. Each story had the same structure. One character provided instructions to another character. The second character attempted to comply by performing an action. At the end of each story, a puppet, played by a second experimenter (or the experimenter, in the case of adult participants), briefly recalled the story plot and asked whether what the second character did was okay, reminding the participant of the first character's instructions. The participant's task was to respond "yes" or "no" (verbally in the case of children, or on a response sheet in the case of adults) and occasionally provide a justification.

In a sample control item, Ruby is drawing pictures of animals, and her brother Max enters, wanting to help draw something. Ruby gives Max a piece of paper and tells him to draw a cat. Max draws a dog, giggles, and says, *"Here you go!"* The puppet then says, *"That was a story about Max and Ruby, and Max wanted to help Ruby draw. I remember that Ruby asked Max to draw a cat. Is what Max did ok?"* Such items were responded to without difficulty, with participants saying "no" across the board.

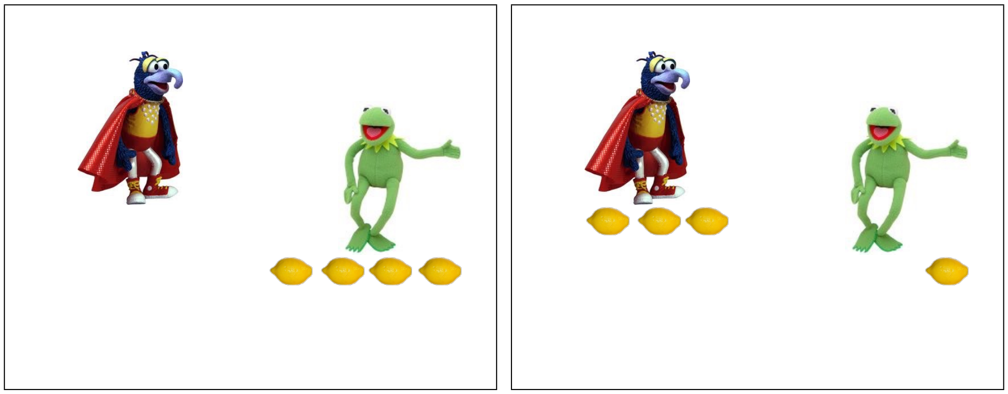
The experimental session began with two training items. The test session that followed consisted of six test items and four control items, all with the same basic structure, as described above. Each test item involved instructions from the second character that featured a modal and a numerical expression. There were two conditions, based on the type of modal: an UPPER BOUND condition that used the existential modal *allowed to*, and a LOWER BOUND condition that used sentences with the universal modal *have to*. Participants were evenly divided between the two conditions.

Within each condition, the six test items were further divided into two groups. In one group of three test items, the numerical expression composed directly with a plural count noun (e.g., *two books/carrots/lemons*). In a second group of three test items, the numerical expression was part of a measure phrase in a pseudopartitive construction and the noun was a mass noun (e.g., *two feet of water*). We examined both types of constructions to determine if the effects witnessed with numerals and plural count nouns extended to expressions of quantity in which the measurement did not involve cardinality. (Previous research demonstrates that by age four, children can properly interpret such expressions of measurement; Syrett 2013; Syrett and Schwarzschild 2009.) The numeral was always *two*.

Within each group of three test items, there was a scenario in which the second character performed an action that involved the exact amount required by the first character (=2), another in which the action involved an excess of the amount (always one more, in the case of the numeral/count noun examples) (>2), and another in which the action involved less than the amount (always one less, in the case of the measure phrase/mass noun examples) (<2). In this way, we manipulated quantity, and could determine — based on participants' responses — if they accessed an upper bounded, lower bounded or upper bounded interpretation of the sentence. Each participant saw two items of each quantity type, one version involving numeral/count noun combinations and the other involving measure phrase/mass noun combinations.

In one of the count noun scenarios, Gonzo is making lemonade and needs lemons. He turns to his friend Kermit, who has lots of lemons. Gonzo asks for some lemons, and Kermit is happy to oblige. In the UPPER BOUND condition, Kermit says that they have to share so that he has enough lemons for himself. He tells Gonzo, *"You are allowed to use two lemons."* In the <2 version, Gonzo takes one lemon; in the =2 version, he takes two lemons; and in the >2 version, he takes three. (One quantity version of each story was used across participants, so that the differences in quantity were paired with particular stories.) In the LOWER BOUND condition, Kermit says that he wants to make sure Gonzo has enough lemons to make nice lemony lemonade. He tells Gonzo, *"You have to use two lemons."* As above, Gonzo could take either one, two, or three lemons. Examples of beginning and ending scenes for this item are presenting in Figure 1.

The measure phrase plus mass noun scenarios were similar. In one scenario, Sister Bear and Brother Bear are playing outside on a hot day, and Sister Bear asks

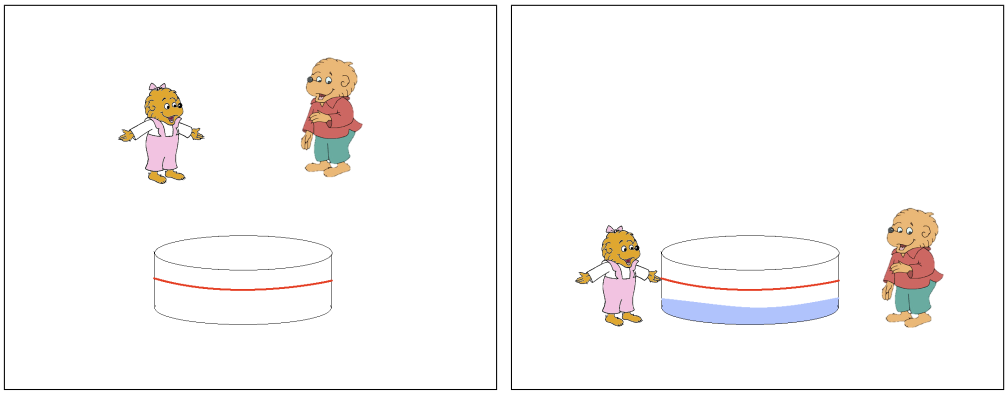


(a) Beginning scene (b) Ending scene

Figure 1: Experiment 1, count noun scenario for *two lemons* (>2)

Brother Bear if she can use the pool. Brother Bear agrees, and says she can fill it up all by herself. She says she has never done this before, and asks how much water to put in. In the upper bound condition, Brother Bear explains, *"You don't want the water to spill out when you splash! From the bottom of the pool to the red line is two feet. So ... you are allowed to fill the pool with two feet of water."* In the lower bound condition, he says *"You want to have plenty of water for splashing around! From the bottom of the pool to the red line is two feet. So ... you have to to fill the pool with two feet of water."* As with the count noun scenarios, Sister Bear could either fill the pool below the red line (<2), just to the red line (=2), or past the red line (>2). Examples of beginning and ending scenes for this item are presenting in Figure 2.

Experimental items using numerals plus count nouns included taking lemons to use for lemonade, reading books before bedtime, and sharing carrots. Items using measure phrases plus mass nouns included filling a pool with water, filling a pitcher with water, and filling a teddy bear with stuffing. As the second character performed the actions with the objects, the experimenter and computerized animation



(a) Beginning scene (b) Ending scene

Figure 2: Experiment 1, mass noun scenario for *two feet of water* (<2)

highlighted each sequential quantity, making it clear that the character had failed to meet, met, or exceeded the target amount. For example, if Gonzo ultimately took three lemons, he took one lemon, paused, took another, paused, and then took another. After each experimental item, the puppet repeated the first character's instructions, including the modal and numerical expression, and asked if what the second character did was okay. Participants then responded "yes" or "no."

RESULTS: UPPER BOUND CONDITION We will discuss the results in the upper and lower bound conditions sequentially. Recall that the dependent measure is the percentage of times the participants said that the second character's actions were "okay" when the quantity was less than, equal to, or greater than two (i.e., the percentage of "yes" responses to the question *"Is what so-and-so did okay?"*). Since all participants, regardless of age group or condition, readily accepted the items in which the target number was met (the =2 items), we focus our attention on a comparison of the <2 and >2 items. In addition, since there was no difference between subjects' behavior on the mass and count noun items in any of our experiments or conditions, we report the results for all test items as a single group.

	QUANTITY		
	<2	=2	>2
ADULTS	78.1%	100.0%	3.1%
CHILDREN	43.8%	100.0%	21.9%

Table 1: Mean percentage acceptance of second character's actions in the upper bound condition in Experiment 1

The results for the upper bound condition are presented in Table 1. A McNemar's test for the upper bound condition looking at overall results found a highly significant difference between the <2 and >2 items for adults (p<.0001, two-tailed) and a marginally significant difference for children (p=.09, two-tailed). Follow-up Wilcoxon tests for this condition for both age groups revealed a highly significant difference for adults, with adults more likely to say "yes" for the <2 items than for the >2 items (W=120, z=3.39, p<.001, two-tailed), but no statistical difference in acceptances for children across <2 and >2 items.

However, when we looked more closely at children's responses, we observed an item effect. With the exception of one item, children almost uniformly rejected the character's responses to the >2 items. The exceptional item involved a story about filling a bear with stuffing. In this story, one character is showing another character how to make a toy bear. The first says "you want your bear to be cuddly, but not too stiff, so you are allowed to use two inches of stuffing," which was measured by putting it into a container marked by a red line. Of the seven acceptances from children in the upper bound condition, six came from this one item, and the lone acceptance from adults in this condition also came from this item. We therefore concluded that this item was producing deviant responses and should be removed from the evaluation of all conditions. In the upper bound condition, this left us with sixteen responses to count noun items and eight responses to the other mass noun item. The revised results are shown in Table 2. A Wilcoxon test comparing children's responses to the adjusted <2 and >2 items revealed a highly significant difference in the same direction as the adults, with children more likely to say "yes" for the <2 items than for the >2 items (W=55, z=2.78, p=.005, two-tailed). In addition, children did not differ from adults in their rates of rejection of the >2 items (U_A=276, z=.24, p=.81, two-tailed).

	QUANTITY		
	<2	=2	>2
ADULTS	87.5%	100.0%	0.0%
CHILDREN	41.7%	100.0%	4.2%

Table 2: Mean percentage acceptance of second character's actions in the upper bound condition in Experiment 1, "Stuffing the Bear" items removed

To provide additional confirmation that the initial numbers reflected the influence of a faulty item, we also ran a follow-up experiment replicating all items in the original upper bound condition except for the "stuffing the bear" item, replacing it with a new >2 mass noun item. The follow-up featured the exact same images, except that the bear scenario was changed so that one character was showing another how to fill a container with stuffing in order to ship building materials. The first says to the second, "You want the building materials to be really protected, but you still want enough room for the building materials to go in and not be cushioned too much, so *you're allowed to use 2 inches of stuffing*." The second character ends up filling the stuffing past the two inch line on the container, as shown in Figure 3.

15 children (8 boys, 7 girls; range 3;11-6;0; Mean 4;11; Median 4;11) participated. This time, there was a clear and significant difference between the >2

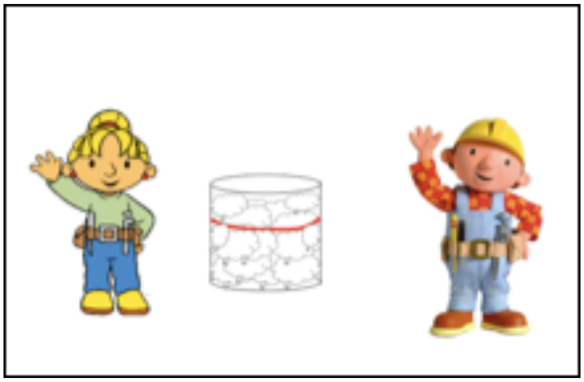


Figure 3: Follow-up scenario for *two inches of stuffing*

and <2 items, with children being more likely to say "yes" to the <2 items than to the >2 items, as shown in Table 3 (W=104, z=2.94, p=.004, two-tailed). These results support our judgment that the mass noun results in Experiment 1 were due to an item effect, and allow us to conclude that children systematically assign an upper bound interpretation to sentences in which a numeral appears in the scope of *allowed to*, regardless of whether it is a bare numeral accompanying a count noun or a numeral in a measure phrase accompanying a mass noun in a pseudopartitive.

	QUANTITY		
	<2	=2	>2
CHILDREN	60.0%	100.0%	10.0%

Table 3: Results of a follow-up experiment correcting for the item effect in Experiment 1

Before turning to the lower bound condition, we wish to note a second significant feature of the children's responses, and a difference between children and adults. Although children are significantly more likely to accept the <2 items than the >2 items, children are significantly less likely to accept the <2 items than the

=2 items (Experiment 1 with item removed: U_A=176, z=3.44, p=.0006, two-tailed; Experiment 1 followup: W=-66, z=-2.91, p=.004, two-tailed), while adults are not (Experiment with item removed: U_A=288, z=1.58, p=.114, two-tailed), and children are less likely than adults to accept the <2 items (U_A=372, z=-1.72, p=.09, two-tailed, marginally significant).[6]

RESULTS: LOWER BOUND CONDITION Table 4 presents the results for the lower bound condition of Experiment 1, showing percentages of acceptance for the original set of experimental items, minus the problematic "Stuffing the Bear" items. (We did not include a lower bound condition in our follow-up experiment, where the primary goal was to confirm the item effect.) A McNemar's test taking into account the responses to the <2 and >2 items for each of the age groups (collapsing over count and mass items) revealed no significance for either group (children: p=.69, adults: p=.39, two-tailed). Follow-up Wilcoxon tests likewise revealed no difference between items for adults or for children for either the full or revised set of items; in particular, both groups strongly favored the upper bounded interpretation of numerals in this condition.

DISCUSSION Once we correct for the item effect in the upper bound condition, the results of Experiment 1 and the follow-up clearly demonstrate that, at an age in which children tend not to calculate upper bounding implicatures in tasks similar to ours (i.e., truth-value judgment tasks), they (together with adults) systematically

[6]Austin, Sanchez, Syrett, Lingwall, and Pérez-Cortes 2015 report similar results in a study of how monolingual English and Spanish-English bilingual children interpret sentences of the form *'You may take two/all of the N'* in scenarios in which a conversational participant took less than the amount indicated by the numeral. While the child participants from different language backgrounds diverged on other scalar target items, they patterned the same with these control sentences, diverging from adult controls.

	QUANTITY		
	<2	=2	>2
ADULTS	12.5%	96.9%	33.3%
CHILDREN	16.7%	100.0%	20.8%

Table 4: Mean percentage acceptance of second character's actions in the lower bound condition in Experiment 1, excluding the "Stuffing the Bear" items

assign upper bounded interpretations to sentences in which numerals appear in the scope of an existential root modal. This result is expected on the scopal analysis of numerals, in which such readings are derived semantically by assigning the sentence a parse in which the numeral takes scope over the modal, but it is unexpected on the local analysis of numerals, in which such readings can only be derived pragmatically. These results therefore provide prima facie evidence in support of the scopal analysis of numerals and against the local analysis. Our results in the lower bound condition, which showed that both children and adults favor upper bounded interpretations, are consistent with the scopal analysis, and provide a further argument against a purely pragmatic account of upper bound implications, which cannot derive such interpretations at all. On the other hand, the fact that lower bound interpretations were so rare is unexpected, on any account of numerals. We investigate the possibility that this result was an artifact of the experimental setup in Experiment 2, presented in the next section.

There are three ways that the proponent of the local analysis could challenge our conclusion that the results in the upper bound condition of Experiment 1 argue in favor of the scopal analysis of numerals. The first challenge says that children's failure to calculate upper bounding implicatures does not extend to the case of numerals: the robust literature demonstrating children's capacity to interpret numerals as having upper bounds when other scalar terms do not does not indicate that numerals are semantically upper bounded, but rather that children have no trouble calculating scalar implicatures for numerals. If this is correct, then children's behavior in the upper bound condition simply reflects the implicature system at work, and so does not argue in favor of a semantic analysis of upper bounds (scopal, local or other). Working out the details of this challenge in a way that is consistent with the general pattern of numeral interpretation under modals requires some space, so we address it separately in Section 4.

The second challenge, suggested to us by one anonymous reviewer, accepts that children's failure to calculate upper bounding implicatures extends to numerals, but denies that upper bounding implicatures need to be invoked to explain children's behavior in the upper bound condition of Experiment 1. Instead, their behavior can be explained in terms of a more general kind of reasoning about what is appropriate or inappropriate behavior in the context, given what they have been told. Specifically, when a child hears an utterance like *"Gonzo is allowed to use two lemons,"* she assigns it the interpretation predicted by the local analysis — *this is what is allowed: Gonzo uses exactly two lemons* — and since there is uncertainty about how to distribute the lemons, she follows a "safe strategy" and rejects scenarios in which he takes three lemons, since it is not known whether or not this is allowed.

The problem with this explanation is that it leads us to expect similar judgments about *all* scenarios in which the character in the story does something that is not known to be allowed; i.e. we expect similar judgments on the <2 and the >2 items. But this is not what we saw: children accepted <2 items significantly more than >2 items (W=55, z=2.78, p=.005, two-tailed). At the same time, children accepted <2 items significantly less than =2 items (U_A=176, z=3.44, p=.0006, two-tailed), and less than adults accepted <2 items (U_A=372, z=-1.72, p=.09, two-tailed, m.s.), though their rejection of >2 items was not different from adults (U_A=276, z=.24, p=.81, two-tailed). The full picture that emerges from these results, then, is that children are adult-like in their interpretation of sentences involving numerals and existential modals, with one exception: they are less likely than adults to accept scenarios in which the character does less than what is explicitly allowed. Or, to put it another way, *they are less likely than adults to calculate free choice inferences for* *numerals.*[7]

This, we believe, provides an argument against the kind of explanation suggested by the reviewer, and support for the scopal analysis. The free choice inferences of a sentence like *"Gonzo is allowed to use two lemons"* are that Gonzo is allowed to use exactly two lemons, that he is allowed to use exactly one lemon, and

[7] At first glance, this appears to be a surprising result, since there is evidence that children are able to calculate free choice inferences for disjunctions like (i) even when they cannot calculate upper bounding implicatures (Zhou, Romoli, and Crain 2013; Tieu, Romoli, Zhou, and Crain 2015). That is, children hear (i) as communicating (iia) and (iib) but not (iic), in contrast to adults.

(i) Gonzo can have cake or ice cream.

(ii) a. Gonzo can have cake.

b. Gonzo can have ice cream.

c. Gonzo can't have cake and ice cream

But in fact, it is not clear that free choice inferences for numerals and free choice inferences for disjunction are derived in the same way. For example, it has been argued that free choice inferences with disjunction are derived by essentially the same principles that derive scalar implicatures (see e.g. Fox 2007; Franke 2011; Bar-Lev and Fox 2017), in such a way that excluding the possibility that an individual disjunct holds exclusively entails the possibility of the other disjuncts. For example, given (i), if it's not the case that Gonzo can only have cake, then it must be the case that he can have ice cream, and vice-versa. And unlike the computation of the scalar implicatures for disjunction, this process does not require access to lexical alternatives, which the references cited above identify as the key factor in explaining the child language data. Numerals are different. First, any calculation will necessarily involve access to lexical alternatives, a point to which we return in Section 4. Second, even allowing for consideration of lexical alternatives, excluding the possibility that a particular alternative holds exclusively says nothing about the possibility of any of other relevant alternatives: it does not follow from the fact that Gonzo cannot use exactly n lemons that he can use exactly m lemons for any numerical alternatives n, m. Evidently something else (or something in addition) must be involved in the derivation of free choice inferences for numerals, though figuring out what this is must be left for another paper.

that he is allowed to use no lemons. As we saw in Section 1.2, neither the scopal nor the local analysis derives these as an entailment, which means that they are evidently derived as implicatures. And indeed they are cancellable: there's a salient reading of a sentence like *"Students are allowed to take two classes per quarter."* that does not grant permission to take one class or no classes. How precisely such implicatures are derived is not a question that we can answer here (see note 7 for some considerations); what we can say is that the difference between children's behavior on the <2 items and the >2 items, and the difference between children and adults on the <2 items, is exactly what we would expect if free choice inferences are derived pragmatically while upper bounding inferences are derived semantically. The scopal analysis introduces this asymmetry; the local analysis does not.

The third challenge to our conclusion, suggested to us by a second anonymous reviewer, also accepts that children's failure to calculate upper bounding implicatures extends to numerals in the general case, but supposes that the specific contextual setup of our Experiment 1 made such implicatures available. The reasoning goes like this:

(i) In the upper bound condition, it was made contextually salient that there are constraints on what is allowed, specifically that there were limits (not made explicit) on how much/many of some stuff/items a character was allowed to use for some task.

(ii) Children's behavior on upper bounding implicatures improves when it is contextually salient that there are constraints on what is allowed.

(iii) Given (i) and (ii), we can't exclude the possibility that children's rejection of >2 items in the upper bound condition is due to an upper bound implicature, rather than to the semantics of the crucial sentences.

This reasoning is valid, but whether it is sound depends on the whether premise (ii) is true, and we know of no study providing direct evidence for this conclusion. Papafragou and Tantalou (2004) show that children's performance on scalar implicatures improves when they assess actual conversational interactions compared to assessments of descriptions of events acted out in a story, but our study was of the latter type, not the former. Skordos and Papafragou (2016) show that children's performance on scalar implicatures with *some* improves in the latter type of task with prior exposure to examples involving alternative quantity terms, regardless of whether those terms are true scalar alternatives to *some* or not (the authors found facilitation with both *all* and *none*), but only when it was salient that judgments were about quantity as opposed to object type. But this study shows only that salience is a necessary condition for implicature calculation, not that it is a sufficient one, and our study did not include exposure to alternative quantity terms other than the target numerals. Finally, while there is robust evidence that congruence with the question under discussion can influence children's behavior on interpretive tasks such as access to different readings of scopally ambiguous sentences (Gualmini, Hulsey, Hacquard, and Fox 2008), we know of no study demonstrating that this factor alone (in the absence of e.g. also making salient scalar alternatives) can improve implicature calculation in child language.

That said, in order provide a quick assessment of the reviewer's suggestion that our results crucially depended on the contextual setup of the experiment, we added a single experimental item to the end of an independent study run in March 2019, which asked children for judgments on *"allowed ... n"* sentences in contexts that provided no explicit information about whether there were constraints on what is allowed. The participants saw a slide with three bears, and the experimenter said *"Look at these bears. They would like some balls."* She then introduced a second slide in which a character appeared with a stack of balls, and reported that the character said *"You're allowed to have n balls,"* where n was *two* or *three*. Participants then saw a third slide with the same three bears, this time in possession of two, three or four balls. For each bear, the experimenter asked, *"Is what they took ok?"* If a child needed the character's prompt repeated, the experimenter repeated it once.

The results for a total of 15 children tested (average age 4 years 7 months) are shown in Table 5, where percentages indicate the number of *"yes"* answers to the experimenter's question.

		QUANTITY		
		2	3	4
NUMERAL	*two*	100%	0%	0%
	three	13%	100.0%	0%

Table 5: Upper bound implications in the absence of contextually salient constraints on what is allowed

These results mirror the results we obtained in the full study: children hear *"allowed ... n"* sentences as ruling out values higher than n, indicating that they have computed an upper bound, even when the context does not make it salient that there are constraints on what is allowed. At the same time, they are relatively unlikely to accept lower values, indicating a failure to compute free choice inferences.

3.2 Experiment 2

Even given the considerations articulated in the discussion of Experiment 1, our conclusion that the results in the upper bound condition provide evidence for the scopal analysis of numerals is tempered somewhat by the results we obtained in the lower bound condition. On the one hand, these results provide further evidence that the upper bounded interpretation of numerals is derived semantically. As we pointed out in section 1.1, standard neo-Gricean analyses of upper bounded meanings fail to derive the correct truth conditions for sentences in which numerals are embedded under universal modals, yet both children and adults robustly assigned such meanings to these sentences in the lower bound condition, systematically rejecting the >2 items that made the upper bounded reading false and the lower bounded meaning true. On the other hand, if children's behavior on the upper bound condition indicates that the scopal analysis is correct, and in particular indicates that they can assign numerals wide scope relative to existential root modals, then these results are puzzling. All things being equal, children (and adults) should also be able to assign numerals wide scope relative to universal root modals as well, which would in turn derive lower bounded truth conditions and lead to a higher rate of acceptance of the >2 items.

An explanation for these results, consistent with the scopal analysis, is that there was a bias in favor of parses that deliver upper bounded truth conditions which our experiment did not control for — e.g., a preference for parses that deliver stronger meanings.[8] If this explanation is correct, we hypothesized, then manipulating the context to promote a lower bound interpretation should cause it to emerge more clearly, in a parallel way, for both children and adults. We tested this hypothesis in a second experiment, whose purpose was, primarily, to manipulate the experimental contexts in the test session in order to highlight the lower bounded

[8] A preference for parses that deliver stronger interpretations in child language is not inconsistent with limited scalar implicature calculation in child language. The former involves a comparison of alternative meanings or structures that are derived by the grammar from the same set of lexical items, while the latter involves access to lexical alternatives and/or reasoning about the communicative intentions of other agents. Recent studies of children's interpretation of disjunction indicate that they have no problem with the first kind of task even when they struggle with the second kind (Zhou et al. 2013; Tieu et al. 2015; Singh, Wexler, Astle-Rahim, Kamawar, and Fox 2016).

interpretation of the sentences in which the modal *have to* and the numeral *two* interacted, thereby inducing a higher percentage of acceptances by children and adults when the character in the story exceeds the specified amount.

PARTICIPANTS 14 children (8 boys, 6 girls; range: 4;1-5;0; Mean: 4;6, Median 4;6) and 32 undergraduates participated. Data from three additional children were excluded due to a "yes" bias, and data from two additional adults due to native speaker status. More adults than children were run, merely as a result of participant recruitment methods for the in-lab undergraduate population. Reducing the adult sample size by randomly selecting a subset, the size of which would be on par with the child sample, would not alter the results, since adults were so consistent in their responses. We therefore chose to maintain the entire adult sample for analysis.

MATERIALS AND PROCEDURE The experimental task was similar to the one described in Experiment 1 for the lower bound condition, but with some crucial changes made in order to induce more willingness on the part of the participants to accept the character's actions when s/he took or did more than the specified amount. We outline these changes here.

First, we noticed that the requirements of the character in the set of stories included in Experiment 1 were often mandated by an authority figure. For example, the Man in the Yellow Hat tells Curious George that he has to or is allowed to read two stories before bed, or older Brother Bear tells younger Sister bear to fill the pool a certain amount. We suspected that the participants might have thought that the second character would be more likely to hear such a figure as issuing strict commands, thereby favoring an upper bounded interpretation and decreasing the likelihood of participants accepting scenarios in which the second character behaved in accord with a lower bounded interpretation. To address this potential issue, we made the status of the characters more similar.

Second, in two of our scenarios, the first character was sharing some of his objects with the second, who would not be able to return them afterwards. For example, Kermit gave Gonzo lemons to use for lemonade, and Peter Rabbit gave Benjamin Bunny carrots to eat. We noticed that some children did not want the second character to take too many pieces of food away from the first character, thereby favoring a scenario in which the second character took as little as possible while still obeying the first character, resulting in an upper bounded reading. To address this potential issue, we made sure that none of the new stories involved this element.

Third, we also noticed that the stories also often involved effecting a change upon an external object, often without any apparent consequence for the second character. For example, Gonzo had to use a certain number of lemons for his lemonade, but doing so did not necessarily have consequences for him or reflect something about him. We suspected that if we made it clear that meeting — or exceeding — the given standard had positive consequences for the second character, participants might be more willing to accept >2 scenarios. We therefore changed the scenarios so that in performing the action, the second character was demonstrating a property that was desirable (e.g., being tall enough to go on a ride, eating enough vegetables to be healthy, swinging on enough vines to be fast, and so on). As a result of this change, instead of asking if what the second character did was okay, we asked if they would be able to do something (e.g., go on the ride, play, etc.).

We made two additional changes that did not involve contextual manipulations. First, we reduced *have to* to *hafta* based on an intuition that the non-reduced form might favor an upper bounded reading. And second, we used the numeral *three* instead of *two*, because our stories involved the age and height of characters who were more likely to be compared to being three feet high and three years old.

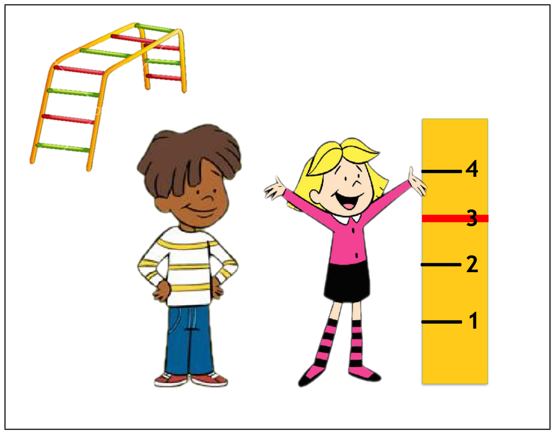


Figure 4: Experiment 3, >3 mass noun scenario

In an example scenario, Emily Elizabeth is playing with her friend Charley at the playground. Usually she plays on the swingset or on the slide, but today, she wants to play on the monkey bars. Charley plays on them all the time. He tells her *"They're pretty high. You hafta reach three feet to be able play on them."* He then asks her how high she can reach. They stand by a chart with a red line marking three feet. Emily Elizabeth reaches, and she says, reaching her arms above the three-foot line, *"Look, Charley, I can reach four feet!"* The puppet then briefly recaps what happened in the story and what Charley said, and asks the participant, *"Can Emily Elizabeth play on the monkey bars?"* The participant's job was to say "yes" or "no" and to provide a justification for his or her answer. A snapshot from this test scenario is presented in Figure 4.

RESULTS The results for experiment 2, compared with those of the lower bound condition in Experiment 1 (all items), are presented in Table 6. It is immediately apparent that the story manipulations had an effect: the percentage of acceptances increased sharply, with the difference between the results for Experiments 1 and 2 significant for both children and adults (Mann-Whitney tests, adults: U_A=4, z=5.5, p<.0001; children: U_A=198, z=-3.55, p<.001). Notice that the manipulations did not increase acceptance across the board: participants were unwilling to accept amounts that were less than the amount indicated by the numeral or measure phrase, but were now willing to accept both the exact amount and amounts that were greater. Both children and adults were more likely to accept the character's actions in the =2 scenarios than in the <2 scenarios (children: W=-300, z=-4.28, p<.0001; adults: W=-2080,z=-6.95 p<.0001) but there was no difference between the =2 and >2 scenarios (children: W=0, n.s.; adults: W=1, n.s.).

Although the difference in responses between Experiments 1 and 2 is not significant for the <2 items themselves (adults: p=.29, children: p=.15, two-tailed), the standard deviation in the responses was different between the two experiments (adults: .577 v. 0, children: .730 v. .267), demonstrating more consistency in rejection of the character's actions in Experiment 2. Moreover, a comparison of the participants' responses between Experiment 2 and the version of Experiment 1 in which the deviant 'stuffing' item was removed reveal a similar trend: while children and adults were significantly more likely to accept the character's actions in Experiment 2 than in Experiment 1 for the >2 items (children: U_A=106, z=4.21, p<.0001; adults: U_A=268, z=4.68, p<.0001), they did not differ from each other in their rate of acceptance for the <2 items (U_A=324, z=-0.73, p=.466).

DISCUSSION Experiment 2 demonstrates that both children and adults assign lower bounded interpretations to sentences in which numerals are embedded under universal root modals when the experimental scenario promotes such readings. Together with the results of Experiment 1, these findings confirm that children have full capacity with both lower and upper bounded interpretations of numerals in sentences involving universal modals, in line with the earlier results from Musolino

	QUANTITY: EXP. 1			QUANTITY: EXP. 2		
	<2	=2	>2	<2	=2	>2
ADULTS	12.5%	96.9%	33.3%	0.0%	100.0%	98.4%
CHILDREN	16.7%	100.0%	20.8%	3.6%	89.3%	78.6%

Table 6: Mean percentage acceptance of second character's actions in the lower bound condition of Experiment 1 and Experiment 2

(2004) that we discussed in section 2. Moreover, as noted above, the results of the lower bound condition in Experiment 1 provide additional evidence against the traditional neo-Gricean pragmatic analysis of upper bounded interpretations of numerals, since such an analysis cannot deliver the upper bounded interpretation of numerals in the scope of universal modals, which both children and adults favored in this study. This in turn means that such interpretations are derived semantically, and the results of the upper bound condition in Experiment 1 provide evidence that favors the scopal variant of a semantic account over the local one. Finally, the fact that children and adults show parallel behavior in all of our tasks can also be taken as evidence for the scopal analysis, assuming children and adults both have equal command of the grammatical principles generating these two readings — namely principles of scope assignment, as has been argued by e.g. Lidz and Musolino (2002); Musolino and Lidz (2006); Syrett and Lidz (2009).

4 Degree quantification vs. exhaustification plus scale salience

We have argued that Experiments 1 and 2 provide further support in general for the conclusion that upper bounded readings of sentences containing numerals are derived semantically, and provide support in particular for a scopal analysis of the semantics of numerals over a local analysis. The crucial examples are those in which a numeral is embedded under an existential root modal, such as *"You are allowed to read two books,"* which are naturally understood as imposing an upper bound. In the local analysis, this reading can only be derived pragmatically as a scalar implicature, even assuming a upper bounded semantics for the numeral; in the scopal analysis, it is derived semantically by scoping the numeral above the modal. Our Experiment 1 demonstrated that children quite naturally assign an upper bounded interpretation to such sentences at an age in which they have been shown to fail to compute scalar implicatures in similar tasks. If this result indicates that such readings are derived in a way that does not engage the mechanisms involved in calculating scalar implicatures, then, since the scopal analysis is the only existing semantic analysis of numerals that derives the right meaning without engaging these mechanisms, our results indicate that the scopal analysis is the correct one.

Clearly, a key element of this argument is the assumption that the large body of experimental work demonstrating children's reduced capacity to calculate scalar implicatures is general, and applies to all sentences containing scalar terms, including numerals. This assumption has been challenged by Barner and Bachrach (2010), however, who interpret children's differential behavior with numerals and other scalar terms not as indicating not that upper bounded interpretations of numerals are semantic, but rather that numerals are different from other scalar terms in supporting scalar implicature calculation. More precisely, Barner and Bachrach (B&B) suggest that children do not have a reduced capacity to calculate scalar implicatures at all; instead, their non-adult-like behavior with scalar terms like *some*, *start* and so forth indicates a failure to construct the scalar alternatives for these expressions that the implicature mechanism needs in order to generate upper bounding implications. Numerals, on the other hand, are different. Children explicitly learn numerals as members of an ordered list (*one, two, three, four, ...*), so the scales they occupy have increased salience compared to quantificational scales (*some*, *all*), aspectual scales (*start*, *finish*), and so forth. Because numeral scales are cognitively salient, children are able to construct scalar alternatives for sentences containing numerals, which then feed into the implicature mechanism and generate upper bounding implicatures.

B&B further argue that the developmental path of numeral acquisition is most consistent with a lower bounded semantics for numerals. It is well-established that children acquire adult-like competence with numerals gradually (see e.g. Wynn 1990, 1992). In the "*one*-knower" stage, children know that the word *one* applies to groups of cardinality one, and not to groups of greater sizes, but do not show similar competence with higher numerals. They then move to a "*two*-knower" stage in which they display adult-like competence with the words *one* and *two* but not higher numerals. This pattern continues until they jump to adult-like competence with all numerals and become "cardinal(ity) principle (CP) knowers" (Gelman and Gallistel 1978; Wynn 1990, 1992), typically around age three and a half to four. B&B observe that children who are n-knowers actually display greater than chance competence with $n + 1$, which they interpret to indicate that children actually have acquired a lower bounded meaning for $n + 1$ that is used to pragmatically compute an upper bounded meaning for n. If n-knowers instead had acquired an upper bounded meaning for n, they claim, there would be no reason for them to show greater than chance performance with $n + 1$, which they should at this point not have acquired at all.

B&B's account of numeral meaning, then, is essentially an updated version of the classic analysis from Horn 1972: numerals introduce lower bounded truth conditions, and upper bounded interpretations emerge as a scalar implicature that stronger alternatives are false. The crucial difference between numerals and other scalar terms in child language is that only the former actually have alternatives; alternatives for non-numeral scalar terms do not emerge until later in development.

B&B do not take a stand on whether the implicature mechanism involves reasoning about speaker meaning, as in the traditional neo-Gricean approach, or a silent exhaustivity operator, as in the "grammatical" approach to implicature (Chierchia 2006; Fox 2007; Chierchia et al. 2012; Spector 2013, inter alia), but the fact that children preferred upper bounded interpretations of numerals embedded under universal modals in Experiment 1 (see also Geurts 2006 and the discussion of this point in note 1) indicates that they would need to commit to the latter view.

For the purpose of working out the predictions of this analysis, let us assume that the semantic contribution of the exhaustification operator is something along the lines of (15), where $\subset$ is asymmetric entailment: exhaustification of ϕ gives back the conjunction of ϕ and the denial of all of its stronger alternatives.[9]

(15) $[\![exh\ \phi]\!] = \phi \wedge \forall\psi[[\psi \in \text{ALT}(\phi) \wedge \psi \subset \phi] \rightarrow \neg\psi]$

Next assume that numerals denote numbers and compose with a parameterized existential determiner MANY, as in the Nouwen-style analysis that we used to exemplify the local account of upper bounded meaning, with the only difference being that there is no ambiguity; there is only the "weak" MANY, which gives the lower bounded truth conditions for a sentence like (16a) in (16b).

(16) a. Gonzo used two lemons.

b. $\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2]$

Finally, assume with B&B that the (relevant) alternatives for sentences with numerals are variants that differ just in the numeral. Exhaustification of (16a) derives (17a), which is equivalent to (17b), and represents the upper bounded interpretation.

[9] The assumption that exclusion of alternatives is based on asymmetric entailment is a simplification for the purposes of illustrating the analysis. There is a rich and nuanced literature on the question of what this relation actually is; see e.g. Gazdar 1977; Hirschberg 1985; Sauerland 2004; Fox 2007; Bar-Lev and Fox 2017.

(17) a. $\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \forall \psi[[\psi \in \{\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = n] \mid n \in \mathbb{N}\} \wedge \psi \subset \exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2]] \rightarrow \neg \psi$

b. $\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \neg \exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) > 2]]$

With these assumptions in hand, the correct upper bounded interpretation of an example like *Gonzo has to use three lemons* is derived by inserting the exhaustivity operator in the embedded clause, as in (18a) (with reconstruction of the subject), which provides the content in (17) as the argument to the modal, as desired:

(18) a. has to [exh [Gonzo use three lemons]]

b. $\Box[\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \neg \exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) > 2]]]$

Alternatively, inserting exh above the modal, as in (19a) derives a lower bounded interpretation: it's necessary for Gonzo to use (at least) three lemons, and not necessary for him to use more.

(19) a. exh [has to [Gonzo use three lemons]]

b. $\Box \exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \neg \Box \exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) > 2]$

Exhaustivity plus scale salience thus predicts that sentences with numerals and universal modals should have exactly the range of interpretations in child language that are predicted by the scopal analysis, namely the ones that we saw in our experiments.

But more than that, an analysis based on exhaustification and scale salience also makes the same predictions as the scopal analysis about sentences involving numerals and existential modals. The two potential parses of *Gonzo is allowed* *to use three lemons* are shown in (20a) and (21a); the crucial one is (21a), which derives the upper bounded truth conditions in (21b).

(20) a. allowed to [exh [Gonzo use two lemons]]

b. $\Diamond[\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \neg\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) > 2]]]$

(21) a. exh [allowed to [Gonzo use two lemons]]

b. $\Diamond\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) = 2] \wedge \neg\Diamond\exists x[\textbf{used}(x)(\textbf{g}) \wedge \textbf{lemons}(x) \wedge \#(x) > 2]$

The predictions of the scopal analysis and an analysis based on exhaustification and scale salience are therefore the same for the crucial child language examples involving numerals and existential modals: the former predicts upper bounded interpretations because such meanings are derived without invoking the implicature system; the latter because children have access to alternatives based on numerals at an age when they do not have access to alternatives based on other scalar expressions. What, then, can we conclude about the semantics of numerals from our study?

First, it should be acknowledged that an analysis based on exhaustification is fundamentally a *semantic* analysis of bounding implications, not a pragmatic one, since bounding implications are introduced compositionally by exh. It is moreover a *scopal* analysis, since the crucial factor in deriving different kinds of bounding implications is the relative scope of exh and (in these examples) universal vs. existential modals. All other things being equal, the exhaustification analysis derives truth conditions for a structure like (22a), where Num denotes the number n and is the basis for calculation of the alternatives to S, that are identical to the truth conditions derived by degree quantifier analysis for (22b), where Num denotes the degree quantifier $\lambda P.max\{m \mid P(m)\} = n$.

(22) a. [... exh [$_S$ ... Num ...] ...]

b. [... Num [$_S$ ... t ...] ...]

The two approaches are thus similar both in the general claim that bounding implications are a scopal phenomenon and in their empirical predictions. Where they differ theoretically is in how the meanings are built up. In the exhaustivity analysis, there are three moving parts, each of which has plausible independent justification: exhaustification, calculation of alternatives, and a semantics for numerals that delivers lower bounded content. In the degree quantifier analysis, all of these pieces are effectively built into the denotation of the numeral. This theoretical difference points to the crucial potential empirical difference between the approaches: the analyses make different predictions about the interpretation of numerals in contexts in which exhaustification or alternative calculation is blocked or disrupted.

This, of course, is the very distinction that we relied on when we interpreted children's behavior in the upper bound condition of Experiment 1 as evidence in support of the scopal analysis of numerals and against the local analysis, and that same reasoning would likewise favor the degree quantifier analysis over the exhaustification analysis if it were the case that children's behavior with non-numeral scalar terms indicates that their general capacity to exhaustify or compute alternatives is disrupted. So the crucial question is whether the systematic difference between the interpretation of numerals and other scalar terms by children reflects the fact that children have acquired a meaning for numerals that has effectively "precompiled" exhaustification over numbers (degree quantification), or rather indicates that they have variable capacity for the calculation of alternatives (exhaustification and scale salience).

When we step back and look at the larger picture, we believe that there are reasons to endorse the former account. The most compelling reason, in our view, is that an account of the child language data in terms of scale salience says nothing about differences between numerals and other scalar terms in adults, who presumably are capable of computing alternatives for the full range of scalar terms, yet such differences are well documented. For example, in addition to the many examples which show that upper bounded readings of numerals are retained in grammatical contexts in which upper bounded readings of other scalar terms disappear (see Kennedy 2013 for an overview), Huang et al. (2013) provide evidence that adults assign upper bounded interpretations to numerals but not to other scalar terms in a task designed to suppress implicature calculation. Additionally, Marty et al. (2013) show that under working memory load, upper bounded readings of non-numeral scalar terms decrease, plausibly because memory load impacts the calculation of alternatives, but under the same conditions, upper bounded readings of numerals actually increase.[10] In order to account for such differences in terms of scale salience, one would need to say that, in adult language, alternatives based on numeral scales are different from those based on other quantitative scales in that the former are always accessible, and are insensitive to whatever contextual or processing factors may disrupt access to the latter. But this move effectively stipulates that exhaustification and alternative calculation happen in sentences containing numerals regardless of factors that otherwise influence such calculations, which significantly weakens the appeal of this type of analysis. On the other hand, this difference is exactly what we expect in the degree quantifier analysis, in which these calculations

[10] Panizza, Chierchia, and Clifton (2009), in contrast, found a penalty for upper bounded meanings in a reading time study, and a preference for lower bounded meanings in downward entailing contexts. The latter result is consistent with a general preference for stronger meanings, given the possibility of deriving lower bounded meanings from the upper bounded degree quantifier meaning (see note 3), but the former finding appears to conflict with results like those described in Huang et al. (2013) and Marty et al. (2013), as Panizza et al. themselves note.

are "built in" to the meaning of the numeral. We submit that the most theoretically parsimonious account of the child language pattern is one in which children acquire this very same meaning.[11]

The degree quantifier meaning is, moreover, fully compatible with Barner and Bachrach's observations about the developmental path of numeral acquisition. If a child has successfully acquired the degree quantifier meaning for a particular numeral, say *two*, she has associated it with a denotation of the sort shown in (23)

(23) $\lambda P.max\{n \mid P(n)\} = 2$

Here '2' stands for a model-theoretic object: the unique degree that represents the value of the '#' function when applied to pluralities consisting of the join of two atomic objects, i.e. the number two. There are, therefore, two key parts to the acquisition of numeral semantics: assigning the appropriate quantificational denotations to the appropriate numerals, *and* building up the set of model-theoretic objects on which those denotations are based, i.e. learning numbers. Given that the former is dependent on the latter, we can explain B&B's observations by supposing that in the early stages of numeral acquisition (before they become CP knowers), children ini-

[11] The fact that children fail to compute free choice inferences with numerals also constitutes a potential argument against an analysis based on scale salience. As we pointed out in note 7, the account of free choice inferences for numerals, unlike e.g. the account of free choice inferences for disjunction, requires access to lexical alternatives. If children have access to lexical alternatives for numerals per scale salience, and if the computation of free choice inferences for numerals involves the same principles as the computation of upper bounding inferences, then children should be equally good at computing both kinds of inferences. But this is not what we found in Experiment 1: they are significantly better at upper bounding inferences than at free choice inferences. However, we also pointed out in note 7 that the account of free choice inferences with numerals may involve a different kind of computation from upper bounding inferences. Whether children's failure to reliably compute free choice inferences with numerals constitutes an argument against scale salience or not must wait on a resolution of this question.

tially analyze numerals as denoting numbers. This position is consistent with that of Wynn (1992), who argued based on experimental evidence that, even at a very early age, children seem to know that numerals pick out quantities rather than individuals or properties of individuals (see also Bloom and Wynn 1997; Syrett, Musolino, and Gelman 2012), even when they don't know which numerosities they pick out. This position is moreover plausible syntactically, since expressions that denote atomic types α generally have the same distributions as their quantificational counterparts of type $\langle\langle\alpha, t\rangle, t\rangle$; and semantically, since saturation of the noun phrase-internal degree position with a number-denoting numeral derives lower bounded truth conditions, as we observed earlier when discussing Nouwen's (2010) version of the local analysis. We may even assume with B&B that the move from a number denotation for numerals (e.g., 2 for *two*) to the corresponding degree quantifier denotation (the one in (23)) correlates with the acquisition of the number denotation for the next expression in the counting list (3 for *three*), since it is precisely in virtue of the identification of $n + 1$ as a potential, greater value than n that the maximality component of the degree quantifier denotation gains its informational force.[12]

5 Conclusion

In this paper, we compared two semantic accounts of bounding implications in sentences containing cardinal numerals, one in which the distinction between lower bounded and upper bounded interpretations is due to lexical ambiguity, underspecification or other compositional factors that are "local" to the numeral and the expression that introduces the things that it counts, and one in which it reflects the sco-

[12]An open question is how this account of the developmental path — as well as Barner and Bachrach's, to which it is largely isomorphic — relates to Carey's (2004) suggestion that children initially organize numerals according to a more basic system for representing sets of objects.

pal interactions of numerals qua degree quantifiers and other expressions in the sentence. We demonstrated that the two types of analyses make different predictions about sentences in which numerals are embedded under existential root modals, with the local analysis deriving upper bounding via the implicature mechanisms, and the scopal analysis deriving the upper bound semantically, via scope taking, just as in other contexts. We then described a set of experiments which examined how a population of speakers that does not readily calculate scalar implicatures — young children — interpreted the crucial examples. These experiments demonstrate that children behave just like adults in assigning upper bounds to the relevant cases, as predicted by the scopal analysis but not by the local analysis. Finally, we considered an alternative account of the child language data based on exhaustification and numeral scale salience, and rejected it because, unlike the scopal analysis, it does not generalize to an explanation of the use and interpretation of numerals in adult language.

References

Austin, Jenniver, Liliana Sanchez, Kristen Syrett, Anne Lingwall, and Silvia Pérez-Cortes. 2015. Quantity implicatures in English and monolingual Spanish-English bilingual children. Poster presented at BUCLD 40.

Bar-Lev, Moshe, and Danny Fox. 2017. Universal free choice and innocent exclusion. In *Proceedings of Semantics and Linguistic Theory 27*, ed. Dan Burgdorf, Jacob Collard, Sireemas Maspong, and Brynhildur Stefánsdóttir, 95–115. Linguistic Society of America.

Barner, David, and Asaf Bachrach. 2010. Inference and exact numerical representation in early language development. *Cognitive Psychology* 60:40–62.

Barner, David, Neon Brooks, and Alan Bale. 2010. Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition* 118:84–93.

Bloom, Paul, and Karen Wynn. 1997. Linguistic cues in the acquisition of number words. *Journal of Child Language* 24:511–533.

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25:93–139.

Bultinck, Bert. 2005. *Numerous meanings: The meaning of English cardinals and the legacy of Paul Grice*. Amsterdam: Elsevier.

Carey, Susan. 2004. Bootstrapping and the origin of concepts. *Daedalus* 59–68.

Carston, Robyn. 1998. Informativeness, relevance and scalar implicature. In *Relevance theory: Applications and implications*, ed. Robyn Carston and S. Uchida, 179–236. Amsterdam: John Benjamins.

Chierchia, Gennaro. 2006. Broaden your views. implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37:535–590.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, and Luisa Meroni. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedngs of the 25th Boston University Conference on Language Development*, ed. Anna H.-J. Do, L. Domingues, and A. Johansen, 157–168. Somerville, MA: Cascadilla Press.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In *Semantics: An international handbook of natural language meaning*, ed. Claudia Maienborn, Klaus von Heusinger, and Paul Portner, volume 3. de Gruyter.

Crain, Stephen, and Rosalind Thornton. 1998. *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*.

Cambridge, Mass.: MIT Press.

Cresswell, M. J. 1976. The semantics of degree. In *Montague grammar*, ed. Barbara Partee, 261–292. New York: Academic Press.

Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, ed. Uli Sauerland and Penka Stateva, 71–120. Palgrave MacMillan.

Franke, Michael. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4:1–82.

Frege, Gottlob. 1980 [1884]. *The foundations of arithmetic (Grundlagen der Arithmetik)*. Evanston, IL: Northwestern University Press.

Gazdar, Gerald. 1977. Implicature, presupposition and logical form. Doctoral Dissertation, University of Sussex.

Gelman, Rochel, and C.R. Gallistel. 1978. *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Geurts, Bart. 2006. Take 'five'. In *Non-definiteness and plurality*, ed. Svetlana Vogeleer and Liliane Tasmowski, 311–329. Amsterdam: Benjamins.

Grice, H. Paul. 1975. Logic and conversation. In *Studies in the way of words*, ed. H. Paul Grice. Cambridge, MA: Harvard University Press.

Gualmini, Andrea, Stephen Crain, Luisa Meroni, Gennaro Chierchia, and Maria Teresa Guasti. 2001. At the semantics/pragmatics interface in child language. In *Proceedings from SALT XI*, ed. R. Hastings, Brendan Jackson, and Zsofia Zvolensky, 231–247. Ithaca, NY: CLC Publications.

Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard, and Danny Fox. 2008. The Question-Answer Requirement for scope assignment. *Natural Language Semantics* 16:205–237.

Guasti, Maria Teresa, Gennaro Chierchia, Stephen Crain, Francesa Foppolo, Andrea Gualmini, and Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20:667–696.

Hackl, Martin. 2001. Comparative quantifiers and plural predication. In *Proceedings of WCCFL XX*, ed. K. Megerdoomian and L.A. Bar-el, 234–247. Somerville, MA: Cascadilla Pres.

Hirschberg, Julia. 1985. A theory of scalar implicature. Doctoral Dissertation, University of Pennsylvania.

Horn, Laurence. 1972. On the semantic properties of logical operators in English. Doctoral Dissertation, University of California, Los Angeles.

Horn, Laurence. 1992. The said and the unsaid. In *Proceedings of Semantics and Linguistic Theory 2*, ed. Chris Barker and David Dowty, 163–192. Columbus, OH: OSU Working Papers in Linguistics.

Huang, Yi Ting, Elizabeth Spelke, and Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9:105–129.

Hurewitz, Felicia, Anna Papafragou, Lila Gleitman, and Rochel Gelman. 2006. Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development* 2:77–96.

Ionin, Tania, and Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23:315–360.

Kennedy, Christopher. 2013. A scalar semantics for scalar readings of number words. In *From grammar to meaning: The spontaneous logicality of language*, ed. Ivano Caponigro and Carlo Cecchetto, 172–200. Cambridge University Press.

Kennedy, Christopher. 2015. A “de-Fregean” semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8:1–44.

Koenig, Jean-Pierre. 1991. Scalar predicates and negation: Punctual semantics and interval interpretations. In *Proceedings of CLS 27*, ed. Lise Dobrin, Lynn Nichols, and Rosa Rodriguez, 140–155. Chicago Linguistic Society, Chicago, IL: Chicago Linguistic Society.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and contextual expression*, ed. Renate Bartsch, Johann van Benthem, and Peter van Emde Boas, 75–115. Stanford, CA: CSLI Publications.

Krifka, Manfred. 1998. At least some determiners aren’t determiners. In *The semantics/pragmatics interface from different points of view*, ed. Ken Turner, 257–291. Oxford, UK: Elsevier.

Lidz, Jeffrey, and Julien Musolino. 2002. Children’s command of quantification. *Cognition* 84:113–154.

Marty, Paul, Emmanuel Chemla, and Benjamin Spector. 2013. Interpreting numerals and scalar items under memory load. *Lingua* 133:152–162.

Musolino, Julien. 2004. The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition* 93:1–41.

Musolino, Julien, and Jeffrey Lidz. 2006. Why children’t aren’t universally successful with quantification. *Linguistics* 44:817–852.

Nouwen, Rick. 2010. Two kinds of modified numerals. *Semantics and Pragmatics* 3:1–41.

Noveck, Ira. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78:165–188.

Panizza, Daniele, Gennaro Chierchia, and Charles Jr. Clifton. 2009. On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language* 503-518.

Papafragou, Anna. 2006. From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language* 33:721–757.

Papafragou, Anna, and Julien Musolino. 2003. Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognitiong* 86:253–282.

Papafragou, Anna, and Niki Tantalou. 2004. Children's computation of implicatures. *Language Acquisition* 12:71–82.

Sadock, Jerrold. 1984. Whither Radical Pragmatics. In *Meaning, form and use in context: Linguistic applications*, ed. Deborah Schiffren, 139–149. Washington, DC: Georgetown University Press.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27:267–391.

Scharten, Rosemarijn. 1997. Exhaustive interpretation: A discourse-semantic account. Doctoral Dissertation, Katholieke Universiteit Nijmegen.

Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar, and Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 305-352.

Skordos, Dimitrios, and Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153:6–18.

Smith, Carol. 1980. Quantifiers and question answering in young children. *Journal of Experimental Child Psychology* 30:191–205.

Spector, Benjamin. 2013. Bare numerals and scalar implicatures. *Language and Linguistics Compass* 7:273–294.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3:1–77.

Stiller, Alex, Noah Goodman, and Michael Frank. 2015. Ad-hoc implicature in preschool children. *Language Learning and Development* 11.

Syrett, Kristen. 2013. The role of cardinality in the interpretation of measurement expressions. *Language Acquisition* 30:228–240.

Syrett, Kristen, and Jeffrey Lidz. 2009. QR in child grammar: Evidence from Antecedent-Contained Deletion. *Language Acquisition: A Journal of Developmental Linguistics* 16:67–81.

Syrett, Kristen, Julien Musolino, and Rochel Gelman. 2012. How can syntax support number word acquisition? *Language Learning and Development* 146-176.

Syrett, Kristen, and Roger Schwarzschild. 2009. The representation and processing of measure phrases by four-year-olds. Paper presented at the 34th Annual Boston University Conference on Language Development.

Tieu, Lyn, Jacopo Romoli, Peng Zhou, and Stephen Crain. 2015. Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics* 33:269–298.

Wynn, Karen. 1990. Children's understanding of counting. *Cognition* 36:155–193.

Wynn, Karen. 1992. Children's acquisition of number words and the counting system. *Cognitive Psychology* 24:220–251.

Zhou, Peng, Jacopo Romoli, and Stephen Crain. 2013. Children's knowledge of free choice inferences. In *Proceedings of Semantics and Linguistic Theory 23*, ed. Todd Snider, 632–651. Linguistic Society of America.

Appendix: Complete List of Lead-Ins and Target Sentences for Experiment 1, Experiment 1 follow-up, and Experiment 2

Experiment 1

Control Items

The choice of lexical item in brackets was designed to favor a "yes" or "no" response.

(24) a. Remember, Ruby asked Max to draw a [cat/dog]. Is what Max did ok?

b. Remember, Dora wanted some fruit and asked Boots to give her [an apple/a banana]. Is what Boots did okay?

c. Remember, Mickey told Goofy to use the [red/blue] Play-Doh to make a car. Is what Goofy did ok?

d. Remember, Joe told Blue to draw a [star/flower]. Is what Blue did okay?

Test items

Numeral-Count Noun scenarios

(25) *Making Lemonade*

a. Gonzo was making lemonade and needed to use Kermit's lemons. Remember, Kermit said to Gonzo, "I need to leave some lemons for myself so ... you *are allowed to* use two lemons." Is what Gonzo did okay?

b. Gonzo was making lemonade and needed to use Kermit's lemons. Remember, Kermit said to Gonzo, "You want it to taste lemony so ... you *have to* use two lemons." Is what Gonzo did okay?

(26) *Books before Bedtime*

a. The Man in the Yellow Hat wanted George to go to bed soon, but George wanted to read first. Remember, The Man in the Yellow Hat told George, "You have to go to bed soon so ... you *are allowed to* read two books." Is what George did okay?

b. George wanted to go to bed, but The Man in the Yellow Hat wanted him to read first. Remember, The Man in the Yellow Hat said to George, "It's important to practice reading so ... you *have to* read two books." Is what George did okay?

(27) *Bunnies and Carrots*

a. Benjamin Bunny wanted to eat some of Peter Rabbit's carrots. Remember, Peter Rabbit said to him "I need to keep enough carrots to make dinner so ... you *are allowed to* have two carrots." Is what Benjamin Bunny did okay?

b. Benjamin Bunny needed energy for his trip. Remember, Peter Rabbit said to him "You want to have enough energy for the whole trip so ... you *have to* have two carrots." Is what Benjamin Bunny did okay?

Measure Phrase-Mass Noun scenarios

(28) *Filling the Pitcher*

a. Elmo wanted to help get the picnic lunch ready by filling a pitcher of water. Remember, Big Bird said to Elmo, "We don't want the pitcher to be too heavy to carry so ... you *are allowed to* fill the pitcher with two inches of water." Is what Elmo did ok?

b. Elmo wanted to help get the picnic lunch ready by filling a pitcher of water. Remember, Big Bird said to Elmo, "We want there to be enough for everyone so ... you *have to* fill the pitcher with two inches of water." Is what Elmo did ok?

(29) *Filling the Pool*

a. Sister Bear was learning how to fill the pool up on her own. Remember, Brother Bear said, "You don't want the water to spill out when you splash so ... you are *allowed to* fill the pool with two feet of water." Is what Sister Bear did ok?

b. Sister Bear learning how to fill the pool up on her own. Remember, Brother Bear said, "You want enough water to be able to splash so ... you *have to* fill the pool with two feet of water." Is what Sister Bear did ok?

(30) *Stuffing for Bears* (Faulty item in original Experiment 1)

a. Bob was showing Wendy how to make a toy bear. Remember, Bob said, "You want your bear to be cuddly, but not too stiff so ... you are *allowed to* use two inches of stuffing." Is what Wendy did ok?

b. Bob was showing Wendy how to make a toy bear. Remember, Bob said, "Stuffing is very important in order to make your bear cuddly so ... you *have to* use two inches of stuffing." Is what Wendy did ok?

(31) *Packing building materials* (replacement item in follow-up to Experiment 1)
Bob was showing Wendy how to pack stuffing in a container to ship building materials. Remember, Bob said, "You want the building materials to be really protected, but you still want enough room for the building materials to go in and not be cushioned too much, so you're *allowed to* use 2 inches of stuffing." Is what Wendy did ok?

Experiment 2

Control Items

The choice of lexical item in brackets was designed to favor a "yes" or "no" response.

(32) a. Remember, Ruby told Max that if he drew a [cat/dog] for her, then he could have a turn in the sandbox. Can Max have a turn in the sandbox?

b. Remember, Dora asked Boots to find [an apple/a banana] in his basket. Can Boots be in charge of making the fruit salad?

c. Remember, Mickey told Goofy to use the [red/blue] Play-Doh to make his racecar. Can Goofy race his car?

d. Remember, Joe told Blue to draw a [star/flower]. Can Blue use the special chalk on the sidewalk?

Test Items

Numeral-Count Noun scenarios

(33) *Vines*

Marvin wanted to play with Tarzan in the jungle, but Tarzan can only take fast monkeys into the jungle with him. Remember, Tarzan said you *hafta* swing on three vines in a row to show you are a fast monkey. Can Marvin play with Tarzan in the jungle?

(34) *Lily Pads*

Freddy's brother Frankie invited him to play in the deep pond with the big kid frogs, but only strong jumpers can play there. Remember, Frankie said you *hafta* jump over three lily pads to show you are a strong jumper. Can Freddy go play in the deep pond?

(35) *Vegetables*

Remember, Diego said that kids in his class *hafta* eat three vegetables in order to get a "healthy kid" sticker. Can Diego get a sticker?

Measure Phrase scenarios

(36) *Board Game*

Kevin and Amanda weren't sure if they could play the board game. Remember, the box said you *hafta* be three years old to play. Can Kevin and Amanda play the game?

(37) *Monkey Bars*

Emily Elizabeth wanted to play on the monkey bars for the first time, but she wasn't sure if she could reach them. Remember, Charley said you *hafta* reach three feet up to be able to play on the monkey bars. Can Emily Elizabeth play on the monkey bars?

(38) *Roller Coaster*

Elmo wasn't sure if he could go on the roller coaster, so Big Bird was helping him measure himself. Remember, the sign said you *hafta* be three feet tall to ride. Can Elmo go on the rollercoaster?