Numerals denote degree quantifiers: Evidence from child language

Christopher Kennedy and Kristen Syrett
University of Chicago and Rutgers University

December 17, 2020

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 maximizing 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.

1 Numerals and 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 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. The first class of approaches, which we will refer to as

[1] 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*.

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. The basic idea is that the numeral in an example like (4) is ambiguous such that the sentence can be interpreted either as in (4a), giving lower-bounded truth conditions, or as in (4b), where ∃! means "there is a unique x," giving upper-bounded truth conditions.[2]

(4) Mookie got three hits.

a. $\exists x[\mathbf{hits}(x) \wedge \#(x) = 3]$

b. $\exists ! x[\mathbf{hits}(x) \wedge \#(x) = 3]$

On this view, the difference in meaning between (2) and (3) is just the difference between parsing the prejacent as in (4a) or as in (4b), respectively.

The second class of approaches is one in which bounding entailments at the sentential level do not reflect ambiguity or underspecification at the lexical level, but instead emerge from scopal interactions; we refer to such approaches as SCOPAL ANALYSES, of which there are two variants. One version of the scopal analysis comes from the so-called "grammatical" analysis of scalar implicature, which allows for a compositional implementation of the the classic analysis from Horn 1972: numerals introduce lower bounded truth conditions, and upper bounding implicatures are introduced by a silent, alternative-sensitive exhaustification operator in the syntax (see e.g. Chierchia 2006; Fox 2007; Chierchia, Fox, and Spector 2012; Spector 2013). For present purposes, we may assume a semantics for exhaustification as in (5), where $\subset$ is asymmetric entailment: exhaustification of ϕ gives back the conjunction of ϕ and the denial of all of its stronger alternatives.[3]

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

[2] To keep our logical representations as perspicuous as possible, we omit the relation contributed by the verb, since this plays no crucial role in distinguishing between the different analyses of numerals we are considering. A complete version of (4a) (and *mutatis mutandis*, the other examples considered in this paper) would look like (i).

(i) $\exists x[\mathbf{hits}(x) \wedge \#(x) = 3 \wedge \mathbf{got}(x)(\mathbf{m})]$

[3] The assumption that exclusion of alternatives is based on asymmetric entailment is a gross but hopefully benign oversimplification, for the purposes of illustrating this type of 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), but this does not bear on the issues discussed in this paper.

Assuming that the (relevant) alternatives for sentences involving numerals are variants that differ just in the numeral, composition of *exh* with (4), which is assumed to have only the lower-bounded interpretation in (4a), derives (6a), which can be equivalently and more perspicuously written as (6b).

(6) *exh* [Mookie got three hits]

a. $\exists x[\textbf{hits}(x) \wedge \#(x) = 3] \wedge \forall \psi[[\psi \in \{\exists x[\textbf{hits}(x) \wedge \#(x) = n] \mid n \in \mathbb{N}\} \wedge \psi \subset \exists x[\textbf{hits}(x) \wedge \#(x) = 3]] \rightarrow \neg \psi$

b. $\exists x[\textbf{hits}(x) \wedge \#(x) = 3] \wedge \neg \exists x[\textbf{hits}(x) \wedge \#(x) > 3]]$

In this approach, different interpretations of *Mookie has to get three hits* in (2) and (3) correspond to different attachment sites for *exh*: whether it takes scope over the modal or vice-versa. The lower bounded interpretation in (2) is derived when *exh* is inserted above the modal, as in (7); the upper bounded interpretation in (3) is derived when it is inserted below the modal, in the embedded clause, as in (8).

(7) a. *exh* [has to [Mookie get three hits]]

b. $\Box \exists x[\textbf{hits}(x) \wedge \#(x) = 3] \wedge \neg \Box \exists x[\textbf{hits}(x) \wedge \#(x) > 3]$

(8) a. has to [*exh* [Mookie get three hits]]

b. $\Box[\exists x[\textbf{hits}(x) \wedge \#(x) = 3] \wedge \neg \exists x[\textbf{hits}(x) \wedge \#(x) > 3]]]$

A second version of the scopal analysis treats numerals as scope-taking expressions in their own right, specifically as generalized quantifiers over degrees, of which numbers are a special case (Kennedy 2013, 2015; Buccola and Spector 2016; cf. Frege 1980 [1884], Scharten 1997, von Stechow's (1984, p. 56) treatment of measure phrases, and Solt's (2015) treatment of the quantity terms *much* and *many*). The numeral *three*, for example, has the denotation in (9): it is true of a property of degrees if the maximal degree that satisfies it is the number three.

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

As shown by (10), this denotation delivers two-sided truth conditions as the default interpretation of simple sentences like *Mookie got three hits*, but a lower-bounded interpretation can be derived by lowering the quantificational meaning of the numeral to a singular term (number-denoting) meaning by the application of Partee's (1987) BE and IOTA type-shifts (Kennedy 2015).

(10) a. three [Mookie got t hits]

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

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 (11), and the upper bounded interpretation in (3) arises when the numeral takes scope below the modal, as in (12).[4]

[4](11b) 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. (11b) thus places a lower bound of three on the number of hits that we find in each world that satisfies the modal claim.

(11) a. three [has to [Mookie get t hits]]

b. $max\{n \mid \Box[\exists x[\mathbf{hits}(x) < \wedge \#(x) = n]]\} = 3$

(12) a. has to [three [Mookie get t hits]]

b. $\Box[max\{n \mid \exists x[\mathbf{hits}(x) \wedge \#(x) = n]\} = 3]$

All other things being equal, the exhaustification and degree quantifier analyses derive identical truth conditions for structures like (13a-b), where Num denotes the number n and is the basis for calculation of the alternatives to S in the former, and where Num denotes the degree quantifier $\lambda P.max\{m \mid P(m)\} = n$ in the latter.

(13) 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: a semantics for numerals that delivers lower-bounded content, exhaustification, and calculation of alternatives, with the latter two corresponding to the quantity implicature system. In the degree quantifier analysis, all of these pieces are effectively built into the denotation of the numeral.

Our goal in this paper is to examine a set of data that allows us to draw a distinction between these three approaches — local analyses, an exhaustification-based scopal analysis, and the degree quantifier analysis — and, we claim, argues in favor of the latter. The crucial facts involve sentences in which a numeral is embedded in the complement of an existential modal, such as (14).

(14) 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 (14) 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 degree quantifier analysis, the upper bound reading of (14) is simply a consequence of the lexical and compositional semantics of the expressions in the sentence: it corresponds to a parse in which the numeral takes scope over the modal:

(15) a. three [can [Mookie make t errors]]

b. $max\{n \mid \Diamond[\exists x[\mathbf{errors}(x) \wedge \#(x) = n]]\} = 3$

(15b) 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 (16b), which require merely that there is a world in the modal domain in which Mookie makes exactly three errors.

(16) a. can [three [Mookie make t errors]]

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

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 exhaustification analysis, the two readings likewise correspond to a scopal interaction, in this case between *exh* and the modal, with wide scope of *exh* in (17) delivering upper bounds:

(17) a. *exh* [can [Mookie make three errors]]

b. $\Diamond\exists x[\textbf{hits}(x) \wedge \#(x) = 3] \wedge \neg\Diamond\exists x[\textbf{hits}(x) \wedge \#(x) > 3]$

(18) a. can [*exh* [Mookie make three errors]]

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

The difference between this analysis and the degree quantifier analysis is that it does not rely only on lexical and compositional semantics, but also on alternative calculation: if, for some reason, the alternatives to the prejacent were not computed, the upper-bounded truth conditions in (17b) would not be derived.

Finally, on the local analysis, there are two possible interpretations of (14), shown in (19a-b).

(19) a. $\Diamond[\exists x[\textbf{errors}(x) \wedge \#(x) = 3]]$

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

Neither (19a) nor (19b) entails an upper bound. (19b) is logically equivalent to (16), and has the same weak truth conditions. (19a) 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 (19a) can be strengthened to an upper bounded interpretation via reasoning involving the Maxim of Quantity: (19a) (but not (19b)) 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.

The crucial difference between the three approaches to upper bounding inferences, then, is that only the degree quantifier analysis derives an upper bounded interpretation for (14) strictly in terms of the lexical and compositional semantics of the expressions involved: both the exhaustification and local analyses must *also* invoke whatever mechanisms are involved in the calculation of quantity implicatures, which are grammatical in the former, and (we assume) fully pragmatic in the latter.[5] The analyses therefore make different predictions about how sentences like (14) will be evaluated if the quantity implicature mechanisms are suppressed or otherwise inactive: the degree quantifier analysis predicts that upper-bounding inferences will be retained, all other things being equal; the other two analyses predict that they will disappear. We can therefore distinguish between the two approaches

[5] We take it that the semantic difference postulated by the local analysis presumes a fully pragmatic theory of implicature calculation, since there would be no reason for the language learner to posit an upper-bounded semantics for numerals if upper-bounds could be independently derived from exhaustification of the lower-bounded semantics.

by examining how sentences like (14) are understood by a population that has competence with quantification but has difficulty with quantity implicatures. In the next section, we describe an experiment involving one such population.

2 Experiment: Upper bounds and existential modals 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 an 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).

The asymmetry between upper bounded interpretations of numerals vs. other scalar terms in child language is now viewed by many as a central argument for a fully semantic account of the former. If this is right, then we already have reason to reject an exhaustification-based account to upper bounding inferences, since it relies on the implicature system — though Barner and Bachrach (2010) propose a different interpretation of the asymmetry that is consistent with such an account. We return to Barner and Bachrach's proposal in section 3.1 below; in this section, we describe an experiment designed to test the predictions of the local and degree quantifier analyses of numerals which assumes that children's difficulty with implicature extends to all scalar terms, including numerals.

The inspiration for our experiment comes from a study by Musolino (2004). In Musolino's experiment, four- and five-year old children were exposed to scenarios in which a character had to perform an action with multiple objects, and were told that the character would win a prize under certain conditions. These conditions were described using sentences with the numeral *two*, including sentences in which *two* was embedded under an existential modal, as in (20).

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

Musolino was interested specifically in whether children correctly understand (20) as not imposing a *lower* bound, and he found that indeed, in scenarios in which the Troll missed one hoop, children who were told (20) said that it should get the prize more than 80% of the time. Musolino did not test for children's judgments about sentences like (20) in scenarios in which the *upper* bound was exceeded, however (e.g., scenarios in which the Troll missed three or more hoops); our experiment was designed to introduce this condition, since it is in precisely such scenarios that the predictions of the three semantic analyses of numerals discussed in the previous section come apart. The local analysis (as well as the exhaustification analysis) can only derive an upper-bounded interpretation of sentences like (20) values by implicature, and so predicts that children (of the relevant age) should fail to reject such sentences as descriptions of scenarios in which the upper bound is surpassed. In contrast, the degree quantifier analysis derives an upper-bounded interpretation without invoking the implicature system, by scoping the numeral above the modal, and so predicts that children should reject such sentences as descriptions of scenarios in which the upper bound is surpassed.[6]

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

[6] Our experimental stimuli, as well as the results of a separate experiment demonstrating that children and adults assign the same range of interpretations to sentences in which numerals are embedded under universal root modals, are available at https://semanticsarchive.net/Archive/2E3Y2FjO.

"yes" or "no" (verbally in the case of children, or on a response sheet in the case of adults) and occasionally provide a justification.

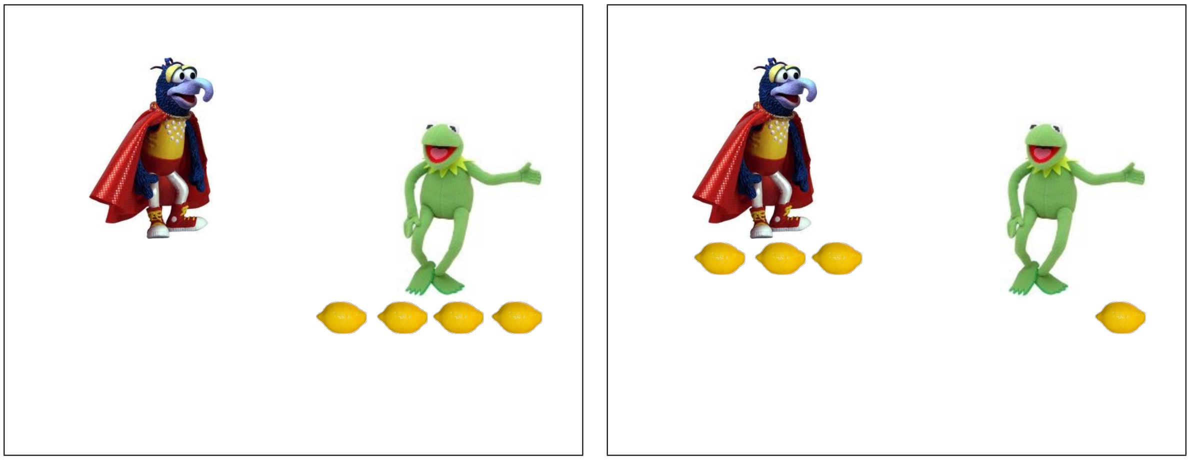
The experimental session began with two training items. The test session that followed consisted of six test items and four control items, all with the same basic structure. Each test item involved instructions from the second character that featured an existential modal and a numerical expression. Half of these involved numerals plus plural count nouns (e.g., *two books/carrots/lemons*), and half involved measure phrases plus mass nouns in a pseudopartitive constructions (e.g., *two feet of water*); there was no significant difference between the two types of nominals.

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 — what kind of bounding implications they took the sentence to convey. 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.

For example, in one scenario, 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.) Examples of beginning and ending scenes for this item are presented in Figure 1. 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 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.

The results are presented in Table 1. A McNemar's test 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,



(a) Beginning scene (b) Ending scene

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

	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

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).

To provide additional confirmation that the initial numbers reflected the influence of a faulty item, we also ran a follow-up experiment on children that replicated all items in the original experiment except for the "stuffing the bear" item, replacing it with a new >2 mass

	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, "Stuffing the Bear" items removed

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*." In the >2 condition, the second character ends up filling the stuffing past the two inch line on the container.

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 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 our initial results 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*.

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

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

Before turning to the discussion, we wish to note a second noteworthy 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, they are significantly less likely to accept the <2 items than the =2 items (main experiment with faulty item removed: U_A=176, z=3.44, p=.0006, two-tailed; follow-up experiment: W=-66, z=-2.91, p=.004, two-tailed), while adults are not (main experiment with item removed: U_A=288, z=1.58, p=.114, two-tailed). Overall, children are less likely than adults to accept the <2 items (U_A=372, z=-1.72, p=.09, two-tailed, marginally significant). We will have more to say about this result in section 3.2.[7]

[7] Syrett, Austin, and Sanchez 2020 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.

Discussion Once we corrected for the item effect, the results of our experiment 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 assign upper bounded 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 from the lexical and compositional semantics by assigning the sentence a parse in which the numeral takes scope over the modal, but it is unexpected on the exhaustification and local analyses, in which such readings can only be derived using the quantity implicature system. In the next section, we consider and reject three ways that a proponent of the exhaustification or local analyses of upper-bounded interpretations of numerals could challenge the conclusion that our experimental results argue in favor of the degree quantifier analysis.

3 Alternative interpretations of our experimental results

3.1 Scale salience

The first challenge to our interpretation of our experimental results 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 should not be taken to indicate that numerals are semantically upper bounded, but rather that children have no trouble calculating scalar implicatures for numerals. And indeed, Barner and Bachrach (2010) have argued precisely this. Specifically, Barner and Bachrach (B&B) propose 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. If this is correct, then children's behavior in our experiment may simply reflect the implicature system at work, and so does not argue in favor of the degree quantifier analysis.

When we step back and look at the larger picture, we believe that there are reasons to resist this account of the child language data. The most compelling reason, in our view, is that this account 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.[8]

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 are "pre-compiled" in the lexical semantics of the numeral, so to speak. The most theoretically parsimonious account of the child language pattern, then, is one in which children acquire this very same meaning.

B&B also 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.

But the degree quantifier meaning is fully compatible with B&B'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 (21).

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

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,

[8] 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 (Kennedy 2015), 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.

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 initially 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. It is also 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. 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 (21)) 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.[9]

3.2 "Playing it safe"

The second challenge 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 our experiment. 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 one of the interpretations 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 "plays it safe" 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

[9] 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.

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 unmodified numerals.

This, we believe, provides an argument against the "play it safe" account of upper-bounding inferences suggested above. 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 that he is allowed to use no lemons. In fact, none of the analyses under consideration derive these inferences as entailments, which means that they must be pragmatic in nature. 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 inferences are computed is not a question that we can answer here; 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 degree quantifier analysis introduces this asymmetry; the exhaustification and local analyses do not.

3.3 Contextual factors

The third challenge to our conclusion 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 made such implicatures available. The reasoning goes like this:

(i) In the experiment, it was made contextually salient that there are constraints on what is allowed, specifically that there were limits (not made explicit) on the quantity of 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 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 (without e.g. also making salient scalar alternatives accessible) can improve implicature calculation in child language.

That said, in order to make a quick test of whether our results crucially depended on the contextual setup of the experiment, we added a single experimental item to the end of an independent experiment, 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 4, where percentages indicate the number of *"yes"* answers to the experimenter's question. 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 remain relatively unlikely to accept lower values, indicating the same reluctance to draw free choice inferences that we saw in our experiment.

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

Table 4: Rejection of higher values in the absence of contextually salient constraints on what is allowed

4 Conclusion

We began this paper by observing that sentences in which logical operators compose with propositions that involve upper-bounded interpretations of numerals are a problem for the traditional neo-Gricean analysis of upper-bounding inferences in numerals, and considered three different semantic analyses of numerals that eliminate this problem: "local analyses," in which numerals introduce two-sided content via composition with their nominal arguments, and two scopal-analyses, one based on exhaustification and a second based on maximizing degree quantification. We then observed that when a numeral is embedded under an existential root modal, only the degree quantifier analysis can derive upper-bounded interpretations without invoking the mechanisms involved in scalar implicature calculation. This led us to examine how children interpet such sentences at an age at which they are known to have problems with scalar implicature, and we found that they are adult-like in their ability to interpret such sentences as entailing upper-bounds, a result that we believe to favor the degree quantification analysis of numerals, even after consideration of alternative interpretations of the child language data.

A question that remains, if this conclusion is correct, is *why* numerals take on meanings as maximizing degree quantifiers. This question is particularly salient if we also accept the two key claims of the so-called "grammatical" theory of scalar implicature (both of which have plausible justification), namely that natural language includes a compositional mechanism for exhaustification (*exh* or the equivalent) and a mechanism for the generation of alternatives. As we have seen, these mechanisms together can derive exactly the same kinds of truth conditions for sentences with numerals as maximizing degree quantification, in basically the same way (i.e., through scopal interactions); the difference is that the meaning components are "distributed" in the exhaustification account and "pre-compiled" in the degree quantifier account. We have argued that the child language data speaks in favor analyzing numerals as degree quantifiers, and this semantic analysis receives further support from the studies of adult language that we cited in Section 1, such as the robust preservation of upper-bounded interpretations of numerals under negation and in other downward entailing contexts. Evidently there is something special about numerals that leads to conventionalization of maximization.

Acknowledgments

We are very happy to be able to contribute this paper to a volume in honor of Stephanie Solt, whose work has engaged and inspired us over the years in ways that know no upper bounds, and we send her big congratulations on the acquisition of her very own number. We are grateful to several anonymous reviewers for constructive feedback on this work, to the research assistants in the Laboratory for Developmental Language Studies at Rutgers University, and to audiences at the Boston University Conference on Language Development, the Rutgers Center for Cognitive Science, The Ohio State University, and the Hebrew University of Jerusalem.

References

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.

Buccola, Brian, and Benjamin Spector. 2016. Modified numerals and maximality. *Linguistics and Philosophy* 39:151–199.

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 *Proceedings 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, MA: MIT 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.

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.

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. 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.

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.

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. *Cognition* 86:253–282.

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

Partee, Barbara. 1987. Noun phrase interpretation and type shifting principles. In *Studies in Discourse Representation and the Theory of Generalized Quantifiers*, ed. Geroen Groenendijk, Dik de Jongh, and Martin Stokhof, 115–143. Dordrecht, Reidel, Netherlands: Foris.

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.

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.

Solt, Stephanie. 2015. Q-adjectives and the semantics of quantity. *Journal of Semantics* 32:221–273.

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, Jennifer Austin, and Liliana Sanchez. 2020. Establishing upper bounds in english monolingual and heritage spanish-english bilingual language development. *Language Acquisition* .

Syrett, Kristen, Julien Musolino, and Rochel Gelman. 2012. How can syntax support number word acquisition? *Language Learning and Development* 8:146–176.

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.