Numerals denote degree quantifiers: Evidence from child language

Christopher Kennedy and Kristen Syrett
University of Chicago and Rutgers University

January 20, 2016

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 scalar implicatures of the sort seen with other scalar terms. However, not all semantic analyses of numerals designed to introduce upper bounding entailments in simple sentences make the same predictions about upper bounding implications in other sentences. Here we focus on the case of upper bounded construals of numerals embedded under existential root modals, which are derived as entailments by a semantic analysis of numerals as generalized quantifiers over degrees (Kennedy 2013, 2015), but can only be derived as scalar implicatures by other semantic analyses. We provide experimental evidence from child language that upper bounding implications in such cases are entailments and not implicatures, thereby providing support for the degree quantifier analysis.

1 Numerals and bounding implications

1.1 Two semantic accounts of upper bounding

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 Dustin got on the last day of the season; i.e., (1) is taken to mean that Dustin got exactly three hits.

(1) Dustin 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 Dustin 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) Dustin 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 "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 Dustin has to get is actually expected, given the use of the universal modal *have to*. The same reasoning that derives the implicature that Dustin didn't get four (or more) hits from an utterance of (1) derives the implicature that Dustin 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) Dustin has to get three hits on the last day of the season in order to finish with a batting average of .345.

Since Dustin'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 Dustin doesn't have to get four or more hits to finish with an average of .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 a two-sided interpretation of a numeral, are one instance

[1]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 exhaustivity operator in the syntax, as in Chierchia 2006; Fox 2007; Chierchia, Fox, and Spector 2012; Spector 2013; etc. (But 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 adults.

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 kind of approach includes several types of analyses that are distinct from each other in many respects, but share the assumption that bounding inferences are introduced "locally" 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 are determiners that introduce two-sided content exclusively (Koenig 1991; Breheny 2008); analyses in which numerals or some part of the larger nominal constructions in which they appear are lexically ambiguous between two-sided 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 the lexical 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 lexical 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 comes in two versions: 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. $[\![three\ \text{MANY}_w]\!] = \lambda P \lambda Q.\exists x[P(x) \wedge \#(x) = 3 \wedge Q(x)]$
b. $[\![three\ \text{MANY}_s]\!] = \lambda P \lambda Q.\exists! x[P(x) \wedge \#(x) = 3 \wedge Q(x)]$

(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. (5b) introduces two-sided 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. $[\![\text{Dustin has to get three MANY}_w \text{ hits}]\!] = \Box[\exists x[\textbf{hits}(x) \wedge \#(x) = 3 \wedge \textbf{got}(x)(\textbf{dustin})]]$
b. $[\![\text{Dustin has to get three MANY}_s \text{ hits}]\!] = \Box[\exists! x[\textbf{hits}(x) \wedge \#(x) = 3 \wedge \textbf{got}(x)(\textbf{dustin})]]$

The second approach is one in which bounding entailments at the sentential level arise not directly through the composition of an ambiguous or underspecified numerical constituent with its nominal argument, but rather through scopal interactions between the numeral and other expressions in the sentence. On this view, numerals are neither determiners nor cardinality predicates nor singular terms (denoting numbers), but 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

[2]Two comments. First, in analyses in which numerals unambiguously introduce two-sided 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 makes the same predictions as the ambiguity analysis about 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*.

“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) ⟦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 derive the two-sided truth conditions in (8c).[3]

(8) a. three [Dustin got t MANY hits]
b. ⟦three⟧$(\lambda n.\exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{dustin})])$
c. $max\{n \mid \exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{dustin})]\} = 3$

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 scope over the universal modal *have to*, as in (9a), and the upper-bounded/two-sided 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 [Dustin get t MANY hits]]⟧ =
$max\{n \mid \Box[\exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{dustin})]]\} = 3$
b. ⟦has to [three [Dustin get t MANY hits]]⟧ =
$\Box[max\{n \mid \exists x[\textbf{hits}(x) \wedge \#(x) = n \wedge \textbf{got}(x)(\textbf{dustin})]\} = 3]$

1.2 Upper bounding 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 believe, argues in favor of the degree quantifier analysis. The crucial facts involve sentences in which a numeral is embedded in the complement of an existential modal, such as (10).

[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). Experimental results reported in Marty et al. 2013 suggest that this is a non-default, marked option.

[4] (9a) says that three is the maximum n such that in every world in the modal domain (worlds in which Dustin wins the batting title), there’s a group of hits of size n that Dustin 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.

(10) Dustin 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 Dustin can make, but unlike (1) and (3), it does not impose a lower bound: it allows for the possibility of Dustin making two, one or zero errors.

In the degree quantifier analysis of numerals, the upper bound reading of (10) is derived compositionally by scoping the numeral over the modal, which returns the truth conditions shown (11).

(11) $[\![$three [can [Dustin make t MANY errors]]$]\!]$ =
$max\{n \mid \Diamond[\exists x[\textbf{errors}(x) \wedge \#(x) = n \wedge \textbf{make}(x)(\textbf{dustin})]]\} = 3$

(11) says that three is the maximum n such that there is a world in the relevant modal domain in which Dustin makes at least n errors, which rules out the possibility that he makes more than three errors, but allows for him to make fewer. 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 Dustin makes exactly three errors.

(12) $[\![$can [three [Dustin make t MANY erros]]$]\!]$ =
$\Diamond[max\{n \mid \exists x[\textbf{errors}(x) \wedge \#(x) = n \wedge \textbf{make}(x)(\textbf{dustin})]\} = 3]$

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

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

(13) a. $[\![$Dustin can make three MANY$_w$ errors$]\!]$ =
$\Diamond[\exists x[\textbf{errors}(x) \wedge \#(x) = 3 \wedge \textbf{make}(x)(\textbf{dustin})]]$

b. $[\![$Dustin can make three MANY$_s$ errors$]\!]$ =
$\Diamond[\exists! x[\textbf{errors}(x) \wedge \#(x) = 3 \wedge \textbf{make}(x)(\textbf{dustin})]]$

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 Dustin cannot make four, five, etc. errors, which derives the upper bound.

The crucial difference between the two approaches, then, is that the degree quantifier analysis derives an upper-bounded interpretation for (10) semantically, by scoping the numeral over the modal, while the lexical ambiguity 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 degree quantifier analysis predicts that upper-bounding inferences will be retained, all other things being equal; the lexical ambiguity 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 difficultly computing upper-bounding implicatures in contexts in which adults (virtually automatically) generate such meanings in accord with neo-Gricean reasoning based on the Maxim of Quantity (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 using ad-hoc and non-lexical-based scales (Barner et al. 2010; Papafragou and Tantalou 2004; Stiller, Goodman, and Frank 2015) 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 two-sided 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 two-sided interpretations of numerals and for a semantic one.[5]

It is not the case that children always assign two-sided 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*
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

[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 a grammatical theory of scalar implicature make similar predictions to the degree quantifier 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 degree quantifier analysis.

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 lexical ambiguity and degree quantifier analyses of numeral meaning. 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 degree quantifier analysis, this meaning is derived semantically; in the ambiguity analysis, it can only be derived pragmatically via Quantity reasoning. Given that children generally do not calculate upper bounding implicatures, the lexical ambiguity analysis predicts that they should have difficulty assigning upper-bounded interpretations to sentences like (14b). The degree quantifier analysis, on the other hand, predicts that they should have no trouble with such a meaning, assuming that they have acquired whatever general principles are necessary for generating scopal interactions in the first place.

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 ($>n$). The crucial question concerns judgments about existential root modal sentences like (14b) as descriptions of $>n$ scenarios. The degree quantifier analysis predicts that such sentences should be systematically rejected as descriptions of these scenarios; the lexical ambiguity analysis predicts that they should be accepted, since they are semantically true and only false in virtue of strengthening by scalar implicature. In the next section, we present experimental results that confirm the predictions of the degree quantifier analysis of numeral meaning and bounding implications.

3 Experiments

3.1 Experiment 1

The purpose of Experiment 1 was to determine whether children assign an upper bounded interpretation to sentences in which a numeral is embedded under an existential root modal, by asking whether they reject such sentences as descriptions of scenarios in which the upper bound is surpassed. The crucial sentences were compared with sentences involving a universal root modal, which should give rise to 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 using Powerpoint on an iMac for adults and on a MacBook Pro for children. 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.

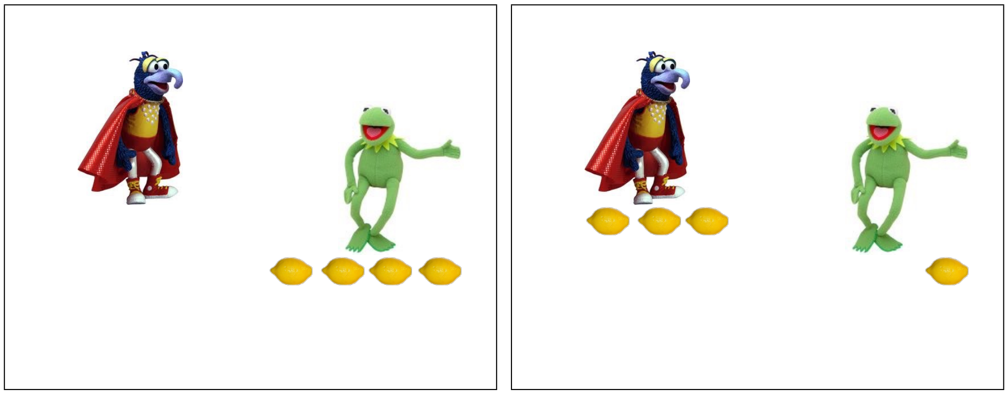
The experimental session began with two training items. The test session that followed consisted of six test items and four control items, all with the same basic structure, as described above. Each test item involved instructions from the second character that featured a modal and a numerical expression. 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 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 two-sided 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.

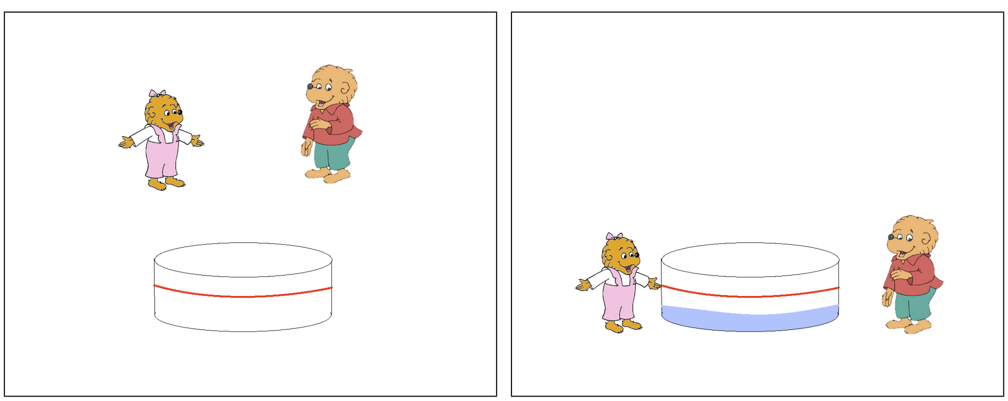
The measure phrase plus mass noun scenarios were similar. In one scenario, Sister Bear and Brother Bear are playing outside on a hot day, and Sister Bear asks Brother Bear if she can use the pool. Brother Bear agrees, and says she can fill it up



(a) Beginning scene (b) Ending scene

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

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.



(a) Beginning scene (b) Ending scene

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

	UPPER BOUND		
	<2	=2	>2
Adults	78.1%	100.0%	3.1%
MASS	68.8%	100.0%	6.3%
COUNT	87.5%	100.0%	0.0%
Children	43.8%	100.0%	21.9%
MASS	43.8%	100.0%	43.8%
COUNT	43.8%	100.0%	0.0%

Table 1: Mean percentage acceptance of second character's actions for the mass and count items 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. (See Appendix B for responses.) The exceptional item involved the 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. We suspect that children deemed it permissible for the character to exceed the "two inches of stuffing" limit in this scenario either because the stuffing could be pushed down to go below the two inch line, or because there were no negative consequences for filling a bear with additional stuffing: the bear was still soft. This kind of reasoning was reflected in one child's justification for acceptance, as shown in the following exchange ("C" is the child; "E" is the experimenter):

(15) C: (nods) Because she wanted a bear.
E: Do you think she used the right amount of stuffing?
C: Yeah.
E: How come?
C: Because the bear's soft.

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.

	UPPER BOUND (REVISED)		
	<2	=2	>2
Adults	87.5%	100.0%	0.0%
MASS	87.5%	100.0%	0.0%
COUNT	87.5%	100.0%	0.0%
Children	41.7%	100.0%	4.2%
MASS	37.5%	100.0%	12.5%
COUNT	43.8%	100.0%	0.0%

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

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 (now 41.7% v. 4.2%, respectively) (W=66, z=2.91, p=.004, two-tailed).[6]

We now turn to the lower bound condition. Table 3 presents the results, showing percentages of acceptance for the full set of experimental items and for the revised set that we evaluated in the upper bound condition, which excludes the problematic “Stuffing the Bear” items.

	LOWER BOUND (ALL ITEMS)			LOWER BOUND (REVISED)		
	<2	=2	>2	<2	=2	>2
Adults	12.5%	96.9%	25.0%	12.5%	96.9%	33.3%
MASS	6.3%	100.0%	6.3%	0.0%	100.0%	12.5%
COUNT	18.8%	93.8%	43.8%	18.8%	93.8%	43.8%
Children	25.0%	100.0%	31.3%	16.7%	100.0%	20.8%
MASS	31.3%	100.0%	50.0%	12.5%	100.0%	37.5%
COUNT	18.8%	100.0%	12.5%	18.8%	100.0%	12.5%

Table 3: Mean percentage acceptance of second character’s actions for the mass and count items in the lower bound condition in Experiment 1, with and without the “Stuffing the Bear” items

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.

DISCUSSION Once we correct for the item effect in the upper bound condition, the results of Experiment 1 indicate that children (as well as adults) systematically assign upper bound interpretations to sentences in which numerals appear in the scope of an existential root modal. This result is expected on the degree quantifier 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. In contrast, the result is unexpected on semantic analysis of numerals in which such readings can only be derived pragmatically, given independent experimental results showing

[6]Children’s lack of willingness to accept the target sentence for the <2 items (relative to that of the adults) is also attested in Austin, Sanchez, Syrett, Lingwall, and Pérez-Cortes 2015, for English and Spanish-English bilingual children of the same age, for control sentences such as, *‘You may take {all, two} of the N’* in a context in which a conversational participant takes less than the amount indicated by the quantifier in the speaker’s utterance.

that children in this age group tend not to calculate upper bounding implicatures in tasks like this one.

However, this conclusion is tempered somewhat by the results we obtained in the lower bound condition of Experiment 1. On the degree quantifier analysis, assigning a numeral wide scope over a universal root modal should result in lower-bounded sentential truth conditions, yet both children and adults in our study were reluctant to accept the character's actions for the >2 items. One possible explanation for this result is a preference to assign narrow scope to the numeral, since this derives stronger, two-sided truth conditions for the sentence. And in fact, justifications provided by the subjects who rejected the character's actions in the >2 cases indicated that the character took or did too much, didn't do or take the right amount, or took/did more than what was needed, suggested, indicated, or said. (See Appendix B.) If this explanation is correct, we hypothesized, then manipulating the context to promote a lower bound interpretation should cause it to emerge. We tested this hypothesis in Experiment 2.

3.2 Experiment 2

The purpose of Experiment 2 was to manipulate the experimental contexts in the test session in order to highlight the lower-bounded 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 needed to do exactly what the first character said, thereby favoring a two-sided interpretation and decreasing the likelihood of participants accepting scenarios when the second character did or took more than was specified. 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 a two-sided 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.).

Fourth, we reduced *have to* to *hafta* based on an intuition that the non-reduced form might favor a two-sided reading. (See related discussion in Grodner et al. (2010, footnote 3) concerning scalar inferences with reduced *some* (*sm*) versus the partitive *some of the*. The latter is claimed to disambiguate to the strong, upper-bounded reading.)

Finally, 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.

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

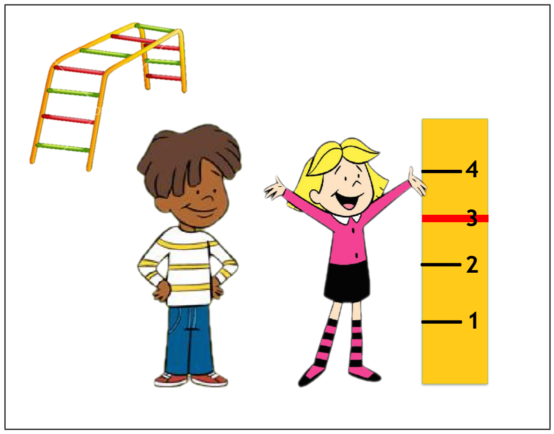


Figure 3: Experiment 3, >3 mass noun scenario

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 3.

RESULTS The results for Experiment 2, compared with those of the lower bound condition in Experiment 1 (all items), are presented in Table 4. It is immediately apparent that the story manipulations designed to increase the salience of the lower bound and permissibility of not meeting the exact numeral worked: the percentage of acceptance for all items increased sharply in Experiment 2. The difference between the results for Experiments 1 and 2 is 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). Examples of justifications from adults and children are included in Appendix C.

Notice that the manipulations did not increase acceptance across the board: participants were unwilling to accept amounts *less than* the target amount, but were now willing to accept both the exact amount (regardless of whether it was referred to with a numeral word, e.g., *two*, or a pseudopartitive MP). Although the difference in responses between Experiments 1 and 2 is not significant for the <2 item (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.

	LOWER BOUND EXP. 1: *have to*			LOWER BOUND EXP. 2: *hafta*		
	<2	=2	>2	<2	=2	>2
Adults	12.5%	96.9%	25.0%	0.0%	100.0%	98.4%
MASS	6.3%	100.0%	6.3%	0.0%	100.0%	100.0%
COUNT	18.8%	93.8%	43.8%	0.0%	93.8%	96.9%
Children	25.0%	100.0%	31.3%	3.6%	89.3%	78.6%
MASS	31.3%	100.0%	50.0%	0.0%	92.9%	78.6%
COUNT	18.8%	100.0%	12.5%	7.1%	85.7%	100.0%

Table 4: Mean percentage acceptance of second character's actions in the lower bound condition of Experiment 1 and Experiment 2

.267), demonstrating more consistency in rejection of the character's actions in Experiment 2.

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 context supports such readings. As noted in the discussion of Experiment 1, lower bounded interpretations are weaker than two-sided interpretations, so our findings are consistent with a general preference for stronger readings. In the absence of contextual support for the weaker, lower bounded reading in Experiment 1, the two sided reading was strongly preferred; with the addition of contextual factors supporting a lower bounded reading in Experiment 2, this reading emerged.

It is also worth noting that the parallel behavior of adults and children in Experiment 2 and the lower bound condition of Experiment 1 provides further support for the general hypothesis that two-sided interpretations of sentences containing numerals are semantic, not pragmatic in nature, as both the degree quantifier and lexical ambiguity analyses maintain. If such interpretations were pragmatic, as in traditional neo-Gricean accounts, the fact that children as well as adults showed a preference for a two-sided interpretation in Experiment 1 would be surprising, given the general finding that children are less likely to calculate scalar implicatures than adults. If, on the other hand, both two-sided and lower bounded interpretations are generated semantically, and 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) — then their parallel behavior is expected.

4 Degree quantification vs. scale salience

We have argued that Experiments 1 and 2 provide further support for the conclusion that two-sided readings of sentences containing numerals are derived semantically, and provide support in particular for a degree quantifier analysis of the semantics of numerals over a lexical ambiguity 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 lexical ambiguity analysis, this reading can only be derived pragmatically as a scalar implicature, even assuming a two-sided semantics for the numeral; in the degree quantifier 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. We conclude from this result that such readings are derived in a way that does not engage the mechanisms involved in calculating scalar implicatures; since the degree quantifier analysis is the only existing semantic analysis of numerals that derives the right meaning without engaging these mechanisms, the results lead us to conclude that this analysis is the correct one.

A crucial assumption of our argument is that the large body of experimental work demonstrating children's reduced capacity to calculate scalar implicatures is general, and applies to all sentences containing a weak scalar term, including numerals (given the assumption that they can denote such terms, as in the lexical ambiguity analysis). This assumption has been challenged in recent work by Barner and Bachrach (2010), however, who interpret children's differential behavior with numerals and other scalar terms not as indicating that two-sided meanings for the former do not involve implicature, but rather as indicating that children's reduced capacity to calculate scalar implicatures does not extend to numerals. Specifically, 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* and *start* indicates a failure to construct the scalar alternatives for these expressions that the implicature mechanism needs in order to generate upper bounding implicatures. 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 *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 *one* and *two* but not higher numerals. This pattern continues until they jump to adult-like competence with all numerals and become "cardinal knowers," typically around age *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 a two-sided meaning for n. If n-knowers instead had acquired a two-sided 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 two-sided interpretations emerge as a scalar implicature that stronger alternatives are false.[7] 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. If B&B are correct, then children's behavior in the upper bound condition of Experiment 1 is exactly as expected, and the argument for the degree quantifier semantics of numerals appears to vanish: the degree quantifier analysis and Barner and Bachrach's "scale salience" version of the classic analysis make exactly the same predictions about the crucial examples with existential modals.

When we step back and look at the larger picture, however, we believe that there are reasons to resist both B&B's analysis of the difference between numerals and other scalar terms in child language in terms of scale salience and access to alternatives and their conclusion that the developmental path of numeral acquisition is incompatible with the acquisition of exact numeral meanings. First and foremost, B&B's analysis says nothing about differences between numerals and other scalar terms in *adult* language, yet such differences are well documented. For example, Huang et al. (2013) provide evidence that adults assign two-sided interpretations to numerals but not to other scalar terms in a task designed to suppress implica-

[7] Barner and Bachrach do not take a stand on the nature of the implicature mechanism (neo-Gricean vs. grammatical), but the fact that children preferred two-sided 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 a grammatical theory in which implicatures are generated by inserting a silent exhaustivity operator into the syntactic representation (Chierchia 2006; Fox 2007; Chierchia et al. 2012; Spector 2013, inter alia).

ture calculation, and Marty et al. (2013) show that under working memory load, two-sided readings of non-numeral scalar terms decrease, but two-sided readings of numerals actually increase. It is highly unlikely that these results reflect a difference in adults' capacity to construct scalar alternatives for numerals vs. other scalars; instead, taken alongside a host of examples showing that two-sided readings of numerals are retained in grammatical contexts in which two-sided readings of other scalar terms disappear (see Kennedy 2013 for an overview), they reinforce the conclusion that two-sided meanings of sentences containing numerals in adult language do not arise via scalar implicature (regardless of whether the mechanisms of implicature calculation are neo-Gricean or grammatical), but instead reflect the fact that the semantic content of the numeral introduces two-sided truth conditions. If this is correct, then certainly the most theoretically parsimonious account of the child language pattern is one in which children acquire this very same meaning; our experiments indicate that this meaning is the one posited by the degree quantifier analysis.

This meaning is, moreover, fully compatible with B&Bs 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 (16)

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

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 numerals: 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 saying that in the early stages of numeral acquisition (before they become cardinal principle knowers), children initially analyze numerals as denoting numbers. This is 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, saturation of the numeral's degree position with a number derives lower-bounded truth conditions, as we observed earlier when discussing Nouwen's (2010) lexical ambiguity analysis. We may even assume with B&B that the move from a singular term denotation (e.g., 2 for *two*) to the corresponding degree quantifier denotation (the one in (16)) correlates with the acquisition of the singular term 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.

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 or underspecification, and one in which it reflects the scope-taking options of numerals qua degree quantifiers. We demonstrated that the two types of analyses make different predictions about sentences in which numerals are embedded under existential modals, with the ambiguity account deriving upper bounding via the implicature mechanisms, and the degree quantifier account 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 they behave just like adults in assigning upper bounds to the relevant cases, as predicted by the degree quantifier analysis but not by the ambiguity analysis. Finally, we considered an alternative, pragmatic account of the child language data, and rejected it because, unlike the degree quantifier analysis, it does not generalize to an explanation of the use and interpretation of numerals in adult language.

Appendix A: Complete List of Lead-Ins and Target Sentences for Experiments 1 and 2

Experiment 1

Control Items

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

(17) 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

(18) *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."

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?

(19) *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."

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?

(20) *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 to carrots to make dinner so ... you *are allowed to* have two carrots."

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

(21) *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."

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?

(22) *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."

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?

(23) *Stuffing for Bears*

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."

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?

Experiment 2

Control Items

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

(24) 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

(25) *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?

(26) *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?

(27) *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

(28) *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?

(29) *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?

(30) *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?

Appendix B: Sample Justifications from Experiment 1 (C: Child, E: Experimenter

Justifications of rejections in the >2 scenarios from CHILDREN in the UPPER BOUND condition

Numeral-Count Noun Scenarios (*Bunnies and Carrots*)

(31) C: No, because he took three.
E: That wasn't okay?
C: (Shakes head no)
E: Why? What did Peter say?
C: Only two.

(32) C: No.
E: How come?
C: Cause he took three.
E: And how many did Peter say he could have?
C: Two.

(33) C: No, he took too much!
E: I remember he said you're allowed to take 2, how many did he take?
C: Three
E: And that's not ok?
C: No

(34) C: No, because he took three

E: So what's wrong with three? — What did Peter tell him?
C: Two
E: So is it ok?
C: No

Measure Phrase-Mass Noun Scenarios (*Filling the Pool*)

(35) C: No, because she didn't fill it up to the red line.
E: What did she do?
C: Filled all the way up.
E: Is that not what Brother ask?
C: No.

(36) C: No.
E: How come?
C: Cause she did it over.
E: She did? Where did Brother Bear say to fill the pool to?
C: To the red line all the way to there. (pointing to red line)
E: And where did she fill it to?
C: Almost to the top.

(37) C: No, because it's past the line.

(38) C: No, she did too high, it's past.

(39) C: No, because it's like up there. (pointing above line)

(40) C: No, because he skipped it. (pointing to line)
E: He skipped it?
C: Yeah.
E: He went too far?
C: Yeah too far.

Justifications of rejections in the <2 scenarios from CHILDREN in the UPPER BOUND condition

Numeral-Count Noun Scenario (*Books before Bedtime*)

(41) C: No, he read one.
E: Is what he did okay?
C: [shakes head no]
E: Why?
C: He only picked one.

E: How many was he suppose to pick?
C: Two!
C: No because he has one.
E: So reading one is not ok?
C: No.

Measure Phrase-Mass Noun Scenario (*Stuffing for Bears*)

(42) C: She did one. [presumably one inch, although not labeled in stimuli]
E: So is that ok?
C: No, because it's right here [points below line] not here [points to line].
E: That's not enough?
C: No.
No, because she didn't go up to this level.
She filled it up lower than the red line.

Justifications of rejections in the >2 scenarios from CHILDREN in the LOWER BOUND condition

Numeral-Count Noun Scenario (*Bunnies and Carrots*)

(43) C: No, he took three carrots
E: What did Peter say?
C: Take two.
C: No, because the bunny took three
E: Oh, three? Can you count them for me? [child counts three carrots]
E: And that's not ok?
C: No.

Numeral-Count Noun Scenario (*Stuffing for Bears*)

(44) C: No, She did a little taller than the red line
E: It's not ok to use more than two inches?
C: No
C: No, because she put too much
E: Would it be ok to put this much? [points below line]
C: That would be too little. You have to use this much. [points to line]

Justifications of rejections in the >2 scenarios from ADULTS in the LOWER BOUND condition

Numeral-Count Noun Scenario (*Bunnies and Carrots*)

(45) a. He took more carrots than offered to him.
b. He took an extra carrot without asking.
c. Benjamin picked up 3 carrots when he only needed 2.

Measure Phrase-Mass Noun Scenario (*Stuffing for Bears*)

(46) a. She passed the red line. The bear will have too much stuffing.
b. She filled the jar up with too much stuffing.
c. Wendy didn't use the right amount because she filled it up more than 2 inches.

Justifications of acceptances in the >2 scenarios from ADULTS in the LOWER BOUND condition

Numeral-Count Noun Scenarios (*Various items*)

(47) a. Because he [Benjamin Bunny] just asked that he eat [carrots], so as long as he took at least 2 he was okay.
b. The man in the yellow hat told him [Curious George] to read at least 2 books.
c. Although he [Benjamin Bunny] took more than what was recommended, he still took "enough" by Peter Rabbit's standards

Measure Phrase-Mass Noun Scenarios (*Various items*)

(48) a. Wendy used enough [stuffing] so her bear wouldn't be floppy.
b. Although too much [water in the pool], Brother Bear said 2 feet would be enough to splash in and she [Sister Bear] has two feet.

Appendix C: Sample Justifications from Experiment 2

*Justifications of acceptances in the >2 scenarios from CHILDREN in the revised LOWER BOUND (*hafta*) condition*

Numeral-Count Noun Scenario (*Vines*)

(49) a. Yes, because he swang more than three vines.
b. Yes, because he swinged on those.
c. Yes, he swung three vines..on four!

d. Yes, because he did all three vines.

Measure Phrase Scenario (*Monkey Bars*)

(50) a. Yes, because she can reach.
b. Yes, because she's much taller.
c. Yes, because she's more than three feet.

*Justifications of acceptances in the >3 scenarios from ADULTS in the revised LOWER BOUND (*hafta*) condition*

Numeral-Count Noun Scenario (*Vines*)

(51) a. He swung more than three.
b. Because he was able to exceed Tarzan's expectations
c. Marvin met more than required by Tarzan.
d. Not only did he swing on 3 vines, but swung on 4.
e. He swung on 3 vines plus a 4th.

Measure Phrase Scenario (*Monkey Bars*)

(52) a. Emily Elizabeth can reach high enough.
b. She can reach higher than 3 ft.
c. Emily Elizabeth meets the 3 ft condition.
d. She can because she could reach 4, when 3 was the minimum

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.

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.

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.

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. The grammatical view of scalar implicature and the relationship between semantics and pragmatics. 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.

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

Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary, and Michael K. Tanenhaus. 2010. "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116:42 – 55.

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.

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.

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.

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.

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.

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

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, 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.

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.

Kennedy
Department of Linguistics
University of Chicago
1010 E. 59th St.
Chicago, IL 60637
ck@uchicago.edu

Syrett
Department of Linguistics
Rutgers, The State University of New Jersey - New Brunswick
18 Seminary Place
New Brunswick, NJ 08901
kristen.syrett@rutgers.edu