**RESEARCH**

# Numerals under negation: Empirical findings

## Stephanie Solt[1] and Brandon Waldon[2]

[1] Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, DE

[2] Stanford University, Department of Linguistics, Stanford, CA, US

Corresponding author: Stephanie Solt (solt@leibniz-zas.de)

Despite a vast literature on the semantics and pragmatics of cardinal numerals, it has gone largely unnoticed that they exhibit a variety of polarity sensitivity, in that they require contextual support to occur felicitously in the scope of sentential negation. We present the results of a corpus analysis and two experiments that demonstrate that negated cardinals are acceptable when the negated value has been asserted or otherwise explicitly mentioned in the preceding discourse context, but unacceptable when such a value is neither mentioned nor inferable from that context. In this, bare cardinals exhibit both similarities to and differences from other types of numerical expressions. We propose an account of our findings based on the notion of convexity of linguistic meanings (Gärdenfors 2004) and discuss the implications for the semantics of numerical expressions more generally.

## 1 The puzzle of numerals under negation

Cardinal numerals have been an active topic of research in semantics and pragmatics for nearly half a century. In this paper, we investigate a pattern that has gone largely unnoticed in this time, namely that unmodified cardinals exhibit a variety of polarity sensitivity.

Much of the work in this area has centered around competing analyses of sentences such as (1):

(1)     Lisa has 40 sheep.

There is little disagreement that in a neutral context, the utterance of (1) would typically communicate that Lisa has exactly forty sheep; that is, it would convey both a lower bound and an upper bound on the number of sheep that she has. Where theories differ is in how this interpretation arises. On one approach, which might be called the one-sided analysis (Horn 1972; Gazdar 1979; Levinson 1983; Horn 1989; van Rooij & Schulz 2004), only the lower bound is contributed by the lexical semantics of the number word; the upper bound is derived via a scalar implicature of the same sort that is responsible for the inference from an utterance of *some* that 'not all' obtains. Two-sided analyses, on the other hand (Sadock 1984; Koenig 1991; Scharten 1997; Geurts 2006; Breheny 2008; Spector 2013; Kennedy 2013; 2015), claim that the 'exactly' interpretation of the numeral in (1) follows from a semantically encoded – rather than pragmatically derived – upper bound.

Going back to at least the early work of Horn (1972), the interaction of cardinal numeral expressions with negation has figured prominently in attempts to resolve this debate. Central to Horn's argumentation for a one-sided analysis is the observation, attributed to Jespersen (1933), that a numerical expression in the scope of sentential negation typically

has a 'less than' interpretation; thus when presented out of the blue, (2) tends to convey that Lisa has fewer than 40 sheep:

(2)      Lisa doesn't have 40 sheep.

This is expected on the one-sided analysis, in that scalar implicatures typically fail to arise in the scope of negation and other downward entailing environments.

Other evidence from negation, however, has been put forward in favor of the two-sided analysis of number-word meaning (see Horn 1992; 1996; Scharten 1997; Breheny 2008; Spector 2013; Kennedy 2013; 2015). With the right context and intonation, the upper bound can also be negated, as evidenced by the acceptability of (3b), and its parallel feel to (3a):

(3)      Do you have three children?
         a.   No, I don't have three children. I have two.
         b.   No, I don't have three children. I have four.

This is problematic on the one-sided account: (3b) should be contradictory if we understand the negative statement as 'it is not the case that I have at least three children'.

It has also been observed that in the scope of negation, numerals behave differently from other scalar expressions that receive an upper bound through pragmatic processes, such as *many* in the example below:

(4)      a.  ??Neither of them read many of the articles on the syllabus. Kim read one and
             Lee read them all.
         b.   Neither of them read three of the articles on the syllabus. Kim read two and
             Lee read four.

Such divergences – bolstered by findings from language processing (Huang & Snedeker 2009; Panizza et al. 2012; Marty et al. 2013) and acquisition (Huang et al. 2013) which show that the upper-bounded interpretation of numerals is more accessible than that of scalar items such as *some* – have been taken as evidence that the two-sided 'exact' interpretation of numerals is lexically encoded rather than pragmatically derived. For Breheny (2008) in particular, the exact interpretation is the only one made available by the semantics; but more commonly, numerals are proposed to be in some way ambiguous between 'exact' and 'at least' interpretations (see Geurts 2006; Spector 2013; Kennedy 2015).

In this and other work, it has been observed that the context in which a numerical expression occurs has an impact on the interpretation it may receive. In particular, examples cited to demonstrate the availability of the 'at least' reading often involve a context in which the numerical value has been previously mentioned or is otherwise salient. An oft-repeated example originating in Gazdar (1979) is that *John has 3 children* readily allows an 'at least' interpretation when three represents some sort of threshold in the context, here perhaps the minimum number of children needed to qualify for government benefits. The role of context is pursued most extensively by Scharten (1997), who argues that the crucial factor is information structure. Specifically, when a numerical expression occurs in comment position, the paradigm case being when it serves as the answer to an explicit or implicit *how many* question, it necessarily receives an exact interpretation; but when it occurs in non-comment position, it may get the 'at least' interpretation.

Yet despite the extensive research in this area, a pattern that has not to our knowledge been explicitly discussed is that in the absence of a context that makes the numerical

value salient, a negated number word is simply infelicitous.[1] By way of example, a wide variety of numerical expressions may serve as appropriate answers to a *how many* question (5a). A simple negated numeral, however, is infelicitous (5b):

(5)     How many sheep does Lisa have?
   a. She has 40 / about 40 / more than 40 / fewer than 40 / at least 40 / at most 40 / between 40 and 50 sheep.
   b. ??She doesn't have 40 sheep.

On the other hand, when the context is such that the numerical value is salient, the very same sentence becomes unobjectionable:

(6)     Context: Farmers with 40 or more sheep qualify for a new government subsidy program.
   Q: Can Lisa apply for the new program?
   A: No. She doesn't have 40 sheep.

In (6), the negated numeral has a one-sided 'at least' interpretation: Lisa cannot apply for the subsidy program because she has fewer than 40 sheep. It is also possible, if somewhat more difficult, to create a context in which a numeral is felicitously negated on the two-sided interpretation. For example:

(7)     Context: The conversational participants know that Fred has 40 sheep.
   Q: Does Lisa have the same number of sheep as Fred?
   A: No. She doesn't have 40 sheep – she has 75.

But when a context such as those in (6) or (7) is lacking, the negated cardinal numeral is decidedly odd. As one way to characterize this pattern, we may say that numerals exhibit a contextually dependent variety of polarity sensitivity.

 From one perspective, the infelicity exemplified in (5b) is not entirely surprising, for two reasons. First, certain modified numerical expressions have been observed to pattern as positive polarity items (PPIs), including the superlative modified numerals *at least n* and *at most n* (Geurts & Nouwen 2007; Cohen & Krifka 2014; Spector 2014) as well as approximative constructions such as *approximately n* and *about n* (Rodríguez 2008; Spector 2014; Solt 2018).

(8)     a. Lisa has at least 40 sheep.
   b. *Lisa doesn't have at least 40 sheep.

(9)     a. Lisa has about / roughly / approximately 40 sheep.
   b. *Lisa doesn't have about / roughly / approximately 40 sheep.

Here too, one has the feeling that the (b) examples would be improved if the numerical value were salient in the context. Perhaps bare numerals should in some way be aligned to this class.

 Secondly, it has long been recognized that negative utterances more generally tend to be odd discourse initially and in neutral contexts, but instead require a context in which the

---

[1] A reviewer notes that it is his/her impression that researchers working in this area are generally aware of the infelicity of numerals under negation. We share this impression, but do not know of any work that discusses it explicitly.

"positive counterpart" has been previously asserted or implied, or is at least in some way under consideration (see Horn 1989 for extensive discussion and references). For example, (10) (based on Ducrot 1973) would be strange if it had not been earlier claimed that Pierre was Marie's cousin; similarly, (11) (from Givón 1978) would be odd if the possibility of the speaker's wife being pregnant had not in any way been raised.

(10)    Pierre isn't Marie's cousin.

(11)    Oh, my wife's not pregnant.

One might then suspect that the infelicity of (5b) – and the contrast to (6) and (7) – is simply another instance of this more general phenomenon.

   This cannot however be the whole story. The reason is that not all numerical expressions are infelicitous under negation in a neutral context, the prime counterexample being comparatively modified numerals:

(12)    How many sheep does Lisa have?
        a.    She doesn't have more / fewer than 40.
        b.    She has no more / fewer than 40.

Thus the unacceptability of (5b) in contrast to the acceptability of the examples in (12) is a fact in need of explanation. Furthermore, because the numerical domain offers such clear examples of expressions that do and do not require contextual support to be felicitously negated, the investigation of these data has the potential to shed light on the discourse constraints on negated utterances more generally.

   The objectives of this paper are twofold. Our first goal is an empirical one: we seek to establish more clearly the facts regarding the acceptability of numerical expressions in the scope of negation. Just how bad are negated bare numerals in neutral contexts? How does this compare to the previously documented PPI status of modified numeral constructions such as *at least n* and *about n*? To what extent does the felicity of negated numerical expressions of different sorts improve in a supportive context in which the numerical value is in some way salient? And what specifically is required of the context? Must the number have been asserted or otherwise mentioned? Or is it sufficient that it be implied, or merely part of the background knowledge of the conversational participants?

   In pursuing answers to these questions, corpus-based methods and especially controlled experimentation have much to offer. Because the acceptability of numerical expressions is dependent on the context, it is difficult to establish the relevant facts via intuition-based approaches alone, because it is all too easy to rescue an otherwise infelicitous example by inferring the appropriate discourse context. Particularly challenging is making comparisons between different sorts of numerical expressions (e.g. bare vs. modified numerals) or different types of discourse contexts. We address this via corpus data illustrating typical uses of numerals under negation, as well as experiments in which both the numerical expression and the discourse context are systematically varied.

   Our second goal is a theoretical one, namely to provide an explanation for the infelicity of negated bare numerals in a neutral context. Previewing our theoretical proposal, we will argue that what goes wrong with an example such as (5b) is that when the numeral takes on its two-sided exact interpretation, its negation specifies a disjoint region on the number line. That is, *not $40_{EXACT}$* specifies values either above or below 40. We will analyze this effect as deriving from a constraint on the assertion of numerical expressions which holds that they must specify a convex region in the space of answers to the current

question under discussion (QUD). The relevance of convexity as a constraint on the meaning of content words was famously established by Gärdenfors (2004); our proposal adds to other recent work (especially Chemla et al. 2019) demonstrating a role for it beyond this domain. As will be argued in Section 5, our proposal accounts not only for the contrast between (5a) and (5b), but also for the role of a supportive context, which we will argue is to change the QUD. Indirectly, our investigation also yields insight into the long-standing debate over one-sided versus two-sided readings of number words.

The organization of the paper is as follows. Section 2 presents corpus data illustrating the types of contexts in which negated numerals are attested. Sections 3 and 4 present the results of two online acceptability judgment studies. Finally, Section 5 develops our theoretical proposal and discusses its broader applicability, and Section 6 concludes.

## 2 Corpus study

As a first step in checking the intuitions discussed above, we collected naturally occurring tokens of bare and modified numerals in the scope of negation, using as a source the Corpus of Contemporary American English (COCA; Davies 2008–).

A limitation of this approach is that the constructions of interest do not lend themselves readily to identification via an automated corpus search. Using a search string of the form "not/n't (modifier) [mc*]" (where [mc*] is the COCA tag for a cardinal numeral) yields some relevant tokens but also a high proportion of irrelevant ones (e.g. *This is really about two families, not about two casinos*). It also fails to capture cases where the negator is separated from the numerical expression (e.g. *They do not have 60 votes in the Senate*). Broadening the search to also capture examples such as these yields a wider variety of good tokens but an even higher proportion of irrelevant ones. This precludes the possibility of reliable quantitative analysis of the frequency of negated numerals or their subtypes. Instead, we opted for a qualitative approach, on which a variety of narrower and broader search strings were utilized to generate possible tokens of negated numerals, and these results were manually reviewed to identify relevant examples. Our goal was thus not to measure the frequency at which numerical expressions occur under negation, but rather to shed light on the sorts of discourse contexts in which such examples are attested.

We begin with negated bare numerals. We observe first that our search strategies yielded many tokens that were very different in character from the examples discussed in Section 1, including: cases in which the numeral is interpreted as taking scope over negation; the negation of *one* to mean 'no' (e.g. *We could not find one clear piece of evidence*); and negated numerals in the complement position of verbs with inherently comparative meanings (e.g. *The entire planting did not exceed 5,000 bushels*), which might be aligned to comparatively modified numerals. Putting such cases aside as not directly relevant, and focusing on those in which a plural cardinality is negated, we find four broad categories of examples:

**i. Denial of assertion** It has been proposed that the prototypical use of negation is denial (Tottie 1991). It is thus not surprising that in some of the examples of negated numerals we found, the negated expression is used to explicitly deny an earlier (positive) assertion in the preceding discourse or the broader context of utterance. The following examples illustrate this: (13) is a denial of a widely publicized claim by Donald Trump that his opponent received three million illegal votes; in (14), there is there is a prior assertion that the truck driver had been convicted of six crimes, which is denied in the passage.

(13)     Contrary to Trump's world of make believe, there **weren't 3 million illegal Hillary Clinton voters**.

(14)     [A] truck driver for the city of Chicago got his job despite admission that he had been convicted of one burglary and five thefts in the past, even though the city had an unofficial policy of not hiring ex-cons. […] Then city officials found out that Felski **didn't have six convictions**, he actually had 22 convictions, and he was fired.

The following example similarly expresses denial of a prior claim, here via constituent rather than sentential negation:

(15)     Eyewitnesses who knew Rohrbough before the shooting – not from subsequent media reports – insist he went down with the first gunfire from the stairs outside the school cafeteria. […] O'Shea describes firing his weapon 60 times, **not 51** as ballistics reports claim.

Note that in both of the previous two examples, the actual value reported is higher than the negated value, meaning that negation must target the two-sided 'exactly' reading of the number word.

   **ii. Explicit contrast/threshold** In some attested examples, the numeral is introduced into the discourse not as part of an assertion that is later denied, but rather associated with some state of affairs to which the speaker/writer intends to make a contrast via the negated numeral. Thus in (16), a contrast is made between the number of potential terrorists and the number of people on the watchlist, whereas (17) expresses a contrast between sales of Fumento's book and the typical sales of books promoted on the Donahue show.

(16)     [Y]ou know, there's, what, 800,000 people on the watch list. Well, there **aren't 800,000 potential terrorists** in America.

(17)     When Donahue does that with your book, you could sell 20,000 to 50,000 additional books in the next weeks. But Fumento's book **didn't sell 50,000 or even 20,000 copies**. In fact, it sold about 12,000.

In (18) the numeral occurs only once in the passage, but nonetheless an explicit contrast is established between Glavine's performance and that of Maddux.

(18)     Perhaps a downside to being part of a talented trio is that at least one member will be overlooked. Among the Braves' Big Three, that most often was Glavine. He **didn't win four consecutive Cy Young awards** like Maddux, and he didn't dominate in the postseason like Smoltz.

A related discourse type involves reference to some contextually relevant numerical threshold. In (19), for example, there is explicit mention of ten years of service as the (minimum) requirement for retirement:

(19)     But many lawmakers could not collect because they, like other state workers, needed 10 years of service to retire. "A lot of legislators in the past **didn't serve 10 years** and weren't eligible for pension," says Morris, the Kansas lawmaker.

**iii. Implicit contrast/threshold** Compare the above examples to the following, in which only one state of affairs is overtly mentioned in the immediate discourse context.

(20)    When Withee made bean collecting into a full-time hobby, he started a bean catalog that resulted in correspondence. He traded beans like collectors trade stamps or baseball cards. "There **weren't 1,200 varieties** of beans back at the time of Christ in this country, there were just a few," he says.

(21)    That's what the Clintons do, and they're very good at it. I mean, that's why **there's not 17 people** running for the Democratic nomination.

Here it is left to the reader to infer what the point of comparison is. In (20) it is implied (though not explicitly stated) that there are now 1,200 varieties of beans, while (21) suggests a contrast to the number of candidates for the Republican nomination.

   **iv. Negation of minimum significant value** In a final type of example, the numeral that is negated appears to represent some minimum value that would count as significant in the given context. The numeral does not correspond to a previous assertion or to some specific threshold or contrastive state of affairs; rather, in these contexts, the negated numeral communicates that the real value is nonzero but low. These might be thought of as 'not even' contexts: the insertion of *even* before the numeral can highlight this aforementioned communicative effect, as in the example below ('doesn't [EVEN] have 10,000 customers').

(22)    The company's goal was to bring financial planning to the masses for what is now a $299 upfront fee plus a $19 monthly subscription. Yet even with nearly $75 million in venture capital money to play with, it **doesn't have 10,000 customers** signed up for its standard plan.

We also note that there are a range of related contexts which make reference to the spatial or temporal domains, these typically involving constituent negation:

(23)    You said you were going to show those to us and you got up and walked to the door and then said, oh, that's right, they're not here. But **not fifteen minutes** before we got here, you told my producer they weren't here. You already knew they weren't here when you got up to get them.

(24)    I turned off the highway and drove a twisting road that finally dropped down to the lake. It wasn't a real lake. The Corps of Engineers had dammed up the Tallahatchie River, and now the town of Como had a lake **not five miles** from the city limits.

We discuss the 'negation of minimum significant value' contexts further in Section 5. For now, we will make the brief comment that there appear to be additional discourse factors governing such examples. Why is 10,000 a significant value in the context of financial planning service customers, or five miles a significant distance from the city limits? This use of negated numerals appears to rely on world knowledge in ways the other context types do not.

   Overall, our corpus investigation suggests that the frequency of negated bare numerals is relatively low in comparison to the frequency of numerical expressions as a whole. In particular, we found few if any negated examples that were not licensed by the discourse or the broader context in one of the ways outlined above.

   We turn now to modified numeral constructions that have been characterized as positive polarity items (see Section 1). With regards to numerals modified by approximators such as *about*, *roughly* and *approximately*, the most common sort of negated example that we find involves cases where they form part of comparative quantifiers, as in (25). Such examples

are discussed in Solt (2014; 2018), who observes that when embedded in comparative quantifiers, approximators shift from positive polarity items to negative polarity items.

(25)     Miami's condo bubble has burst, new home building in south Florida has virtually ground to a halt, and contractors who once cruised 184th St. looking for labor are left seeking work themselves. "Now you **don't see more than about 20 workers** waiting in the mornings," says Ms. Echeverria.

Putting aside such examples, we find the occurrence of approximator-modified numerals under negation to be extremely infrequent. One of the very few such tokens of this sort found is the following, which falls into the 'denial of assertion' category discussed above.

(26)     ELLIOTT: […] This gentleman here. He's 6′7″. How many – how much
         do you weigh?
         GREG: About 140 kilos.
         ELLIOTT: That's about – What? – 300 pounds?
         GREG: **Not about 300**–270.

Similarly, superlative-modified numerals of the form *at least n* and *at most n* were found only rarely in the scope of negation. In many apparent examples of this, the modified numeral scopes covertly over negation, as in (27), the salient interpretation of which is that there were at least nine items that Congress had not acted on. The remaining examples typically involved the negated construction occurring in the scope of another downward entailing operator, for example in the antecedent of a conditional, as in (28); such a configuration has been observed to rescue PPIs in the immediate scope of negation (Spector 2014).

(27)     We are way past the budget deadline, but Congress still **has not acted on at least nine** of the 13 budget items.

(28)     **If you do not have at least 200 mcg of selenium** in your multivitamin, make a trip to the health food store and invest $15 now.

In summary, the results of our corpus study support the initial observation that negated bare numerals require a context in which the numerical value is in some way made salient. They also let us see more clearly that a range of different context types may be sufficient to achieve this: not just one in which the value has been directly asserted, but also ones where it has been introduced as a threshold or point of comparison, or even merely implied as such. We cannot however rule out that other types of uses are also acceptable but simply too infrequently occurring to have been turned up by our search strategies. Our data are also consistent with previous claims that approximator- and superlative-modified numerals are positive polarity items; but here in particular, the data are too sparse to allow us to assess whether these expressions are also sensitive to context. In the next stage of our research we therefore turn to experimental methods to substantiate these findings quantitatively, using the corpus data as a starting point for creating experimental materials.

## 3  Experiment 1

In our first experiment, we assess the acceptability of negated numerical expressions in a range of discourse contexts based on the categories identified in the corpus study reported in Section 2. More specifically, we investigate bare numerals in these contexts, comparing

them to one of the previously described PPI numerical constructions, namely numerals modified by the approximator *about*.

We hypothesize first of all that the more salient the numerical value is in the discourse, the more acceptable the negated numeral will be. Regarding the comparison between bare and approximator-modified numerals, we contrast two possibilities. If the polarity sensitivity of bare numerals is an instance of the same phenomenon characterizing their approximator-modified counterparts, then we would expect that once context is held constant, the acceptability of the two should be equal. If on the other hand the pattern observed to characterize bare numerals has a different nature or source, then we predict differences in their acceptability in some or all discourse contexts.

### 3.1 Participants

A total of 80 workers were recruited via the online workforce marketplace Amazon Mechanical Turk (MTurk). The only inclusion criteria were an acceptance rate of over 95% on prior MTurk "human intelligence tasks" (HITs) and a U.S. internet protocol (IP) address. MTurk workers were paid $1.20 for participation.

### 3.2 Materials

Stimulus items had the form of four-sentence texts, which were modified versions of naturally occurring examples sourced via COCA, or constructed to include typical features of such examples. In each text, the third sentence, which was presented in boldface, contained a negated numerical expression.

Four discourse types were tested (denial of assertion; explicit contrast/threshold; implicit contrast/threshold; unlicensed) in two numerical conditions (bare *n*; *about n*). This resulted in 8 experimental conditions in total. Sample items are shown below:

(29)   ***Denial of assertion:***
When interviewed by investigators shortly after the deadly police shooting, Officer McMillan reported that he had fired his weapon "[about] 20 times over a period of several minutes." This was contradicted by witnesses in the officer's trial. **A ballistics expert testified that the officer had actually fired 37 times, not [about] 20.** Defense attorneys are expected to call further witnesses to dispute this evidence.

(30)   ***Explicit contrast:***
Political commentator James Cusack was asked to write an article of around 400 words for the Wall Street Journal about his hopes for the new president. It had been a polarizing election, and Cusack considered virtually all of the president-elect's policy proposals misguided and needlessly divisive. **He told the paper's editor that he didn't have [about] 400 words to say on the topic.** In fact, he only had four: I hope he fails.

(31)   ***Implicit contrast:***
When Pete Castro was just starting out as a blues guitarist, he played three shows a night in rundown clubs that few fans had ever heard of. Laughing, he recalls: "I remember I'd feel on top of the world if someone came up after my set to buy a CD." **These days, Castro doesn't play to audiences of [about] 15 people anymore.** Last night there were over 10,000 people at his sold-out show at the Lightning Dome.

(32)　*Unlicensed*:
　　　　American Woodmark is a manufacturer of kitchen cabinets and bathroom vanities. Home centers, including Home Depot, Lowe's, and Builder's Square, account for most of the company's business. **The Virginia-based company currently doesn't produce [about] 100 styles of cabinets.** As demand for new housing continues to rise, American Woodmark hopes to expand their range of products in the coming years.

For each discourse type, four stimulus items were constructed. For the first three discourse types, these were based on corpus examples of negated bare numerals; for the fourth discourse type (unlicensed), they were based on examples of bare numerals in positive sentences, modified to add negation. The following provides further details on their structure:

- In the **Denial of Assertion** discourse condition, the first sentence of the stimulus text contained a quoted passage in which the numerical value "[about] *n*" was asserted; this value was denied in the third sentence (see (29)). Thus in this discourse type, the bare and *about* conditions differed in both the first and third sentences. In both cases, the actual value was provided in the third or fourth sentence; in half the items this was greater than the negated value, while in the other half it was lower.
- In the **Explicit Contrast** condition, the first sentence introduced a numerical value as a threshold or point of comparison; the third sentence stated that this value did not obtain, or the threshold was not met (see (30)). In order to more clearly differentiate this condition from the Denial of Assertion condition, and in particular to discourage a quotative interpretation for the numerical expression, in both the bare and *about* versions the value in the first sentence included an approximator other than *about*, e.g. *approximately* or *around*. As above, the actual value was introduced in the third or fourth sentence, and was greater than the negated value in half the items, and lower in half the items.
- In the **Implicit Contrast** condition, the negated value was not mentioned in the preceding text, but a contrast was established between time points or states of affairs, allowing a contrastive value to be inferred (see (31)). Again the actual value was introduced in the third or fourth sentence, and was greater than the negated value in half the items, and lower in half the items.
- Items in the **Unlicensed** condition contained a single negated numerical value in the third sentence, with no form of contrast established in the preceding text, and no actual value provided in the subsequent text (see (32)).

An additional 12 filler/control items were created, again based on naturally occurring examples sourced via COCA, modified as necessary for comprehensibility and consistent structure. Of these, 7 were created to be acceptable; these included occurrences of bare and modified numerals in positive contexts, as well as licensed uses of the polarity items *some*, *any* and *many/much*. The remaining 5 were created to be ungrammatical; these featured unlicensed uses of *some*, *any* and *many/much*. A full list of critical items and controls is provided in Appendix 1 (available as a Supplementary File).

　Items were divided into four lists, each of which included 8 critical items (1 per condition) and 12 filler/control items.

### 3.3 Procedure

The experiment was programmed using the software Ibex (Drummond 2013) and hosted on Ibex Farm, a repository for Ibex experiments. Participants were recruited via Amazon MTurk and were forwarded to the hosting site. After completing the experimental task, they were given a unique code to enter on the MTurk site to receive compensation.

Participants were told they would see paragraphs of text, of the sort that they might read in a book, newspaper, or magazine. They were instructed to rate the acceptability of the bolded sentence on a scale of 1 to 7, with 1 being completely unacceptable and 7 completely acceptable. Participants were instructed to answer based on how natural the sentence sounded rather than on the basis of rules learned in school.

Critical items and fillers were presented in randomized order. In order to encourage participants to read each stimulus text in its entirety, comprehension questions were included after 8 of the filler/control items. Participants were told that if they answered too many of these attention checks incorrectly, their compensation could be negatively affected.

### 3.4 Results

Prior to analysis, data from 6 participants were excluded because they provided incorrect answers to 4 or more of the 8 comprehension questions. All participants received the same compensation regardless of performance on the comprehension questions.

Figure 1 shows the overall results for critical items by approximator condition and controls. A cumulative link mixed model was fitted to the data using the *ordinal* package (Christensen 2015) in *R* (R Core Team 2015), with acceptability rating as dependent variable, condition (Bare, About, Control-Good, Control-Bad) as fixed effect, a random by-participant slope for condition, and random intercepts for participant and item. The reference level was Control-Good. A significant difference was found between Control-Good and Bare ($z = -2.280$, $p < 0.05$), About ($z = -4.782$, $p < 0.001$), and Control-Bad ($z = -4.415$, $p < 0.001$). Planned post hoc testing via the *lsmeans* package (Lenth 2016) using Tukey correction for multiple comparison further found significant differences between Bare and About (z-ratio $= -8.961$, $p < 0.001$) and Bare and Control-Bad (z ratio $= -3.090$, $p < 0.05$), but no significant difference between About and Control-Bad (z ratio $= -0.765$, $p = 0.87$).

Figure 2 shows results for critical items broken out by discourse type and approximator condition. A cumulative link mixed model was fitted to these data with acceptability rating as dependent variable, discourse type and approximator condition as fixed effects,
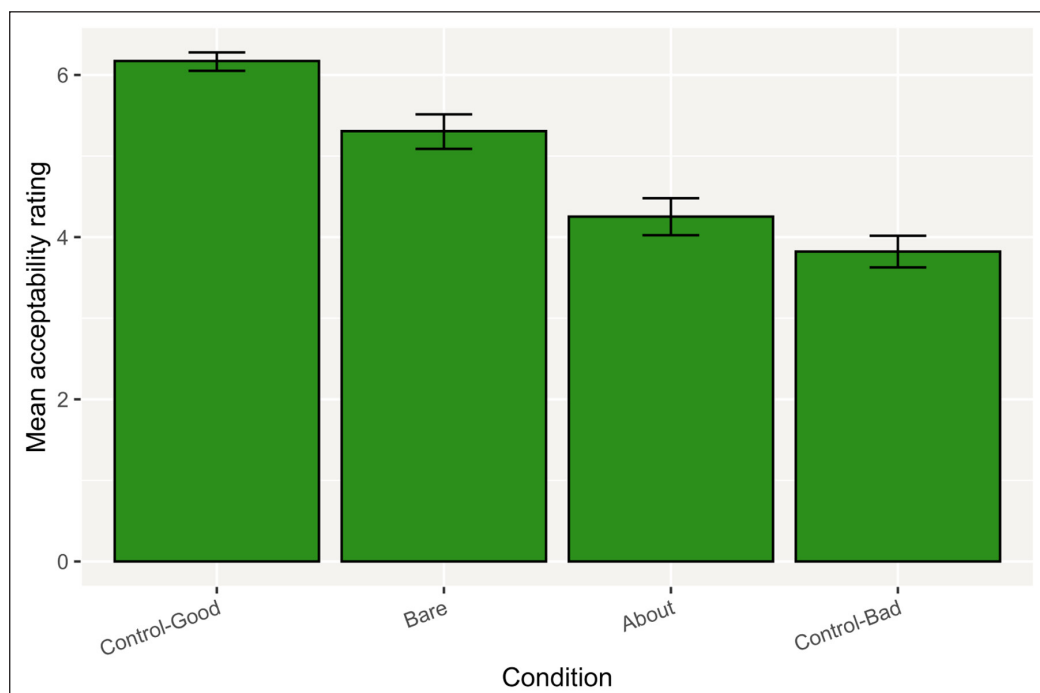


**Figure 1:** Overall results for critical items and controls (Exp. 1).
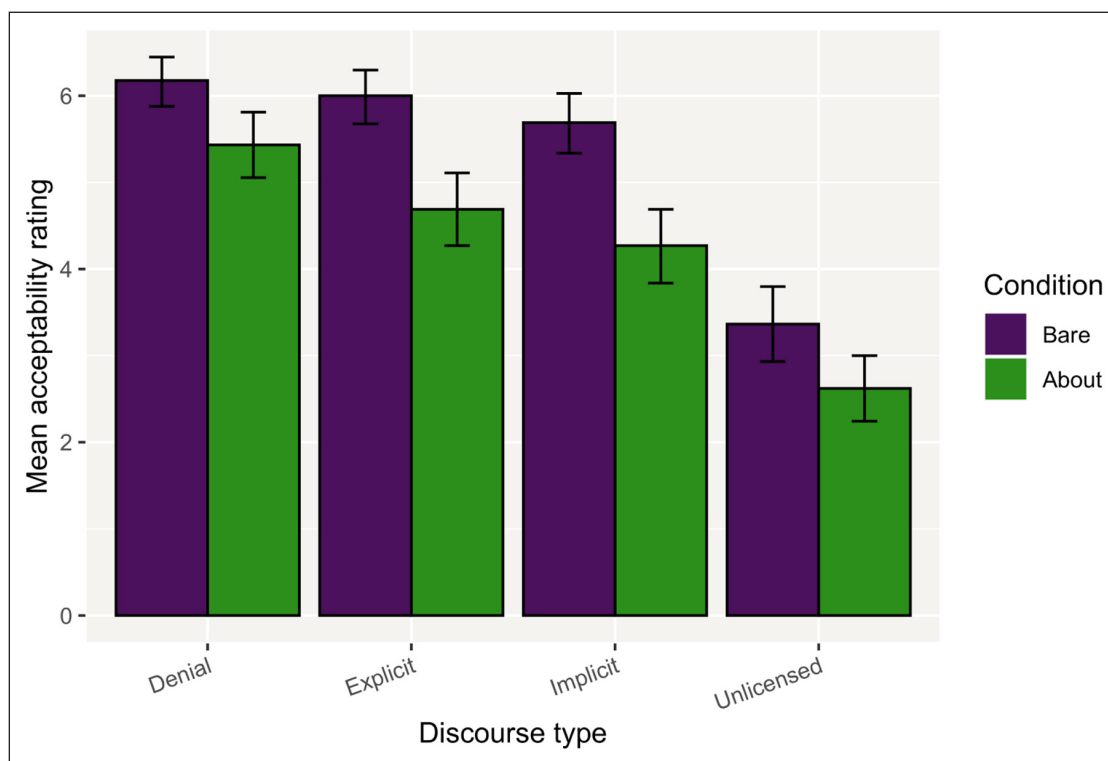
**Figure 2:** Results by discourse type and approximator condition (Exp. 1).

random by-participant slopes for condition and discourse type, and a random by-item slope for condition (along with random intercepts for participant and item). The reference levels were Denial and Bare. The difference between Denial and Implicit was marginally significant (z = –1.918, p < 0.06) while the difference between Denial and Unlicensed was significant (z = –7.982, p < 0.001); the difference between Denial and Explicit was not significant (z = –0.926, p = 0.355). Post hoc testing as above found further significant differences between Explicit and Unlicensed (z ratio = –5.827, p < 0.001) and Implicit and Unlicensed (z ratio = –5.058, p < 0.001). There was also a significant difference between Bare and About (z = –7.982, p < 0.001). The model was not improved by adding a term for the interaction of discourse type and approximator condition.

Finally, the data set was restricted to the discourse types Denial, Explicit and Implicit, for which the stimulus items specified the actual value, which was sometimes higher and sometimes lower than the negated value. A cumulative link mixed model was fitted to these data with acceptability rating as dependent variable, actual value (above vs. below) as fixed effect, random by-participant slopes for condition and discourse type, and a random by-item slope for condition (along with random intercepts for participant and item). No significant effect was found for actual value (z = 0.319, p = 0.750). Additional models demonstrated that this factor was also not significant in interaction with discourse type or approximator condition.

### 3.5 Discussion

At the aggregate level, we find that negated numerals exhibit an intermediate level of acceptability, eliciting ratings that are lower than those for grammatical control items (numerals in positive contexts and licensed polarity items), but higher – at least in the bare numeral case – than those for ungrammatical control items (unlicensed polarity items).

However, the acceptability of numerals under negation varies greatly according to the discourse context in which the negated numerical expression occurs. The most dramatic

difference in acceptability is that between discourses in which the value is explicitly mentioned or implied (Denial, Explicit, Implicit) and those in which it is not even inferable (Unlicensed). These findings thus provide support for our original observations regarding the felicity conditions on negated numerals. The more novel finding is that the differences among the first group of discourse types are relatively small. Numerically, the acceptability ratings are highest for the Denial condition, consistent with the view that denial is the prototypical use of negation, but the differences in ratings only reach marginal significance in the case of the Implicit condition. Thus we have evidence that a negated numeral is somewhat more acceptable when the value has been explicitly mentioned than when it must be inferred; but it also appears that language users are able to rely on fairly subtle contextual cues to accommodate the relevance of a negated numeral.

We further find that across discourse types, bare numerals in the scope of negation are consistently rated as more acceptable than the corresponding numerals modified by the approximator *about*. This argues against aligning bare numerals directly to other PPI numerical constructions. Put differently, while approximator-modified numerals may be affected by the same factors that are in play in the bare numeral case, the addition of an approximator such as *about* would appear to contribute a further layer of resistance to occurrence under negation. Yet both types of expression are affected by discourse context; this pattern too requires explanation.

Finally, no difference in acceptability was found according to whether the actual value was greater than or less than the value specified by the negated numeral. This is consistent with the negated numeral having its 'exact' interpretation. In Section 5 we will in fact argue for a stronger claim: in a neutral context, it is necessarily the doubly bounded 'exactly' interpretation of the numeral that surfaces, and this is the source of the infelicity of such examples.

Before pursuing this idea further, however, we note some potential limitations of the present experiment. First, because the experimental materials were based on naturally occurring examples, the test scenarios in the different discourse conditions were not fully matched. We cannot rule out the possibility that the apparent effect of discourse type that we found was in fact due to other unrelated differences between items. A further limitation is that only two numerical constructions were investigated (bare and *about*), and these only in negative sentences, raising questions regarding the generalizability of these findings. We address both of these points in our second experiment.

## 4 Experiment 2

In our second experiment, we assess the acceptability of a wider range of numerical expressions in both positive and negative sentences, contrasting neutral contexts with those in which the numerical value is made salient in the discourse context. We employ a methodology based on Cummins et al. (2012), in which the stimuli consist of brief question-answer dialogues between two speakers, with the two discourse conditions differing only in the presence or absence of the numerical value in the first speaker's assertion and the corresponding form of the question. In doing so, we seek to rule out the possibility that the effects of discourse type found in Experiment 1 may have been due to extraneous differences between the stimulus items in the four types of contexts tested.

We furthermore broaden the scope of the investigation to compare bare numerals to four modified numeral constructions: *about n, at least n, more than n* and *between m and n*. Including the first two of these allows us to further assess how the polarity sensitivity of bare numerals compares to that of established cases of positive polarity items in the numerical domain. Including the latter two allows us to directly compare a doubly bounded numerical expression (*between m and n*) and a lower-bounded one (*more than n*)

with respect to their acceptability in negative sentences, and thus to assess whether there is something about doubly bounded numerical meanings in particular that results in infelicity under negation.

Based on the literature on negation as well as the results from the previous stages of our research, we predict that negated numerical expressions will in general be less than fully acceptable in neutral contexts, but that their acceptability will improve when the numerical value is made salient or primed in the preceding discourse. We further predict differences by expression type. Specifically, we expect that bare numerals in particular will show an effect of discourse context. The findings of our first experiment lead us to expect that *about n* will be less acceptable under negation than bare numerals, and we hypothesize that similar results will be found for *at least n*, also described in the literature as a PPI. Finally, if the pattern observed for bare numerals derives from some sort of issue with negating their doubly bounded 'exactly' reading, we expect to find similar results for the doubly bounded *beween m and n* but crucially not for the lower bounded *more than n*.

### 4.1 Participants

We recruited a total of 140 participants via Amazon Mechanical Turk (MTurk). The only inclusion criteria were an acceptance rate of over 90% on prior human intelligence tasks (HITs) and a U.S. internet protocol (IP) address. MTurk workers were paid $1.50 for participation.

### 4.2 Materials

Stimulus items had the form of two-person exchanges: an assertion and subsequent question uttered by a first speaker (labeled Speaker A), followed by a response uttered by a second speaker (labeled Speaker B) which contained a numerical expression or indefinite quantifier. Participants' task was to rate the acceptability of Speaker B's response.

Critical items included one of the following 5 numerical constructions in negative sentences: bare *n*, *about n*, *at least n*, *more than n* and *between m and n*. Additionally, two types of control items were included: numerical control items containing the same 5 numerical constructions in positive sentences; and indefinite control items containing the PPI indefinite *some* and the NPI indefinite *any* in positive and negative sentences. This resulted in 14 sentence types (5 numerical + 2 indefinite × 2 polarity). Two discourse conditions were tested, neutral and primed, as illustrated by the following sample items:

(33)  **Neutral:**
      Speaker A: This afternoon, delegates will be arriving to attend the convention. How many copies of the agenda do we have for them?
      Speaker B: **We've printed about 20 copies of the agenda.** /
                 **We haven't printed about 20 copies of the agenda.**

(34)  **Primed – numerical:**
      Speaker A: This afternoon, about 20 delegates will be arriving to attend the convention. Do we have enough copies of the agenda for them?
      Speaker B: Yes. **We've printed about 20 copies of the agenda.** /
                 No. **We haven't printed about 20 copies of the agenda.**

(35)  **Primed – indefinites:**
      Speaker A: This afternoon, delegates will be arriving to attend the convention. Do we have some/any copies of the agenda for them?
      Speaker B: Yes. **We've printed some/any copies of the agenda.** /
                 No. **We haven't printed some/any copies of the agenda.**

In the **neutral** condition, Speaker A's assertion contained no numerical information, and the question was a *how many* question, which was answered by Speaker B with a positive or negative sentence containing a numerical expression or indefinite (i.e. one of the above 14 sentence types). In the **primed** condition for numerical expressions, Speaker A's assertion was identical with the exception of the inclusion of a numerical expression, and the question was a *yes/no* question that referred back to that value; Speaker B's answer was preceded by "Yes" or "No" followed by a positive or negative sentence containing the same numerical expression. For the indefinite control items, the inclusion of *any* in the assertion (as in the numerical items) would have resulted in ungrammaticality; therefore in the case of *some/any* the primed condition featured the indefinite determiner in the question instead.

Fourteen vignettes of the sort illustrated above were created, in both neutral and (minimally different) primed versions. Each sentence type was tested in each vignette. Discourse condition was tested as a between subjects factor, to rule out the possibility that exposure to primed items would cause participants to infer some significance for the numerical value even for unprimed items (see Cummins et al. 2012 for discussion of this as a possible confound). Expression and polarity were within-subjects factors: in both primed and unprimed versions of the experiment, each participant saw each of the 7 expressions of interest in both a positive and a negative sentence in a Latin Square design, for a total of 14 critical items per list (each shown within a unique discourse frame, such that no participant saw the same vignette twice). Additionally, participants saw 14 filler trials. Of these, six were designed to be grammatical (non-numerical expressions, PPIs embedded in positive sentences; NPIs embedded in negative sentences), while eight were created to be ungrammatical (PPIs under negation; NPIs out of the scope of negation). The full stimuli are provided in Appendix 2 (available as a Supplementary File).

### 4.3 Procedure

The experiment was programmed using HTML, CSS, and Javascript. We used GitHub Pages to host the experiment and Submiterator to facilitate Amazon MTurk recruitment and participant compensation.

Participants were instructed at the beginning of the experiment that they would see short dialogues between two individuals, Speaker A and Speaker B, and were asked to read the entire dialogue and then rate the acceptability of Speaker's B response (in bold) on a scale of 1 to 7, with 1 being completely unacceptable and 7 completely acceptable. They were further instructed to judge just the bolded sentence based on how natural it sounded in the dialogue, rather than basing their answer on rules of grammar learned in school.

Critical items and fillers were presented in randomized order. In order to encourage participants to read each stimulus text in its entirety, comprehension questions were included after six of the filler items. Participants were told that if they answered too many of these attention checks incorrectly, their results would not be used, and they would not receive compensation.

### 4.4 Results

Before analysis, data from 14 participants were excluded because they answered incorrectly to 3 or more of the 6 comprehension questions.

Figure 3 displays the results for critical and control items. As seen here, acceptability ratings for numerical expressions in positive sentences are consistently near ceiling, whereas those for the same expressions in negative sentences are lower and vary by expression type and priming condition.
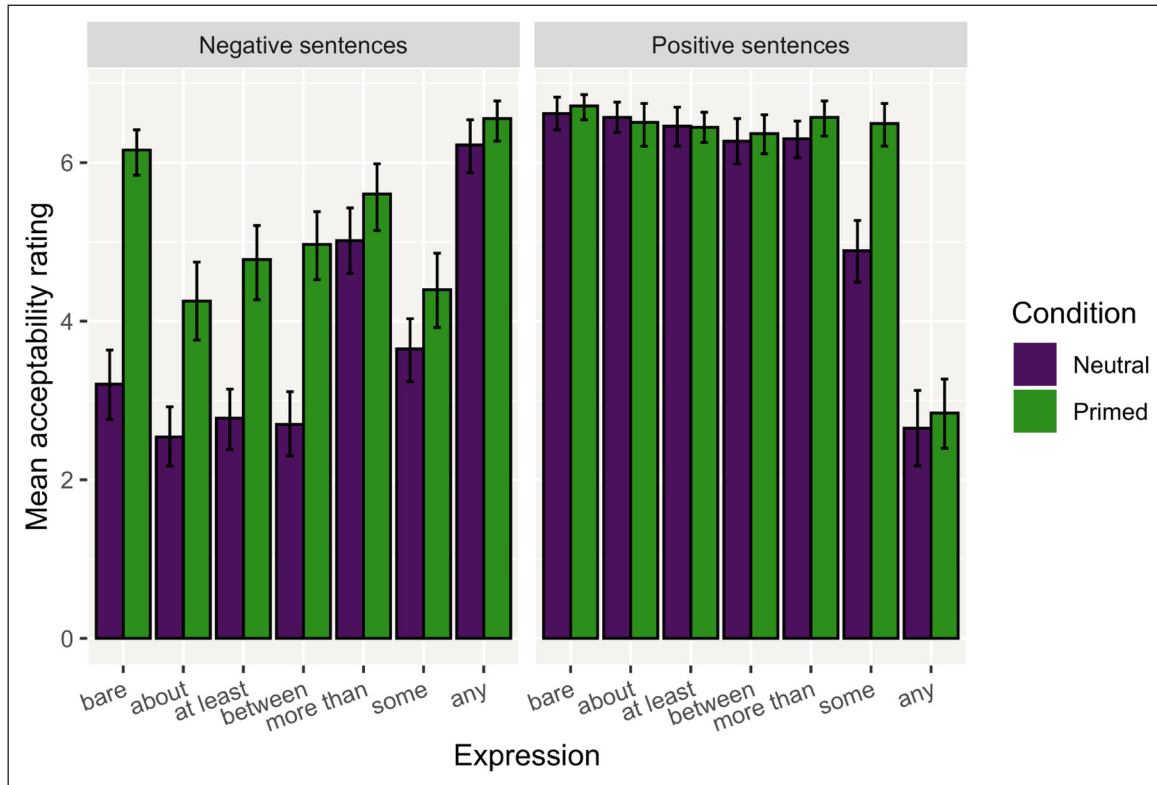
**Figure 3:** Mean acceptability ratings by sentence type and priming (Exp. 2).

To test our hypotheses, a cumulative link mixed model was fitted to the data for negative sentences, with fixed effects for expression, priming and their interaction, random by-item slopes for priming, and random intercepts for item and participant. The reference levels were bare (for expression) and neutral (for priming). As predicted, we found a significant main effect of priming, with higher acceptability in the primed condition ($z = 8.569$, $p < 0.001$). We further found significant main effects of expression, as follows: the NPI indefinite *any* was significantly more acceptable than bare numerals ($z = 7.093$, $p < 0.001$), as was the numerical expression *more than n* ($z = 5.057$, $p < 0.001$). By contrast, the modified numeral expression *about n* was significantly less acceptable than bare ($z = -2.396$, $p < 0.05$), and a near-significant effect in the same direction was found for *between m and n* vs. bare ($z = -1.936$, $p = 0.053$). No significant difference was found between bare and *at least n* or *some*. Finally, and differently from our first experiment, significant interactions of expression and priming were found, with all other expression types exhibiting less sensitivity to priming than bare numerals. This was in particular the case for the NPI *any*, the PPI *some* and the numerical *more than* (*any*: $z = -4.678$, $p < 0.001$; *some*: $z = -6.060$, $p < 0.001$; *more than*: $z = -5.265$, $p < 0.001$); post hoc testing (*lsmeans* package with Tukey correction for multiple comparison) showed no significant difference between primed and neutral conditions for these three expression types. For the remaining expression types there was an effect of priming, but this was significantly less pronounced than that for bare numerals (*about*: $z = -3.238$, $p < 0.01$; *between*: $z = -2.116$, $p < 0.05$; *at least*: $z = -2.715$, $p < 0.01$).

As a control to ensure that the patterns described above were due to the presence of negation, a comparable cumulative link mixed model was fitted to the data for positive sentences. The results were markedly different. There was no main effect of priming. Regarding expression, the most prominent effects were in the indefinite control items, specifically a significantly lower level of acceptability for NPI *any* vs. bare ($z = -10.993$, $p < 0.001$) and a more unexpected lower level of acceptability for PPI *some* ($z = -6.438$, $p < 0.001$) vs. bare,

as well as a significant interaction of *some* and priming (z = 3.688, p < 0.001). These latter effects reflect a lower level of acceptability of positive *some* in the neutral condition, which we attribute to a mild infelicity of answering a *how many* question with *some*, an effect unrelated to the issue under investigation. Among the numerical expression types, the only effects found were significant or near-significant main effects for *more than* and *between*, both of which tending to be less acceptable than bare (*more than*: z = –2.209, p < 0.05; *between*: z = –1.730, p = 0.083); no interactions of expression and priming were found.

### 4.5 Discussion

The results of our second experiment provide further substantiation for the main empirical claim of our paper. In a neutral context, specifically as the answer to a *how many* question, bare numerals are judged to be quite unacceptable in the scope of negation. But their acceptability improves dramatically when the numerical value is introduced in the immediately preceding discourse context.

Both in their degree of acceptability in negated sentences and their sensitivity to discourse context, bare numerals were found to pattern distinctly from all of the other numerical and quantificational expressions investigated. Starting with the polarity-sensitive indefinites that were included as control items, our results were largely as expected: *any* was judged acceptable in negative sentences but highly degraded in positive ones, while the reverse was found for *some* (modulo a moderate decrease in acceptability in positive sentences in the neutral condition, which we attributed to the particular structure of the experimental items). Importantly, in their unlicensed contexts (positive for *any*, negative for *some*), the acceptability of these expressions was not improved when they were mentioned earlier in the discourse context – a direct contrast to what was observed for bare numerals.

Turning to the numerical expressions characterized in the literature as PPIs, namely *about n* and *at least n*, we find their behavior in negative sentences to be qualitatively similar to that of bare numerals, in that they receive low ratings in neutral contexts but improve when their content is made salient in the prior discourse. But *about* is less acceptable overall than bare, and both are less improved by the contextual manipulation than are their bare counterparts. Put differently, the infelicity of bare numerals under negation is almost fully obviated by a supportive discourse context, resulting in acceptability ratings approaching those for numerical expressions in positive sentences; but the same is not the case for the PPIs *about n* and *at least n*.

Particularly interesting is the comparison between the modified numeral expressions *more than n* and *between m and n*. These two are similar in that they both convey ranges of values, and they have been observed to pattern together with respect to certain interpretive phenomena, particularly the absence of ignorance inferences (Nouwen 2010). But they differ in that *more than* has a one-sided or lower-bounded interpretation, whereas *between* has a doubly bounded interpretation, and this difference correlates with a difference in their acceptability in the scope of negation. Specifically, *between* sentences show the same neutral/primed asymmetry observed for bare numerals (though like *about* and *at least* being less acceptable overall and less improved by priming). By contrast, *more than* is relatively acceptable even in the neutral context, and is not improved significantly when its numerical content is made salient in the discourse. From this we conclude that there is something about doubly bounded numerical meanings in particular that results in infelicity when negated in a neutral context.

In the next section, we take this conclusion as the basis for a formal theory of the contextual constraints on numerical utterances, which relies centrally on the notion of convexity of meaning. We apply it to account for the facts relating to bare numerals, which we argue to also have a doubly bounded exact reading in netural contexts.

## 5 Proposal

### 5.1 Convexity and negated numerals

In this section, we develop a formal semantic/pragmatic proposal to account for the patterns of acceptability established in our experimental research.

To recap, the crucial contrasts are the following: a wide range of numerical expressions – including negated ones – can be used to answer a *how many* question. A negated bare numeral, however, cannot (see (36)). But the same negated numeral is fully acceptable when the numerical value has been previously mentioned, or is otherwise salient in or inferable from the broader context (per (37)).

(36)    How many sheep does Lisa have?
      a.    She has 40 sheep.
      b.    She has between 40 and 50 sheep.
      c.    She has more than 40 sheep.
      d.    She doesn't have more than 40 sheep.
      e.    #She doesn't have 40 sheep.

(37)    Fred has exactly 40 sheep. Does Lisa have the same number?
      a.    No. She doesn't have 40 sheep. [She has 25 / 200.]

The central intuition that we pursue here is that what goes wrong with a negated example such as (36e) in the given context is that on the exact interpretation of the numeral, the meaning of the sentence – that is, the set of situations in which it is true – corresponds to a disjoint rather than convex region of the number line. As depicted below, all of the felicitous examples in (36a–d) describe connected or convex numerical ranges, meaning that if two points are in the range, so too are all points between them. But (36e) is true of values either below or above 40, excluding the single point in between.

(38)



$40_{EXACT}$
between 40 and 50
more than 40
not more than 40
not $40_{EXACT}$
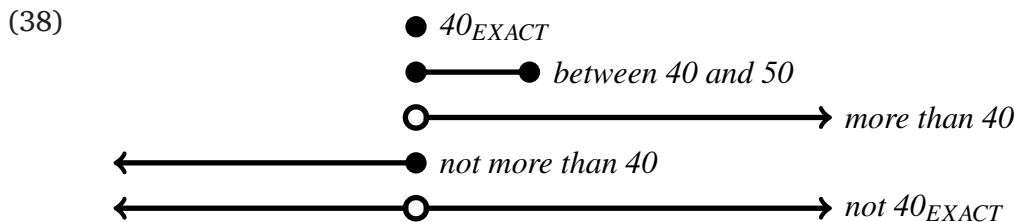
The linguistic relevance of the mathematical property of convexity was established most famously by Gärdenfors (2004; 2014), who argues that the properties expressed by simple words of natural language can largely be analyzed as connected and more specifically convex regions in some conceptual space.[2] Convexity is proposed to facilitate inferencing and concept acquisition, and can be linked to the prototype-based structure of concepts: given an appropriate distance metric, a set of prototypes induces a partition of a conceptual space into a set of convex regions. Originally connectedness and convexity were hypothesized to be constraints on the meaning of content words such as nouns and adjectives. But recently, Chemla et al. (2019) propose that the notion of connectedness can be extended to function words as well, in particular quantifiers, where it can be related to the well-known

---

[2] In an *n*-dimensional space for $n \geq 2$, convexity is a stronger requirement than connectedness. In a 1-dimensional space such as the number line, the two properties coincide. As such, the distinction between them is not crucial for the present purposes. We choose to use the term convexity to better reflect the connection to previous work.

property of monotonicity (Barwise & Cooper 1981). They demonstrate that in an artificial quantifier learning task, the connectedness of a rule facilitates its acquisition, a finding that makes this a possible candidate for a semantic universal.

Our present claim amounts to taking this a step further, in that we propose that convexity also plays a role at the level of sentences uttered in discourse. The above-described interpretation of *not forty* fails to be convex. Its restricted distribution might then be related to informativity and failure of inferencing. The disjoint interpretation of *not forty* is almost maximally uninformative, excluding only a single point on the number line. It furthermore gives no information about the direction in which the true value deviates from that excluded point, greatly limiting the sorts of inferences that might be drawn from its utterance. Such an explanation is in line with proposals put forward in the literature on negation, according to which the infelicity of negative sentences in out of the blue contexts is related to their lack of informativity (e.g. Givón 1978; 1979): whereas *The hat is red* specifies a particular state of affairs, *The hat is not red* is compatible with multiple possibilities (e.g. the hat being blue, black, green, and so forth).

That convexity (or the lack thereof) is in fact the crucial factor underlying the infelicity of negated bare numerals receives support from our experimental findings for *more than n* and *between m and n*. The negation of the former has a convex interpretation, that of the latter a disjoint interpretation; correspondingly, the former can be felicitously negated in neutral contexts, while the latter cannot.

A small additional piece of supporting evidence comes from cases where a numerical expression denotes a scalar endpoint. In describing probabilities or proportions of a whole, even when *100%* receives a punctual or exact meaning, its negation denotes a convex region of the scale, because there are no higher values on the scale, only lower ones. We thus predict that *not 100%* – unlike, say, *not 95%* – should be felicitous in a neutral context, and that is precisely what is seen in examples such as the following:

(39)    How likely is it that our company will be awarded the contract?
         a.    It's 95% / 100% certain.
         b.    It's not 100% certain.
         c.   ??It's not 95% certain.

To formalize our proposal for the role of convexity, and in particular to account for the rescuing effect of prior mention of the numerical value, we adopt the view that the immediate discourse context of an utterance can be represented as a question, the so-called "question under discussion" or QUD (Roberts 1996; 2012), which captures what the discourse is about at a given point. The examples in (36) and (37) feature explicit questions and their answers, which is of course not always the case. We follow authors including van Kuppevelt (1995) and more specifically Scharten (1997) in taking the view that the structure of discourse can be understood as a hierarchically organized set of (generally implicit) questions and their answers. We further adopt a partition semantics for questions (Groenendijk & Stokhof 1984), according to which the meaning of a question – either an explicit one or an implicit QUD – is construed as a partition of the space of logical possibilities.[3]

---

[3] In subsequent work, evidence has been put forward that the denotation of a question cannot itself be a partition; rather, a partition can be derived from some more basic meaning of a question, e.g. a set of possible answers (Heim 1994; Dayal 1996; Fox 2018). We believe that our proposal can be made compatible with developments along these lines, but maintain the partition view in what follows for ease of exposition.

We are now able to characterize what we have somewhat loosely been referring to as a neutral context for a numerical expression as one in which the QUD is an (explicit or implicit) *how many* question. In the case of the small dialogue in (36), the meaning of this question can be expressed as follows:

(40)     $[\![QUD]\!]$ = $[\![$How many sheep does Lisa have?$]\!]$ =

$$
\begin{Bmatrix}
\dots, \\
\lambda w.\ \text{Lisa has (exactly) 38 sheep in } w, \\
\lambda w.\ \text{Lisa has (exactly) 39 sheep in } w, \\
\lambda w.\ \text{Lisa has (exactly) 40 sheep in } w, \\
\lambda w.\ \text{Lisa has (exactly) 41 sheep in } w, \\
\dots
\end{Bmatrix}
$$

As represented in (40), the meaning of a QUD is an unstructured set of propositions. But at least in the case under consideration, a structure can be imposed on it on the basis of the underlying order of the number line. This in particular allows us to establish a "between-ness" relation on members of the set: for any two distinct propostions of the form *λw.Lisa has exactly n sheep in w* and *λw.Lisa has exactly m sheep in w*, a third proposition *λw.Lisa has exactly k sheep in w* is between them iff $n < k < m$ or $m < k < n$. And this in turn allows us to speak of subsets of a set such as (40) as being convex or disjoint: for a QUD denotation on which a between-ness relation is defined, a subset $S \subset [\![QUD]\!]$ is convex iff for all $p,\ q,\ r \in [\![QUD]\!]$, if $p,\ q \in S$ and $r$ is between $p$ and $q$, then also $r \in S$.

With this in place, we propose the following discourse constraint on numerical expressions:

(41)     **Felicity constraint on numerical assertions:** The felicitous assertion of a declarative sentence $\phi$ containing a numerical expression $\alpha$ in a context $C$ requires that $[\![\phi]\!] = \cup S$ for some **convex** subset $S \subset [\![QUD_C]\!]$.

The constraint in (41) has the effect of imposing a matching requirement on the meaning of a numerical sentence and the context of utterance, ensuring that the assertion provides a suitably informative answer to the currently active QUD.

Turning to the possible answers to such a question, we adopt a degree-based semantics for bare and modified numerals (e.g. Nouwen 2010), and further assume that bare numerals have both 'exact' and 'at least' interpretations that are semantically encoded. For concreteness we represent these in the system of Kennedy (2015), according to which the 'exact' interpretation involves a degree quantifier incorporating a maximality operator (42), whereas the 'at least' interpretation is based on type lowering of the quantifier to a type $d$ interpretation which can take scope under an existential quantifier.

(42)     $[\![\text{forty}_{\langle dt,t\rangle}]\!] = \lambda I_{dt}.max_d(I) = 40$

The following then gives the semantics of some of the possible answers in (36).

(43)     a.    $[\![\text{Lisa has 40 sheep}]\!] =$
                 $= \lambda w.max_d(\exists x[sheep(x) \wedge has_w(L,x) \wedge |x| = d]) = 40$
                 $= \cup\{\lambda w.\text{Lisa has (exactly) 40 sheep in } w\}$                    ✔

b.  $[\![$Lisa has between 40 and 50 sheep$]\!]=$
   $=\lambda w. 40 \leq max_d(\exists x[sheep(x) \wedge has_w(L,x) \wedge |x|=d]) \leq 50$
   $=\cup\{\lambda w.$Lisa has (exactly) $n$ sheep in $w$: $40 \leq n \leq 50\}$                    ✔

c.  $[\![$Lisa doesn't have more than 40 sheep$]\!]=$
   $=\lambda w. \neg max_d(\exists x[sheep(x) \wedge has_w(L,x) \wedge |x|=d]) > 40$
   $=\cup\{\lambda w.$ Lisa has (exactly) $n$ sheep in $w$: $n \leq 40\}$                    ✔

d.  $[\![$Lisa doesn't have 40 sheep$]\!]=$
   $=\lambda w. \neg max_d(\exists x[sheep(x) \wedge has_w(L,x) \wedge |x|=d]) = 40$
   $=\cup\{\lambda w.$Lisa has (exactly) $n$ sheep in $w$: $n \neq 40\}$                    ✘

As seen here, (43a–c) each have meanings that are equivalent to the union over some (possibly singleton) convex subset of (40), and therefore satisfy the felicity constraint in (41). But in the case of the negated bare numeral in (43d), there is no such convex subset whose union produces the meaning of the sentence, because that meaning is inherently disjoint. Because the sentence fails to satisfy the constraint in (41), it is infelicitous in the given context.

We turn now to the case where a negated numeral occurs in a supportive discourse context. Following proposals by Scharten (1997) for numerals and Tian et al. (2016) for negated utterances more generally, we take the position that the effect of such a context is to shift the QUD from a *how many* question to a polar question of the form *does n obtain?* In (37) – as in the primed condition in our second experiment – this question is overt. But a question of this sort can also be inferred from a discourse in which the numerical value is mentioned or otherwise made salient.

In a context of this sort, the QUD establishes a simple 2-cell partition, as in the following representation of the question in (37):

(44)    $[\![QUD]\!]=[\![$Does Lisa have (exactly) 40 sheep?$]\!]=$

$$\left\{ \begin{array}{l} \lambda w.\text{Lisa has (exactly) 40 sheep in } w, \\ \lambda w.\text{Lisa does not have (exactly) 40 sheep in } w \end{array} \right\}$$

In this context, unlike the one represented in (40), the meaning of the negative *Lisa doesn't have 40 sheep* does correspond to a convex subset of the QUD, specifically the singleton set containing the negative answer (which is trivially convex). Thus the felicity constraint (41) is satisfied, and the sentence is acceptable.

Recall that our first experiment suggested that the acceptability of negated numerals is somewhat gradient in nature. We can now recast this effect in QUD terms. The easier it is to construct from the context an implicit QUD of the form *does n obtain,* the more acceptable is an assertion of *not n*. In contexts where the numerical value is explicitly asserted or otherwise mentioned, such a question is easily accommodated, whereas when it is only implied, accommodation may be more difficult, resulting in lower acceptability. But when the discourse context is such that no such question can be accommodated, and instead the only implicit QUD that can be inferred is a *how many* question, infelicity results.

### 5.2  Why no 'at least' reading

There is an obvious question that arises at this point. In the above discussion we have assumed the exact interpretation of bare numerals. On this reading, negation produces a disjoint meaning, which we have argued results in infelicity in a neutral context. But as discussed above, cardinal numerals also have an 'at least' reading. In the framework we

have adopted, this is obtained via type lowering of the numeral to an interpretation of type $d$, which can take scope under an existential quantifier (Kennedy 2015):

(45)     $[\![40_d]\!] = IOTA(BE([\![40_{\langle dt,t \rangle}]\!])) = 40$

(46)     $[\![\text{Lisa has 40 sheep}]\!] = \lambda w.\exists x[sheep(x) \wedge has_w(L,x) \wedge |x| = 40]$

The negation of the 'at least' interpretation is convex; the negation of (46), for example, is equivalent to 'less than 40', which can of course be stated in terms of a convex subset of the QUD set in (40). Why then can't a negated bare numeral in an otherwise unlicensed context simply be shifted to this interpretation, thereby eliminating the violation of the felicity constraint?

    We would like to propose that the unavailability of rescue via this route can be attributed to the restricted availability of the 'at least' reading itself. As discussed in Section 1, both linguistic tests and psycholinguistic findings demonstrate that the two-sided exact reading of cardinal numerals is the more salient one, occurring in contexts where other scalar items have only their lower bounded interpretations. On the semantics we have assumed here, this is not unexpected, in that the exact interpretation is the basic one, whereas the 'at least' one is derived from it.

    But beyond this, there is reason to think that the 'at least' reading of cardinals is not just dispreferred, but is also subject to discourse contextual restrictions that rule out its occurrence in precisely those negative contexts where it would be needed to avoid a non-convex interpretation. A proposal that is put forward by van Kuppevelt (1996) and developed further by Scharten (1997) is that when an unmodified numeral occurs in comment position, serving as a partial or complete answer to the current question under discussion, it necessarily receives an exact interpretation, which is truth conditional in nature. The 'at least' reading is only possible when a numeral occurs in non-comment position – that is, when it is part of what is asked (the QUD) rather than part of the answer. Scharten supports this claim with examples such as the following, which demonstrate that the upper bound conveyed by a cardinal numeral may be cancelled when it occurs in topic (non-comment) position (47), but not when it occurs in comment position (48):

(47)     Q:   Who has three cows?
         A:   JOHN has three cows, in fact ten.

(48)     Q:   How many cows does John have?
         A:   John has THREE cows, # in fact ten.

Scharten further proposes that this distinction carries over to negated examples: whether *John doesn't have three children* should be interpreted as 'not exactly three' or 'fewer than three' depends on whether the value three occurred in topic or comment position in the preceding discourse. Interestingly, she does not consider the case corresponding to our neutral condition, where a negated numeral occurs as the answer to a *how many* question (i.e. in comment position) without having been mentioned in the preceding discourse. But we take her theory to predict that here too, the numeral should be interpreted exactly, for the following reason: on Scharten's account, information structure is syntactically encoded, with topic-comment constructions underlyingly containing a specificational predicate *BE* that assigns a value (the comment) to a function (provided by the topic). A *how many* question asks for the (exact) value that equals the cardinality of a set (e.g. the set of Lisa's sheep). A positive answer (e.g. *she has 40*) specifies this value; a negative

answer (e.g. *she doesn't have 40*) then seemingly has to be interpreted as asserting what this value is not.

While we do not endorse Scharten's particular formal implementation, we believe that the intuition behind it is very much correct. The central idea is that in a discourse context in which the overt or inferred QUD is a *how many* question – which is how we have characterized a neutral context – a numerical assertion necessarily says something about an exact value. A positive or negative assertion based on the exact interpretation of the numeral, as in (43a) and (43d), does this; the negative sentence however is ruled out by a violation of the convexity constraint. But an existential sentence of the form in (46) or its negation does not make a statement about an exact value, and is thus also ruled out in this context. There is no possibility of shifting to an 'at least' interpretation in a neutral context, and therefore no rescue from ill-formedness. It is only when the context changes such that the numerical value is in topic position (part of the QUD) that it may be felicitously negated, either because the QUD establishes an 'at least' interpretation for the numeral, and/or because because the QUD is a polar question for which both positive and negative answers are trivially convex.

Here we have to acknowledge that we cannot offer a proposal for how to formalize the discourse restrictions on the availability of the exact and 'at least' interpretations that we have outlined above, especially if one chooses not to adopt Schartens' rather non-standard syntactic and semantic assumptions. We do though note a connection to the observation by Rullmann (1995) that a *how many* question asks for a maximal (or maximally informative) answer. To ask how many sheep Lisa has is to ask what the maximum number $n$ is such that there is a set of $n$ sheep that she owns. This suggests that it would be fruitful to further explore the connections between question meaning on the one hand and numeral interpretation on the other.

We also observe the following independent support for a link between the availability of the 'at least' reading and the possibility of felicitously negating a bare numeral. A negated numeral can be shifted to its lower bounded interpretation grammatically via the focus-sensitive particle *even*, and this shift goes hand-in-hand with an obviation of the infelicity under negation. For example, (49b) unambiguously means that Lisa has fewer than 40 sheep,[4] and in contrast to the minimally different example without *even* is acceptable in the given context.

(49)   a.   How many sheep does Lisa have?
       b.   She doesn't even have 40.

Recall also from the corpus analysis in Section 2 that one sort of naturally occurring example of negated bare numerals involves what we called "negation of a minimum significant value" (see the discussion of examples (22)–(24)). In such examples, the specific value that is negated is not mentioned in nor even inferable from the preceding discourse; rather, the felicity of these examples rests on world knowledge to tell us that the value in question represents some sort of minimum threshold for what would count as significant in the broader context of utterance. Importantly, in this use, the negated numeral necessarily has a 'less than' interpretation; that is, what is conveyed is the negation of the 'at least' reading of the numeral. Thus here too we see a correlation between a shift to a lower-bounded reading and aceptability under negation. We noted in Section 2 that

---

[4] We are not aware of any work that explicitly investigates this effect. We suspect that it derives from the presupposition of *not even p* that *not p* is less likely than all alternatives *not q*, and further that some or all alternatives *not q* are also true (see Karttunen & Peters 1979; Rooth 1985; 1992; Wilkinson 1996; Schwarz 2005; Collins 2016). We are not able to pursue this issue further here.

such uses of negated numerals could be characterized as 'not even' uses, since the effect is similar to what obtains with overt *even*. We therefore suggest that they involve a covert counterpart to *even*, something that has been proposed on independent grounds to play a role in NPI licensing (Krifka 1995; Lahiri 1998; Crnič 2011; Chierchia 2013).

To conclude this section, we take our findings to indirectly support the view that the default interpretation of cardinal numerals – that is, the one that arises in neutral contexts – is the exact one; furthermore, this is also the case when the numeral occurs in the scope of negation. It is by taking this position that we can explain the infelicity of such examples, as well as the obviation of this infelicity when the interpretation is shifted by overt or covert means to the 'at least' one.

We also believe that this discussion sheds light on why it is that the intuitions reported in the literature are just the opposite, namely that bare numerals in the scope of negation have an 'at least' reading. In an out of the blue context, a negated numeral is simply ill formed. Thus to judge such examples, it is necessary to infer an appropriate discourse context, one in which the numerical value is in some way salient. We suspect that it is easier to accommodate a context in which that value corresponds a minimum threshold than one in which it represents an exact point of comparison. In the context of a question of this form, the negated numeral has its 'at least' reading; that is, *not 40* is interpreted as 'less than 40' (cf. the discussion of (6) in Section 1). We thus believe that the observed tendency to interpret negated numerals as lower-bounded does not so much tell us something about the preferred interpretation of the numeral itself, but rather about the most plausible context of utterance.

### 5.3 Beyond bare numerals

Our primary objective in this section has been to account for the patterns characterizing negated bare numerals. In concluding we briefly examine the behavior of the other numerical expressions included in our empirical research, as well as some facts from beyond the numerical domain.

Our experimental results showed that the modified numeral expressions *about n, at least n* and *between m and n* exhibit the same neutral vs. primed difference observed for bare numerals in the scope of negation. For *between* and *about*, the convexity constraint may be relevant: the former and plausibly the latter have two-sided meanings that when negated yield a disjoint interpretation. But *at least n* has a lower-bounded meaning similar to that of *more than n*, and as such the interpretation that arises via negation is entirely consistent with the felicity constraint in (41). A similar example is the disjunctive *n or more*, which was not included in our experiment, but which intuitively exhibits PPI-like behavior similar to *at least*. We must therefore conclude that lack of convexity is not the only source of polarity-based restrictions in the numerical domain; some other mechanism or mechanisms must also be in play. This is further supported by our experimental findings for *about* and *between*: both of these are less acceptable than bare numerals in the scope of negation, and less improved by priming in the discourse context, suggesting that some additional factor must contribute to their degraded status.

We are not in a position to propose a comprehensive theory of polarity sensitivity in the numerical domain, but we briefly review an account of one of these cases, namely *about*, based on Solt (2018). Working within a neo-Gricean alternative-based framework based on Katzir (2007), Solt analyzes the polarity sensitivity of approximator-modified numerals as deriving from competition with the corresponding unmodified numerals. The latter are calculated to be 'better than' the approximator-containing alternatives, being simpler and (in the sense Solt assumes) not definitively different in informativity. The result is an implicature that the unmodified form is not assertable. In the positive case the implicature

is well-formed (the assertion of *about 40* implicates that the speaker is not in a position to assert *(exactly) 40*). In the negative case, however, it results in a contradiction, producing ungrammaticality. Solt does not explicitly address the rescuing effect of discourse context, but her theory might be extended to specify that in a discourse context in which *about n* has previously been mentioned, the bare alternative tends to be removed from contention, thereby eliminating the source of contradiction and the resulting ungrammaticality.

An approach similar to that applied to *about* could potentially be extended to *between* constructions, and perhaps also to *at least*. Alternately, the polarity sensitivity of *at least* may relate in some way to its arguably more complex semantics, which has been proposed to involve modality (Geurts & Nouwen 2007), disjunction (Büring 2008), or an operation over speech acts (Cohen & Krifka 2014). Both Geurts & Nouwen and Cohen & Krifka proposed explanations for the polarity-based restrictions on *at least* based on their particular semantic analyses. Which of the possible analytical approaches will prove most explanatory may depend on what ultimately is determined to be the correct semantic analysis of *at least*.

The felicity constraint in (41) was stated with reference to sentences containing numerical expressions. It is unlikely, though, that such a principle of language use would apply only to such a narrow class of assertions. Thoroughly investigating the possible role of convexity outside the domain of number words would take us beyond the scope of the present paper, but we briefly note that parallel non-numerical examples can also be constructed. In the context of an Olympic ski race, for example, (50a) describes a non-convex region of the space of logical possibilities, encompassing results both better than and worse than third place; correspondingly, it is infelicitous as an answer to a neutral QUD. By contrast, (50b) has a convex interpretation (any place below the top three), and while perhaps somewhat lacking in informativity is considerably better than (50a) in the neutral context. Finally, just as in the numerical case, the addition of *even* removes the infelicity:

(50)    [In the context of an Olympic ski race:] How did Sue do?
    a.  ??She didn't win the bronze medal.
    b.    She didn't win a medal.
    c.    She didn't even win the bronze medal.

Other similar cases can be found. For example, a gradable adjective in combination with a modifier such as *fairly* or *somewhat* has a doubly bounded interpretation (*fairly good* conveys 'moderately but not extremely good'); correspondingly, such modifiers in English as well as other languages are PPIs (see e.g. van Os 1989 for German). Even the infelicity of an example such as *The hat is not red* in a neutral context might be assimilated to this pattern, in that *not red* describes a non-convex region in the color space (Gärdenfors 2004).

The present proposal also opens up a potentially productive line of investigation of facts relating to scalar implicature.[5] It has long been been recognized that weak scalar terms such as *some*, *possible*, *believe* and *or* give rise to upper-bounding scalar implicatures in positive sentences (e.g. *possible* implicates *not certain*), but that these implicatures fail to arise in the scope of negation and other downward-entailing environments (Horn 1972; Gazdar 1979: and ff.). A standard explanation is that the exhaustification mechanism responsible for scalar implicatures only applies if it has a strengthening effect (e.g. Chierchia 2004), which is the case in positive but not negative contexts. But approaching these facts from the perspective of the present proposal suggests a slightly different explanation. The

---

[5] We thank an anonymous reviewer for suggesting that we pursue this connection, and regret that we can only scratch the surface in doing so.

implicature-strengthened interpretation (e.g. 'possible but not certain') is doubly bounded. Thus perhaps this interpretation fails to arise in the scope of negation not simply because it would be less informative than the basic semantic one, but rather because it would be uninformative in a particular way, describing a disjoint rather than convex region of the relevant scale. In fact, exactly this sort of explanation is proposed by Enguehard & Chemla (2019) for the unavailability of certain readings of weak scalar items that could in principle be generated by the application of a covert exhaustification operator: these are blocked, they argue, by a constraint that specifies that parses resulting in non-connected meanings are dispreferred. While the specifics of their account differ from ours, the central idea is very similar. There is, though, a crucial difference between numerals and other scalar items, namely that in the latter case the apparent constraint against non-convex meanings does not result in ungrammaticality under negation but instead a preference for the unenriched lower-bounded interpretation. This is further evidence of the different status of the two-sided interpretation of number words versus that of other scalar terms.

The above brief discussion has suggested that a convexity constraint along the lines of (41) has more general applicability beyond the domain of number words. At the same time, it cannot be inviolable. Speakers of course have occasion to communicate non-convex meanings, and correspondingly languages have ways to express such meanings, notably via disjunction:

(51)    Lisa has either fewer than 40 or more than 50 sheep.

Thus at least to some extent, the felicity condition as we have formulated it here overgenerates.

It is not entirely clear to us at this stage what exceptions there are to the postulated constraint against non-convex meanings in discourse, and thus how exactly the operation of (41) should itself be restricted. At one end of the space of possibilities, we might conclude that (41) must be be construed as applying exclusively to negative utterances. This would however fail to capture the connection to convexity as a constraint on lexical meanings and its possible role in implicature calculation. At the other extreme, it might turn out that the constraint against non-convex meanings in discourse is in operation by default, excluding only some narrow class of exceptions, perhaps limited to disjunction and lexically non-convex meanings (e.g. *an odd/even number of*). With regards to disjunction, Chemla et al. (2019) observe that to ensure that convexity is preserved for the disjunction of two quantifiers would require one of them to necessarily be trivial, rendering the entire disjunction useless. Disjunction thus emerges as a natural way of expressing non-convex meanings. We also note that it is not entirely obvious what discourse constraints there may be on the assertion of non-convex quantificational expressions such as *either fewer than 40 or more than 50* and *an even number of*. In fact, Enguehard & Chemla (2019) mark an example parallel to (51) as degraded, which we suspect reflects difficulty in inferring an appropriate context in which it might be uttered. Thus here too, the operation of a constraint of the sort we have proposed may in fact be in operation. We think that further research – and specifically experimental research – will be necessary to clarify these issues.

## 6 Conclusions

The primary empirical contribution of this paper is to show that bare numerals in the scope of sentential negation are infelicitous in an out of the blue context, but perfectly acceptable if the numerical value is made salient in the discourse context. This finding is we believe conclusively established by our experimental results, which further demon-

strate that bare numerals in this respect pattern subtly but systematically differently from other numerical expressions and non-numerical polarity items. We propose an account for these findings based on a constraint that numerical expressions must provide a convex answer to the current QUD, coupled with a previous proposal that bare cardinal numerals in neutral contexts are necessarily interpreted exactly.

We see several broader implications from the findings and analysis. First, they add to other evidence that the default interpretation of number words – even in negative contexts – is the exact one. Second, they demonstrate that the mathematical notion of convexity, first proposed as a constraint on the possible meaning of content words, is also relevant at the level of discourse. Finally, we believe these findings highlight the importance of investigating patterns of acceptability and interpretation in context. Without taking context into account, the data relating to negated numerical expressions are puzzling; but when we consider such expressions situated in a discourse, the picture is more systematic, and different from how it might initially appear.

## Abbreviations

COCA = Corpus of Contemporary American English, HIT = Human Intelligence Task, MTurk = Amazon Mechanical Turk, NPI = negative polarity item, PPI = positive polarity item, QUD = question under discussion

## Additional Files

The additional files for this article can be found as follows:

- **Appendix 1.** Stimuli for Experiment 1. DOI: https://doi.org/10.5334/gjgl.736.s1
- **Appendix 2.** Stimuli for Experiment 2. DOI: https://doi.org/10.5334/gjgl.736.s2

## Acknowledgements

## Funding Information

## Competing Interests

The authors have no competing interests to declare.

## References

Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4(2). 159–219. DOI: https://doi.org/10.1007/BF00350139

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2). 93–139. DOI: https://doi.org/10.1093/jos/ffm016

Büring, Daniel. 2008. The least *at least* can do. In Charles B. Chang & Hannah J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 114–120. Somerville, MA: Cascadilla Proceedings Project.

Chemla, Emmanuel, Brian Buccola & Isabelle Dautriche. 2019. Connecting content and logical words. *Journal of Semantics*. Published online February 22, 2019. DOI: https://doi.org/10.1093/jos/ffz001

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and beyond* 3. 39–103. Oxford: Oxford University Press.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice and intervention*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199697977.001.0001

Christensen, R. H. B. 2015. Ordinal: Regression models for ordinal data. http://www.cran.r-project.org/package=ordinal/. R package version 2015.6-28.

Cohen, Ariel & Manfred Krifka. 2014. Superlative quantifiers and meta-speech acts. *Linguistics and Philosophy* 37(1). 41–90. DOI: https://doi.org/10.1007/s10988-014-9144-x

Collins, Chris. 2016. *Not even. Natural Language Semantics* 24(4). 291–303. DOI: https://doi.org/10.1007/s11050-016-9124-5

Crnič, Luka. 2011. *Getting even*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Cummins, Chris, Uli Sauerland & Stephanie Solt. 2012. Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy* 35(2). 135–169. DOI: https://doi.org/10.1007/s10988-012-9114-0

Davies, Mark. 2008–. The Corpus of Contemporary American English (COCA): 560 million words, 1990–present. Available online at https://www.english-corpora.org/coca/.

Dayal, Veneeta. 1996. *Locality in WH quantification: Questions and relative clauses in Hindi,* (Studies in Linguistics and Philosophy). Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-4808-5

Drummond, Alex. 2013. Ibex farm. Retrieved from http://spellout.net/ibexfarm/.

Ducrot, Oswald. 1973. *La preuve et le dire*. Paris: Maison Mame.

Enguehard, Émile & Emmanuel Chemla. 2019. Connectedness as a constraint on exhaustification. In press in *Linguistics and Philosophy*.

Fox, Danny. 2018. Partition by exhaustification: Comments on Dayal 1996. In Uli Sauerland & Stephanie Solt (eds.), *Proceedings of Sinn und Bedeutung 22 (ZASPiL 60)* 1. 403–434.

Gärdenfors, Peter. 2004. *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/2076.001.0001

Gärdenfors, Peter. 2014. *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/9629.001.0001

Gazdar, Gerald. 1979. *Pragmatics: implicature, presupposition, and logical form*. New York: Academic Press.

Geurts, Bart. 2006. Take 'five': the meaning and use of a number word. In Svetlana Vogeleer & Liliane Tasmowski (eds.), *Non-definiteness and plurality,* 311–329. Amsterdam and Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/la.95.16geu

Geurts, Bart & Rick Nouwen. 2007. *At least* et al.: The semantics of scalar modifiers. *Language* 83(3). 533–559. DOI: https://doi.org/10.1353/lan.2007.0115

Givón, Talmy. 1978. Negation in language: Pragmatics, function, ontology. In Peter Cole (ed.), *Pragmatics* (Syntax and Semantics), 69–112. New York: Academic Press.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Groenendijk, Jeroen & Martin Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers*. Amsterdam: University of Amsterdam dissertation.

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for *know*. In Rhona Buchalla & Anita Mitwoch (eds.), *Proceedings of the Ninth Annual Conference of the Israeli Association for Theoretical Linguistics and the Workshop on Discourse* 1. 128–144.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*. Los Angeles, CA: University of California, Los Angeles dissertation.

Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.

Horn, Laurence R. 1992. The said and the unsaid. In Chris Barker & David Dowty (eds.), *Proceedings of the Second Conference on Semantics and Linguistic Theory (SALT II)*, 163–192. Columbus, OH: Ohio State University Linguistics Department. DOI: https://doi.org/10.3765/salt.v2i0.3039

Horn, Laurence R. 1996. Presupposition and implicature. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 299–319. Oxford: Basil Blackwell.

Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9(2). 105–129. DOI: https://doi.org/10.1080/15475441.2012.658731

Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology* 58(3). 376–415. DOI: https://doi.org/10.1016/j.cogpsych.2008.09.001

Jespersen, Otto. 1933. *Essentials of English grammar*. Tuscaloosa, AL: University of Alabama Press. Reprinted 1964.

Karttunen, Lauri & Stanley Peters. 1979. Conventional implicature. In Choon-Kyu Oh & David A. Dinneen (eds.), *Syntax and semantics 11: Presupposition*, 1–56. New York: Academic Press.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690. DOI: https://doi.org/10.1007/s10988-008-9029-y

Kennedy, Christopher. 2013. A scalar semantics for scalar readings of number words. In Ivano Caponigro & Carlo Cecchetto (eds.), *From grammar to meaning: The spontaneous logicality of language*, 172–200. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139519328.010

Kennedy, Christopher. 2015. A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8(10). 1–44. DOI: https://doi.org/10.3765/sp.8.10

Koenig, Jean-Pierre. 1991. Scalar predicates and negation: Punctual semantics and interval interpretations. In Lise M. Dobrin, Lynn Nichols & Rosa M. Rodriguez (eds.), *Proceedings of the 27th Meeting of the Chicago Linguistic Society, Part Two: The Parasession on Negation*, 130–144.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25(3–4). 1–49. DOI: https://doi.org/10.3765/salt.v4i0.2462

Lahiri, Utpal. 1998. Focus and negative polarity in Hindi. *Natural Language Semantics* 6(1). 57–123. DOI: https://doi.org/10.1023/A:1008211808250

Lenth, Russell V. 2016. Least-squares means: The R package lsmeans. *Journal of Statistical Software* 69(1). 1–33. DOI: https://doi.org/10.18637/jss.v069.i01

Levinson, Stephen C. 1983. *Pragmatics* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Marty, Paul, Emmanuel Chemla & Benjamin Spector. 2013. Interpreting numerals and scalar items under memory load. *Lingua* 133. 152–163. DOI: https://doi.org/10.1016/j.lingua.2013.03.006

Nouwen, Rick. 2010. Two kinds of modified numerals. *Semantics and Pragmatics* 3(3). 1–41. DOI: https://doi.org/10.3765/sp.3.3

Panizza, Daniele, Yi Ting Huang, Gennaro Chierchia & Jesse Snedeker. 2012. Relevance of polarity for the online interpretation of scalar terms. In Ed Cormany, Satoshi Ito &

David Lutz (eds.), *Proceedings of the 19th Semantics and Linguistic Theory Conference (SALT19)*, 360–378. DOI: https://doi.org/10.3765/salt.v19i0.2530

R Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In Jae-Hak Yoon & Andreas Kathol (eds.), *OSU Working Papers in Linguistics* 49. 91–136. DOI: https://doi.org/10.3765/sp.5.6

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. DOI: https://doi.org/10.14198/ELUA2008.22.06

Rodríguez, Raquel González. 2008. Sobre los modificadores de aproximación y precisión. *ELUA* 22. 111–128. DOI: https://doi.org/10.14198/ELUA2008.22.06

Rooth, Mats. 1985. *Association with focus*. Amherst, MA: University of Massachusetts dissertation.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75–116. DOI: https://doi.org/10.1007/BF02342617

Rullmann, Hotze. 1995. *Maximality in the semantics of wh-constructions*. Amherst, MA: University of Massachusetts dissertation.

Sadock, Jerrold. 1984. Whither radical pragmatics? In Deborah Schiffrin (ed.), *Georgetown University round table on languages and linguistics 1984*, 139–149. Washington, DC: Georgetown University Press.

Scharten, Rosemarijn. 1997. *Exhaustive interpretation: A discourse-semantic account*. Nijmegen: Radboud University dissertation.

Schwarz, Bernhard. 2005. Scalar additive particles in negative contexts. *Natural Language Semantics* 13(2). 125–168. DOI: https://doi.org/10.1007/s11050-004-2441-0

Solt, Stephanie. 2014. An alternative theory of imprecision. In Todd Snider, Sarah D'Antonio & Mia Weigand (eds.), *Proceedings of the 24th Semantics and Linguistic Theory Conference (SALT24)*, 514–533. DOI: https://doi.org/10.3765/salt.v24i0.2446

Solt, Stephanie. 2018. Approximators as a case study of attenuating polarity items. In Sherry Hucklebridge & Max Nelson (eds.), *NELS 48: Proceedings of the 48th Annual Meeting of the North East Linguistic Society* 3. 91–104. Amherst, MA: GLSA.

Spector, Benjamin. 2013. Bare numerals and scalar implicatures. *Language and Linguistics Compass* 7(5). 273–294. DOI: https://doi.org/10.1111/lnc3.12018

Spector, Benjamin. 2014. Global positive polarity items and obligatory exhaustivity. *Semantics and Pragmatics* 7(11). 1–61. DOI: https://doi.org/10.3765/sp.7.11

Tian, Ye, Heather Ferguson & Richard Breheny. 2016. Processing negation without context – why and when we represent the positive argument. *Language, Cognition and Neuroscience* 31(5). DOI: https://doi.org/10.1080/23273798.2016.1140214

Tottie, Gunnel. 1991. *Negation in English speech and writing: A study in variation*. San Diego: Academic Press.

van Kuppevelt, Jan. 1995. Discourse structure, topicality and questioning. *Journal of Linguistics* 31(1). 109–147. DOI: https://doi.org/10.1017/S002222670000058X

van Kuppevelt, Jan. 1996. Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 19(4). 393–443. DOI: https://doi.org/10.1007/BF00630897

van Os, Charles. 1989. *Aspekte der Intensivierung im Deutschen*, vol. 37. Tübingen: Gunter Narr Verlag.

van Rooij, Robert & Katrin Schulz. 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13. 491–519. DOI: https://doi.org/10.1007/s10849-004-2118-6

Wilkinson, Karina. 1996. The scope of *even*. *Natural Language Semantics* 4. 193–215. DOI: https://doi.org/10.1007/BF00372819