Granularity and scalar implicature in numerical expressions

OPEN ACCESS

RESEARCH ARTICLE

Granularity and scalar implicature in numerical expressions

Chris Cummins · Uli Sauerland · Stephanie Solt

Abstract   It has been generally assumed that certain categories of numerical expressions, such as 'more than n', 'at least n', and 'fewer than n', systematically fail to give rise to scalar implicatures in unembedded declarative contexts. Various proposals have been developed to explain this perceived absence. In this paper, we consider the relevance of scale granularity to scalar implicature, and make two novel predictions: first, that scalar implicatures are in fact available from these numerical expressions at the appropriate granularity level, and second, that these implicatures are attenuated if the numeral has been previously mentioned or is otherwise salient in the context. We present novel experimental data in support of both of these predictions, and discuss the implications of this for recent accounts of numerical quantifier usage.

1 Introduction

Modified numerals such as 'more than nine' or 'at least ten' typically seem to convey the impression that the speaker's knowledge about the topic under

C. Cummins (✉)
SFB 673, Alignment in Communication, Universität Bielefeld, Bielefeld, Germany
e-mail: c.r.cummins@gmail.com

U. Sauerland · S. Solt
Zentrum für Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany

S. Solt
Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam, Amsterdam, The Netherlands

discussion is imprecise. By contrast, unmodified numerals such as 'ten' convey precise knowledge. However, on closer inspection, this distinction is not as clear-cut as might be supposed: in particular, modified numerals appear to convey more information than their semantic analyses would suggest. This paper aims to characterise the meaning of these expressions more rigorously, and account for it in terms of semantic and pragmatic theory.

The standard account of unmodified and modified numeral meaning claims that unmodified numerals have an upper bound to their interpretation, while modified numerals such as 'more than n' do not. So, while 'ten' would establish that the cardinality in question is not greater than 10, 'more than nine' and 'at least ten' would not establish such an upper bound on the cardinality in question. Theories differ on whether the upper bound for the unmodified numerals is part of semantics (Breheny 2008 and others) or derived as a pragmatic implicature (Horn 1972 and others). But the claim that modified numerals have no such bound is generally agreed on in the literature: Horn (1972, 1984) endorses this implicitly by referring to the interpretation of a numeral without an upper limit as the 'at least n' interpretation. More recently Krifka (1999) and Fox and Hackl (2006) have explicitly claimed that modified numerals with 'more than' or 'at least', when used in unembedded declarative contexts, generally do not give rise to an implicated upper bound, and provided theoretical accounts that encompass this observation. Specifically, both note a conflict between the theory of scalar implicatures and the lack of an upper bound for these modified numerals. We consider the relevant aspects of their accounts in the following section, and then return to the empirical issue of whether modified numerals establish pragmatic bounds on their interpretation. In Sect. 3, we present a proposal for the interpretation of modified numerals based on scale granularity and salience that predicts an implicated bound in most cases when a modified numeral is used, and specifically one that depends upon granularity level. In Sect. 4, we present experimental evidence in favour of our proposal. We conclude by discussing the implications of these findings for Fox and Hackl's account, and consider the conditions under which quantifiers of this type may felicitously be used.

2 Previous accounts

In this section we present two previous accounts of modified numerals, both of which predict that modified numerals should not generally give rise to pragmatically-communicated bounds on the cardinality they report. Before doing so, we must make some preliminary remarks about the nature of the pragmatic enrichment under discussion, namely scalar implicature.

Scalar implicatures have been much debated in recent years (Chierchia 2004; Chierchia et al. 2008; Geurts 2010, and others; see Sauerland 2012 for a summary). However, the discussion in this paper generally does not depend on the specifics of the approach adopted. Therefore we frame our proposal within a classical framework close to that of Horn (1972), but note that this could be adapted to any of the current approaches. We assume that certain words (such as 'some', 'all',

'or', etc.) are associated with lexical scales (such as ⟨*some, all*⟩ and ⟨*or, and*⟩), which we shall refer to as Horn scales. We further assume that unembedded occurrences of scalar terms can trigger a scalar implicature. That is, if scalar term A has a stronger scale-mate B, such that if replacing A with B in the current sentence S creates a sentence S* that unidirectionally entails S, the speaker of S implicates that S* is false. For example, a speaker of (1) implicates the falsity of (2), where these sentences are linked by the substitution of the scale-mates 'some' and 'all'.

(1) Some member states of the EU are monarchies.

(2) All member states of the EU are monarchies.

In order for this implicature to be conveyed it is also necessary for the speaker of (1) to be epistemically committed—that is, the speaker must be presumed to be knowledgeable about the truth or falsity of the stronger proposition expressed by (2), and to be willing to communicate that knowledge. In this paper we focus on cases where this condition is satisfied.

Krifka (1999) and Fox and Hackl (2006) observe that certain categories of modified numeral systematically fail to admit scalar implicature. Krifka (1999) discusses numerals modified by 'at least', noting that their apparent behaviour is at variance with the predictions of Horn's (1972) account of scalar implicature. Specifically, he observes that (3) does not give rise to the implicature that (4) is false. If it did so, a knowledgeable speaker who uttered (3) would be pragmatically understood to imply that John has exactly three children, as this follows from the truth of (3) coupled with the falsity of (4). This is clearly intuitively incorrect (and is shown by Geurts et al. (2010, p. 138) to be uniformly rejected by experimental participants).

(3) John has at least three children.

(4) John has at least four children.

In attempting to account for (3), Krifka considers the possibility that modified numerals of the type 'at least *n*' do not participate in Horn scales, while bare numerals do. Although this solves the problem, it appears to be an arbitrary and unprincipled distinction: the standard requirements for Horn scale membership are that the terms should be equally lexicalised, from the same semantic field, and in the same register. This appears to apply just as much to the scale ⟨*at least three, at least four*⟩ as to the scale ⟨*three, four*⟩. Krifka (1999, p. 259) goes further, asserting that "[i]f number words form Horn scales, they should do so in any context in which they appear".

Krifka (1999, p. 259) also discusses the idea that 'at least' signals the speaker's unwillingness or inability to give a precise answer. This precludes a scalar implicature from arising, as the epistemic condition on the speaker is not met. Krifka's suggestion is that the notion of the speaker's uncertainty, or reticence, is pragmatically derived from the choice of 'at least *n*' rather than the bare numeral *n*, because the latter would carry the implicature of certainty ('exactly *n*').

This analysis, however, does not appear to generalise to the equally problematic case of 'more than'. Fox and Hackl (2006, p. 540) observe that 'more than n' also typically does not give rise to scalar implicature in an unembedded context.[1] The relevant observation is that, for example, uttering (5) does not trigger the scalar implicature that John has exactly four children, even though comparison with (6) should yield exactly this implicature.

(5) John has more than three children.

(6) John has more than four children.

Nevertheless, it is intuitively plausible that an informed and cooperative speaker could say something like (5), for example in the context of establishing whether John is eligible for certain benefits, or whether he needs a bigger car, etc. Under such circumstances, the implicature that John has exactly four children does indeed seem not to arise.

Moreover, if we turn our attention to larger numbers, we can observe clear examples of implicatures failing to arise from expressions using 'more than' and 'fewer than'. Consider, for instance, (7) and (8). These utterances do not convey the corresponding implicatures (9) and (10). In fact, they do not even convey the weak implicatures (9′) and (10′), which demonstrates that the implicature failure cannot be attributed to the hearer doubting the epistemic commitment of the speaker. An informed speaker seems perfectly entitled to utter (7) without being committed to the rather unlikely proposition found in (9′).

(7) More than 100 people got married today.

(8) Fewer than 20 people have walked on the Moon.

(9) Exactly 101 people got married today.

(10) Exactly 19 people have walked on the Moon.

(9′) The speaker considers it possible that exactly 101 people got married today.

(10′) The speaker considers it possible that exactly 19 people have walked on the Moon.

The need to account for this apparently anomalous behaviour of modified numerals has been a partial motivation for various semantic accounts of numerical quantifiers. For Krifka (1999), this and other factors motivate a rejection of the generalised

---

[1] Fox and Hackl (2006, p. 541, footnote 7) observe that an implicature is available if the context specifies a set of 'relevant alternatives'. In particular, they note that if a speaker is prompted with a choice of responses 'more than 10', 'more than 20', 'more than 30', 'more than 40', then the subsequent use of 'more than 20' yields an implicature that 'not more than 30'. However, they contrast this with the use of 'more than 20' out of the blue, which they contend does not give rise to this kind of enrichment. It is this latter claim that we contest here.

quantifier account of Barwise and Cooper (1981) for expressions such as 'at least *n*'. Krifka's account of the way in which scales are built up in such cases feeds into Geurts and Nouwen's (2007) proposal concerning the modal semantics of superlative quantifiers. Fox and Hackl (2006) attribute this absence of implicature to the failure of a covert exhaustivity operator, and use the facts from this domain as one of the motivations for their proposal of the Universal Density of Measurement.[2] They also report a different pattern of behaviour for modified numerals embedded below modal verbs and other quantifiers, as we will discuss briefly in Sect. 5.

However, despite the absence of implicatures such as (9) and (10), there do appear to be some pragmatic enrichments available from utterances using 'more/ fewer than *n*', and perhaps 'at least/most *n*', that have not been satisfactorily accounted for. Consider an utterance such as (11).

(11) John's birthplace has more than 1000 inhabitants.

If John was born in London, this statement is semantically true but intuitively misleading, because 'more than 1000' seems to convey a quantity that is appreciably less than the actual population of London. However, its precise communicative value is difficult to determine introspectively. Certainly it does not seem to convey the classical scalar implicature (12), and in this sense it patterns with (5), (7) and (8). However, it does seem to convey an implicature like (13), if not a stronger one.

(12) John's birthplace does not have more than 1001 inhabitants.

(13) John's birthplace does not have more than a million inhabitants.

In this paper, we attempt to explain this apparent divergence of behaviour. We argue that 'more than *n*' and 'fewer than *n*' do in fact give rise to scalar implicatures, and that the same is true of superlative quantifiers. However, we propose that these implicatures are restricted by the granularity of the numerical scale and by considerations of contextual salience of the numeral.

Our argument stems from the observation that the potential implicature of a sentence depends on the scalar alternatives of the words in the sentence. For instance, if 'most' also inhabited the ⟨*some*, *all*⟩ scale, (1) would additionally implicate the falsity of (14).

(14) Most member states of the EU are monarchies.

In this paper, we argue that the question of which expressions constitute *bona fide* scalar alternatives to forms such as 'more than 100' is more subtle than has previously been assumed. Specifically, in the following section, we develop the

---

[2] Our proposal is compatible with the specific details of Fox and Hackl's analysis but does not depend on them. For instance, Mayr (in progress) develops an account of the core facts that Fox and Hackl seek to account for, but which extends to modified numerals of the "at least/most *n*" type and does not assume the universal density of measurement. As far as we can see, Mayr's analysis is fully compatible with ours.

proposal that scale membership in the numerical domain is determined by granularity considerations. Thus, the failure of 'more than 100' to implicate 'not more than 101', and similar examples, can be accounted for from first principles and without globally prohibiting the participation of these terms in Horn scales. This coheres with existing observations but makes the novel prediction that certain specific scalar implicatures will be available from modified numerals of this type. We further argue on the grounds of relevance that the prior mention of a numeral may license its reuse in a modified expression, and thus that this category of implicatures should be attenuated in such cases. In Sect. 4, we present novel experimental evidence in support of both these claims. In Sect. 5, we discuss the implications of these findings in determining the use and interpretation of these classes of modified numerals.

3 Granularity and salience

3.1 Granularity

As discussed by Krifka (2002), there is a sense in which both (15) and (16) are simultaneously true.

(15)   The distance from Amsterdam to Vienna is 1000 km.

(16)   The distance from Amsterdam to Vienna is 965 km.

What distinguishes these two sentences is the level of granularity they assume, and hence the level of precision that they are taken to convey. Here granularity can be understood as the density of representation points on a measurement scale. Thus (15) plausibly involves a scale whose points are multiples of 100 km (…800 km… 900 km…1000 km…1100 km….), and thus conveys that the distance between Amsterdam and Vienna is 1000 ± 50 km. But (16) implies a finer-grained scale, perhaps one with units of 1 km, in which case it expresses that the distance between the two cities is 965 ± 0.5 km.

   In the case of cardinal quantities, as in examples (1)–(13) above, the finest-grained scale that makes sense is that in which the representation points coincide with whole numbers: that is, the scale points are one apart. However, as discussed by Krifka (2009), drawing upon work by Curtin (1995), coarser-grained scales are typically also available. Krifka notes that in Western culture, and in particular in languages with a decimal number system, the most frequently used granularity levels are those based on powers of 10. Also possible are scales constructed from these by operations of halving (resulting in the optimal refinement of a scale) and doubling.

   Thus in reporting the results of counting entities, available scales appear to include the following (not intended to be an exhaustive list):

(17) a.   1…2…3…4…5…
      b.   10…20…30…40…
      c.   5…10…15…20…   [derived from (17b) via halving]
      d.   20…40…60…80…   [derived from (17b) via doubling]
      e.   100…200…300…400…
      f.   50…100…150…200…   [derived from (17e) via halving]
      etc.

The points on the coarser-grained scales in (17) are the numbers that we intuitively think of as more 'round': 100, for example, seems rounder than 90, which in turn seems rounder than 93 (see Jansen and Pollmann (2001) for a definition of roundness based on divisibility by powers of 10, 2 and 5). More formally, the representation points on each of these scales are distributed equidistantly. Such a distribution can be accounted for in terms of the amount of information that can be conveyed by choosing one of a limited number of expressions: that is, we can see the distribution of representation points as an attempt to partition up a scale in the most efficient way.

As Krifka remarks, scales with other granularity levels are also in operation in specific domains: notably, in the case of time, one granularity level is based on 15 min intervals, another on 3- or 6-month intervals, and so on. Logarithmic granularity may also be an option in some domains (e.g. 10…100…1000… 10,000…), such as for example the size of populations or settlements.

The use of a coarser granularity level may be favoured on several grounds. For one, the expressions referring to points on scales of coarse granularity (e.g. 'ten', 'one hundred', 'one thousand') are typically shorter than those referring to neighbouring points on scales of finer granularity (e.g. 'eleven', 'one hundred and one', 'one thousand and one'), and therefore should be favoured both in production and comprehension, although this is in principle an issue that can be distinguished from granularity per se. Krifka (2002) notes also that values corresponding to points on coarser-grained scales, in being interpreted less precisely, may be useful for face-saving: if one speaks in approximate terms, one is not committed to such high precision and is less likely to be inadvertently misleading. More generally, it could be argued that expressions occurring on scales of coarse granularity are privileged in terms of availability or default activation level, and therefore can be used at a lower cognitive cost. Evidence for the existence of a processing advantage of this sort is provided by Mason et al. (1996), who show that round 5-digit numbers produced as a result of a mathematical operation are recalled more accurately than non-round numbers, even when the task is to recall the first two digits.

Against these advantages for coarse granularity, there is some loss of precision in the information communicated, and indeed an ambiguity may arise as to the intended granularity level. From a relevance standpoint (Sperber and Wilson 1986), all this means that the use of an expression belonging to a coarser-grained scale results in a saving of cognitive effort, at the cost of a loss of cognitive effect. The prediction would be that the coarse-grained scale is used in circumstances where the saving in effort outweighs the loss of effect, typically because the expression on the coarse-grained scale is much more readily available than alternatives on a

finer-grained scale, and/or because more precise information is not sufficiently useful to the hearer to justify the additional expense involved in encoding and decoding it.

Bearing considerations of granularity in mind, we can look again at expressions of the form 'more than n' and similar, and derive new predictions concerning their pragmatic interpretation. Consider in particular the classical scalar implicature that we have argued does not arise from (11), repeated below for convenience.

(11)   John's birthplace has more than 1000 inhabitants.

From the lack of an implicature to 'not more than 1001', we concluded that the sub-scale of alternatives ⟨*more than 1000*, *more than 1001*⟩ was in some way failing to function. Granularity gives us a principled way of accounting for this. Given that 1001 is not a scale point on the numeral scales of granularity 10, 100 or 1000, while 1000 is, 'more than 1000' and 'more than 1001' are not comparable. Specifically, we would argue that the use of 1001 comes at an additional cognitive cost, which counterbalances the additional cognitive effect that its use conveys (in this case, ruling out the possibility that John's birthplace has exactly 1001 inhabitants). Therefore, the use of 'more than 1000' rather than 'more than 1001' is supported by considerations of relevance, irrespective of whether the speaker knows that the latter holds. Hence, the implicature that 'more than 1001' does not hold—and therefore that exactly 1001 is the case—does not arise.

By contrast, if we consider a scale containing numerals which occur on a scale of the same granularity level, this argument fails. For instance, ⟨*more than 70*, *more than 80*⟩ actually does appear to constitute part of a Horn scale. Contrasting utterances (18) and (19), we note that the numerals used in both correspond to scale points on the numeral scale of granularity 10, and both fail to correspond to scale points on the numeral scale of granularity 100. Hence, from a granularity standpoint, the numerals are matched in level. In addition, (19) is more informative than (18). Therefore, if the speaker is in a position to say (19) instead of (18), relevance considerations mandate that they should do so. Hence we predict that (18) does typically give rise to an implicature that (19) does not hold.

(18)   More than 70 people got married today.

(19)   More than 80 people got married today.

Similarly, while ⟨*more than 1000*, *more than 1001*⟩ does not seem to function as a set of alternatives, ⟨*more than 1000*, *more than 2000*⟩ and ⟨*more than 1000*, *more than 1,000,000*⟩ do seem legitimate options, the former corresponding to a scale with units of 1000, the second to a logarithmic scale, or perhaps one based on the lexicalized values 'thousand' and 'million'. Thus while the utterance of (11) does not give rise to the implicature 'not more than 1001', it does allow inferences such as 'not more than 1 million' or even 'not more than 2000', resulting in the pragmatic infelicity discussed above.

It is not strictly necessary, on this account, that the numerals are precisely matched as to the granularity levels on which they are scale points. In fact, if the more informative statement refers to a scale point on a coarser-grained scale, it will arguably be further favoured by its lower cognitive cost. Thus we predict that (18)–(20) will all give rise to an implicature that (21) does not hold, in that 100 is a point on a coarser-grained scale than 70, 80 or 90.

(20) More than 90 people got married today.

(21) More than 100 people got married today.

Following the analysis of Krifka (2009), who asserts that scales of different granularity levels should align, we note that 100 is a scale point for several granularity levels simultaneously, including 100s, 50s, 10s and units. It follows that, on our account, the implicature arising from the utterance of (21) would be in some sense under-determined. Our prediction would be that the possible pragmatic enrichments of (21) vary according to the granularity level at which the numeral is interpreted. The applicable scalar implicature might be that 'more than 200' does not hold, that 'more than 150' does not hold, that 'more than 110' does not hold, or even that 'more than 101' does not hold. We would further predict that numbers that are scale points for finer-grained scales should give rise to stronger, more informative implicatures, i.e. provide a tight upper bound. To enable this to occur, we must assume that the granularity level is itself pragmatically determined by the hearer, and that it guides the hearer's interpretation as regards implicature.

In sum, our first prediction is that expressions of the form 'more than n' (and, *mutatis mutandis*, 'fewer/less than n') do give rise to scalar implicatures, when granularity considerations are taken into account. Under the standard conditions of speaker informativeness and cooperativity, and in simple declarative contexts, 'more than n' gives rise to the implicature than 'more than m' does not hold, where m is any numeral such that $m > n$ and the coarsest granularity level expressed by m is at least as coarse as than that expressed by n. We validate this prediction empirically in Experiment 1 of this paper.

3.2 Contextual salience

In the discussion earlier, we noted that an utterance such as (22) is apparently infelicitous, given our world knowledge. However, this infelicity seems to be mitigated considerably by the presence of a preceding context such as (23).

(22) London has more than 1000 inhabitants.

(23) Give an example of a settlement with more than 1000 inhabitants.

Given our first prediction, articulated above, how are we to address examples such as this? Without the preceding context, (22) seems to convey an implicature to the effect that London's population is of the same order of magnitude as 1000 (and

certainly fewer than a million). However, this implicature is attenuated by the preceding context, in which the number 1000 is introduced.

By appeal to the notion of relevance, defined in terms of the minimisation of cognitive effort and the maximisation of cognitive effect, we can reconcile this finding with our first prediction. The crucial point here is that the prior mention of the numeral licenses its second mention, because the reuse of the numeral saves cognitive effort for the discourse participants. Without a preceding context of this kind, a hearer could legitimately respond to (22) by drawing a granularity-based inference through the process discussed earlier. However, in the presence of such a context, this reasoning falls down. The hearer's reasoning is instead as follows: the speaker uttered (22), rather than making a statement with a more informative quantifier such as 'more than a million'. Therefore, either the speaker is not in a position to make the stronger statement, or it would not be relevant to do so. However, given the preceding context, (22) is plausibly more relevant than the more informative alternative: it reuses a number that has already been activated, potentially saving the hearer effort,[3] and it answers the question raised in that context, thus achieving the necessary effect. Hence the speaker has grounds for choosing to utter (22) in preference to the stronger alternative, even if the speaker knows the latter is true, and therefore the implicature cannot go through.

Note that this argument does not require that the speaker is obliged to use the quantifier 'more than 1000' in answering (23). In fact, if the speaker were to say 'more than a million', it seems clear that this satisfactorily answers the question while preserving semantic truth. However, the hearer's awareness of the possibility that the speaker might simply be choosing to repeat the numeral in the question, without intending any pragmatic effects, is sufficient to block the implicature arising from (22). The hearer who is alert to this possibility will therefore not draw the scalar inference, and will not find (22) pragmatically anomalous in this context.

By contrast, a speaker who says (24) in answer to (23) appears to be guilty of a pragmatic violation. The speaker's failure to reuse the number in the question has the effect of licensing a scalar implicature of the type we have been discussing. As the scalar implicature is false, the utterance of (24) is infelicitous as an answer to (23), while being perfectly felicitous as an answer to a parallel question ("Give an example of a settlement with more than 7000 inhabitants").

(24) London has more than 7000 inhabitants.

Crucial to this analysis is the idea that the preceding context can license the (re)use of a specific numeral. In principle, the existence of this licensing context should attenuate the scalar implicature arising from the use of that specific numeral. The logic here is clear: scalar implicatures are conveyed by what we choose not to say. If there is no choice, because the preceding context specifies which numeral should be used in the continuation of the dialogue, the use of that numeral cannot give rise to

---

[3] As a anonymous reviewer pointed out, the relevance-theoretic approach presumes that the speaker acts in order to minimise the effort required by the hearer, rather than minimising the speaker's own effort. Hence, in RT terms, we argue that reuse of a numeral reduces the hearer's effort. In practice, the speaker's interests must also be considered (Wilson and Sperber 2002, p. 257), and we presume that reuse of a numeral also serves these interests for the same reason.

an implicature. In practice, there is always some choice, and a weaker argument goes through: if the speaker might not have deliberately chosen to use a particular number, then they might not have been attempting to convey the corresponding implicature. The hearer, attempting to reconstruct the speaker's intention, cannot then rely upon the implicature being intended.

In what way can the preceding context specify that a particular numeral should be used? To answer this question fully would require a general theory of contextual salience, and we do not attempt to offer such a proposal here. However, we posit that one specific way in which this can take place is if a numeral is mentioned in the preceding context. We would then expect that this numeral is heightened in availability and thus preferred for reuse. This reuse would then, according to our theory, give rise to weaker implicatures than would the use of the same numeral in a situation where there was no prior mention of it. In the former case, we propose that the prior mention would constrain the speaker's choice of numeral, and thus attenuate the pragmatic significance of the selected utterance.

Hence, we make the following additional prediction. Scalar implicatures of the type we predicted earlier, arising from expressions of the form 'more than n' and 'less/fewer than n' at the appropriate granularity level, will be attenuated by contextual prior mention of the numeral n, compared to cases in which n has not previously been introduced in the context. We validate this prediction empirically in Experiments 2 and 3 of this paper.

4 Experimental evidence

In this section, we report on a series of three experiments designed to test the two predictions discussed in Sect. 3:

(1) Modified numerals of the form 'more than n' give rise to scalar implicatures that are conditioned by granularity
(2) Granularity-based implicatures will be attenuated by mention of the numeral in the prior context.

4.1 Experiment 1: Range of interpretation of numerical quantifiers

Experiment 1 was designed to test the first of these two predictions. In this experiment, we presented participants with a numerically quantified expression and asked them to estimate the value or range of values they interpreted the expression as conveying. We manipulated the roundness of the numeral (corresponding to the granularity of the scale(s) on which it occurs) and the type of numerical quantifier, as discussed below.

4.1.1 Method

4.1.1.1 Participants The experiment was carried out online via the Amazon Mechanical Turk (MTurk) platform (see Sprouse 2011 for a discussion of the use of

MTurk for the collection of acceptability judgements). A total of 1200 participants were recruited (100 per condition, as described below). The only inclusion criterion was an acceptance rate of over 95 % on prior MTurk tasks. The subjects themselves remained anonymous to us, but as part of the study they reported a few demographic facts about themselves. The gender split was 51 % female, 49 % male. Subjects were paid $0.02 for participation.

4.1.1.2 Materials and procedure   Participants were shown the following stimuli, consisting of a statement including a modified numeral, and were asked to provide an estimate of the number in question:

Information A newspaper reported the following.

"[Numerical expression] people attended the public meeting about the new highway construction project."

Question Based on reading this, how many people do you think attended the meeting?

Between _____ and _____ people attended [range condition].
_____ people attended [single number condition].

Participants were also given an opportunity to write a comment explaining why they answered the way they did.

These materials were used across 12 conditions in which the following parameters were crossed:

- 2 quantifiers:
  - more than n
  - at least n
- 3 levels of roundness of n, corresponding to 3 levels of scale granularity:
  - Coarse granularity: multiple of 100 ($n = 100$)
  - Medium granularity: multiple of 10/non multiple of 100 ($n = 110$)
  - Fine granularity: non-round ($n = 93$)
- 2 question formats:
  - range
  - single number

The 12 conditions were fielded online over the course of roughly 20 days in December 2009 and January 2010, each condition being fielded separately in order to reduce the likelihood of individual participants completing multiple versions of the task.

4.1.1.3 Predictions   If participants draw inferences based on the writer's decision not to use a more informative expression containing a numeral m that participates on a scale of the same (or coarser) granularity than n, we expect to see the following:

- The rounder n is, and thus the coarser-grained the scale is on which it occurs, the higher respondents' estimates will be (relative to n)

- In the range condition, typical estimates will be of the form '$n + 1$ to m' (for 'more than n') or 'n to m' (for 'at least n'), where m is the next higher value on a scale on which n occurs (or a coarser-grained scale).

4.1.2 Results

Incomplete responses were eliminated, as were those that did not consist of a single numeral, including non-numerical responses (e.g. 'many' or 'infinity') as well as those expressing ranges (e.g. 'more than 110'); see Appendix 1 for the number of responses excluded in each condition.

Because the data were not normally distributed, and in particular there were a small number of outliers (see Appendix 1), in this and the following experiments we used non-parametric methods for our analyses.

For each condition, median values were calculated, and the most frequent responses were tallied. Additionally, to facilitate comparison across conditions, median values were restated in terms of numerical 'distance' from the numeral n in the quantifier presented. For example, a response of 140 in the 'more than 100' condition represents a distance of 40 from n. Full results are presented in Tables 1 and 2. Figure 1a, b shows graphically the median estimates and quartiles, expressed in terms of distance from n.

To evaluate the effect of granularity, a Kruskal–Wallis test was used, with distance from n as the independent variable and granularity level (coarse, medium, fine) as predictor. For the quantifier 'more than n' in the range condition, a significant effect of granularity was found for the high estimate ($p < 0.001$); follow-up pairwise comparisons via a Mann–Whitney U test showed significant differences between coarse and fine granularity, coarse and medium, and medium and fine (all $p < 0.001$). In the single number condition, a significant effect of granularity was also observed ($p < 0.001$), with follow-up pairwise comparisons showing a significant difference between coarse and fine ($p < 0.001$), coarse and medium ($p < 0.001$) and medium and fine ($p < 0.05$).

For the quantifier 'at least n' in the range condition, a significant effect of granularity was again found for the high estimate ($p < 0.001$); follow-up pairwise comparisons showed significant differences between coarse and fine granularity ($p < 0.001$) and between coarse and medium granularity ($p < 0.05$); no significant difference was found between medium and fine granularity. In the single number condition, a significant effect of granularity was also observed ($p < 0.05$), with follow-up pairwise tests showing a significant difference between coarse and fine and coarse and medium (both $p < 0.01$) but no significant difference between medium and fine.

Note that the preceding analyses included responses that were inconsistent with the truth conditions of the stimuli sentence. This includes responses such as '60–80' in the 'more than 100' range condition, where it seems the participant has misunderstood the nature of the task, and responses of n in the 'more than n' single number condition (see Appendix 1 for the number of such responses in each condition). When truth-conditionally inconsistent responses are removed, a

Table 1 Number estimated—median (MAD[a])

	More than n			At least n		
	Range condition		Single # condition	Range condition		Single # condition
	Low	High		Low	High	
Coarse: $n = 100$	100 (1.5)	149 (35.6)	110 (14.8)	100 (0.0)	125 (37.1)	101 (5.9)
Medium: $n = 110$	110 (1.5)	127.5 (18.5)	112 (3.0)	110 (0.0)	125 (22.2)	110 (3.0)
Fine: $n = 93$	93 (1.5)	100 (7.4)	94 (1.5)	93 (0.0)	100 (7.4)	93 (3.0)

[a] MAD is the (scaled) median absolute deviation, a measure of dispersion, calculated as the median of the absolute deviations from the median, multiplied by a standard scale factor of 1.4826 for approximation to the standard deviation (Hampel 1974)

Table 2 Most frequent estimates—range condition (# of participants)

	More than n		At least n	
	Low	High	Low	High
Coarse: $n = 100$	100 (40)	150 (24)	100 (63)	150 (18)
	101 (28)	125 (12)	*90 (7)*	125 (15)
		120 (8)	*50 (6)*	120 (9)
		200 (7)		100 (9)
		1000 (5)		200 (5)
Medium: $n = 110$	110 (46)	120 (28)	110 (55)	150 (19)
	111 (31)	150 (24)	*100 (14)*	120 (13)
	100 (9)	200 (7)		130 (12)
		119 (6)		110 (9)
		130 (5)		115 (7)
		115 (5)		200 (7)
				125 (6)
				111 (6)
Fine: $n = 93$	93 (32)	100 (34)	93 (57)	100 (42)
	94 (29)	95 (14)	*90 (16)*	150 (9)
	90 (14)	99 (7)	*50 (5)*	95 (8)
		150 (6)		120 (6)
		125 (5)		93 (6)

Truth-conditionally inconsistent responses in italics

Kruskal–Wallis test shows a significant effect of granularity for the high estimate in the range conditions, for both 'more than n' and 'at least n' ($p < 0.001$). In the single number condition, a significant effect of granularity is found for 'more than n' ($p < 0.001$), but the corresponding effect for 'at least n' falls short of significance ($p = 0.083$).

Turning now to the distribution of individual responses in the range conditions (Table 2), for both 'more than n' and 'at least n', respondents' estimates for the
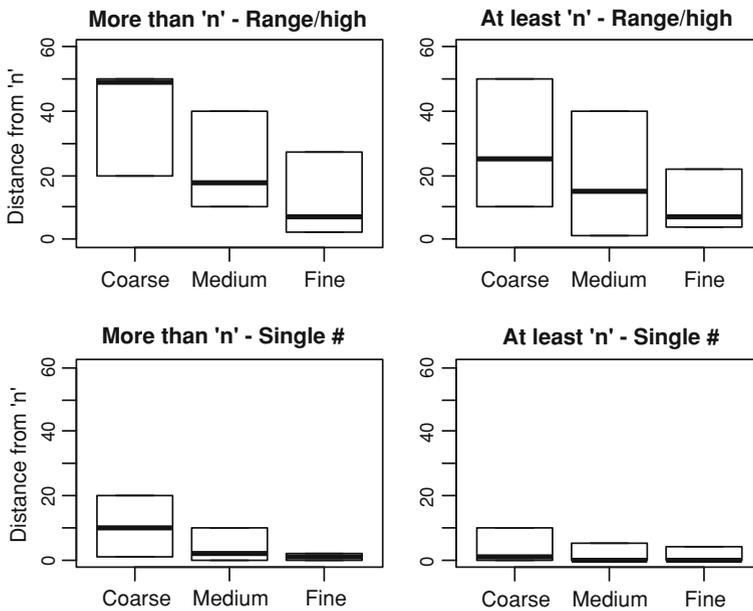
Fig. 1 Number estimated—median and quartiles

lower end of the range were most typically *n* or *n* + 1.[4] Most of the remaining responses fall into the 'truth-conditionally inconsistent' category; only a small proportion of subjects' responses were higher than *n* + 1 (between 5 and 14 per condition).

With regard to the upper end of the range estimated, in the case of 'more than 100'/'at least 100', the most common values given were 150, 125, 120 and 200; these correspond to the next higher point above 100 on scales of granularity 50, 25, 20 and 100, respectively. In the case of 'more than 110'/'at least 110', the most common high estimates were 120 and 150, corresponding to the next points on scales of granularity 10 (on which 110 occurs) and 50 (a coarser grained scale). Finally, in the case of 'more than 93'/'at least 93', the upper end given for the range was most commonly 100, corresponding to the next higher point on scales of granularity 10, 20, and higher.

4.1.3 Discussion

The results of Experiment 1 demonstrate that scale granularity plays a role in the interpretation of numerically quantified expressions of the form 'more than *n*' and 'at least *n*'. The rounder *n* is, and thus the coarser the granularity of the scale(s) on which it occurs, the wider the range of interpretation that subjects allow for 'more than *n*' and 'at least *n*'. Looking just at the median responses, we see that the range

---

[4] Here, we assume that a response of the form 'between *n* and *m*' in the 'more than *n*' condition reflects an exclusive rather than inclusive interpretation of *between*.

of interpretation for 'more than 100' extends on average to 149 (i.e. 49 higher than 100), while that for 'more than 110' on average extends to 127.5 (only 17.5 higher than 110, and actually lower than the upper limit inferred for 'more than 100'), and that for 'more than 93' typically does not extend past 100. Similar patterns are observed in the case of 'at least n', though in this case the differences are less pronounced.

These patterns can be interpreted as arising from scalar implicatures based on granularity. In completing the task, respondents appear to have relied on the inference that a decision not to refer to a higher value on a scale of the same or a coarser level of granularity is an indication that (as far as the speaker/writer knows) the statement with that higher value does not hold. That is, when presented with 'more than n', the hearer typically computes the implicature that 'not more than m', where m is the next higher point on a scale on which n occurs (or on a scale of coarser granularity). The coarser-grained the scale, the more 'distant' m is from n, and thus the higher respondents' estimates relative to n.

The distribution of individual responses further supports this hypothesis. The majority of responses in the range condition were of the form '$n/n + 1$ to m', where m was, as discussed above, the immediately higher point on a scale on which n occurs, or a coarser scale. Thus 'more than 100' typically conveys 'not more than 150', 'more than 110' conveys 'not more than 120' or 'not more than 150', and 'more than 93' conveys 'not more than 100'. While respondents were not consistent as to which scale they used as a basis for inference, their responses overall reflect an influence of granularity.

Additional insight into the reasoning processes involved comes from the responses to the optional comments question. Some, though not all, participants commented on the reasoning behind the answers they provided. In the 'more than 100' condition, responses included "I feel that if there was more than 150, the newspaper would say more than 150", "I chose the above number because I felt had the numbers been higher the paper would have said more than 200", and "I think 125 would be the next increment worthy of mentioning". Similarly, in the 'more than 110' condition, we received responses such as "if it was > 120 it would have been described as such" or "if there were 116 people I assume the writer would've said more than 115." In total, 60 participants gave 'reason' responses, and of those 25 explicitly mentioned an alternative value that the writer could have used but did not. These responses all indicate reasoning based on alternatives, and furthermore reflect awareness of granularity considerations.

While the discussion so far has focused primarily on 'more than n', the results for 'at least n' further corroborate the account we have proposed. The primary difference between the 'at least n' and 'more than n' conditions seems to be that the estimates are on average higher for 'more than n' than they are for 'at least n'. This difference is potentially due to the fact that the use of 'at least n' indicates that the value '(exactly) n' has some non-zero probability of being true (Geurts and Nouwen 2007; Cummins and Katsos 2010; see Nouwen 2008 for a similar analysis of 'no more than n'). However, the general trend appears to be that 'at least n' does not differ substantively from 'more than n'. In the following, we can therefore focus on 'more than n'.

In the case of 'more than 93', we might further ask why participants do not draw the inference that 'exactly 94' is the case, given that 94, like 93, occurs on the scale of unit granularity. Rather, the upper bound is most commonly the next-largest numeral with a coarser granularity level (i.e. 100), rather than at the same level as 93 itself. This possibility itself is not inconsistent with our prediction for granularity effects (which holds that 'more than n' implicates that 'not more than m' does not hold, where m is a numeral that occurs on a scale of the same <u>or coarser</u> granularity than n). However, the absence of unit-granularity implicatures is at variance with our initial hypothesis. We think that this is related to Fox and Hackl's (2006) observation that the enriched interpretation obtained at the finest level of granularity would be futile as it would make 'more than 93' synonymous with '94' (at least with respect to cardinalities). We return to this topic in the general discussion, where we go on to consider the circumstances under which quantifiers such as 'more than 93' may felicitously be employed.

Broadly speaking, however, Experiment 1 supports our first hypothesis. According to these findings, scalar implicatures are available from expressions such as 'more than n', and these implicatures are constrained by granularity. In this respect, our findings cohere with the developing body of research on the role of granularity in quantification.

4.2 Experiment 2: Effect of contextual priming on interpretation of numerical quantifiers

In Experiment 2, we wish to test the second of our two hypotheses, namely that the implicatures due to granularity are attenuated by previous mention of a numeral. Given that Experiment 1 used a methodology which precludes close control of the participants' linguistic and cognitive abilities, we also wish to replicate our findings in a more controlled setting. We further seek to extend the findings to the monotone decreasing quantifier 'fewer than n'. We achieve this by presenting quantified expressions in two contexts, one in which the numeral is previously mentioned ('primed') in the discourse and one in which it is not, and ask our participants to interpret these expressions. We are then able to examine both whether granularity-based implicatures are again exhibited, and, if so, whether prior mention of the numeral influences these implicatures.

4.2.1 Method

4.2.1.1 Participants 45 native adult English-speaking participants were recruited and randomly allocated to one of the versions of the task described below. Their average age was 21.3 years (SD 5.5 years, range 17–46). 27 were female. 26 of the participants were recruited via advertisement in the University of Cambridge, and the remaining 19 by word-of-mouth.

4.2.1.2 Materials and procedure Each participant completed a questionnaire consisting of 16 items. For each item, the participant was presented with a short

dialogue, and asked to provide both a range and their preferred interpretation of the value expressed by a numerically quantified expression in that dialogue. The numeral in the quantified expression was either primed earlier in the dialogue or unprimed. The precise instructions and an example item from this questionnaire in its primed and unprimed versions are as follows.

> Please read the following short dialogues, and answer the questions by filling in a value for each blank space, according to your opinion. Consider each dialogue separately. Assume that participant B is well-informed, telling the truth, and being co-operative in each case.
>
> Primed
> A: We need to sell 60 tickets to cover our costs. How are the ticket sales going?
> B: So far, we've sold fewer than 60 tickets.
> How many tickets have been sold? From …… to ……, most likely …….
>
> Unprimed
> A: We need to sell tickets to cover our costs. How are the ticket sales going?
> B: So far, we've sold fewer than 60 tickets.
> How many tickets have been sold? From …… to ……, most likely …….

The test items were designed to represent the following:

- 2 quantifiers:
  - more than n
  - fewer than n
- 3 levels of roundness of n, corresponding to 3 levels of scale granularity, with 2 numerals per level:
  - Coarse granularity: multiple of 100 ($n = 100, 200$)
  - Medium granularity: multiple of 10/non-multiple of 100 ($n = 60, 80$)
  - Fine granularity: non-round ($n = 77, 93$)
- 2 levels of priming:
  - primed
  - unprimed

12 scenarios of the form illustrated above were developed (2 quantifiers × 3 granularity levels × 2 numerals), with each written in primed and unprimed versions, which were identical except for the inclusion of the numeral in the first sentence of the primed variant (as in the above example). This resulted in 24 test items overall. An additional 4 primed and 4 unprimed filler items used the quantifier 'about n'. Two versions of the questionnaire were developed, each containing one variant (primed or unprimed) of each test and filler item, for a total of 12 test items (6 primed/6 unprimed) and 4 filler items (2 primed/2 unprimed) per questionnaire version. Every participant therefore responded to both quantifiers and each of the three granularity levels, in both primed and unprimed scenarios.

A full set of experimental materials is included as an Appendix to this paper.

4.2.1.3 Predictions We predict the following:

- In the unprimed condition, granularity effects will be observed as in Experiment 1, in that participants' estimates will be more distant from n for values of n occurring on coarser-grained scales
- In the primed condition, participants will not infer granularity-based upper/lower bounds to the value in question, such that estimates will be more distant from n in the primed vs. unprimed condition

4.2.2 Results

Prior to the analysis, 14 responses were removed due to missing or non-numerical answers.

Tables 3, 4 and Figs. 2, 3 show participants' median extreme and most likely estimates, expressed in terms of numerical distance from the value n in the quantifier; here, by extreme we mean the high end of the range in the 'more than n' case and the low end of the range in the 'fewer than n' case. As examples, a high estimate of 150 for a 'more than 100' item represents a distance of 50 from 100; a low estimate of 80 for a 'fewer than 100' item represents a distance of 20 from 100.

As was the case in Experiment 1, participants' estimates were more distant from n at coarser levels of granularity. We also observe higher estimates in the primed versus unprimed conditions.

A Kruskal–Wallis test (again chosen due to the non-normal distribution of the data) shows a significant effect of granularity on the extreme estimate ($p < 0.001$), with post hoc pairwise comparisons (Mann–Whitney U) showing significant differences between coarse and fine, medium and fine, and medium and coarse granularity (all $p < 0.001$). A significant overall effect of granularity is also found in the most likely estimates ($p < 0.001$), with all of the pairwise comparisons again significant ($p < 0.001$).

With regard to the effects of priming, a Mann–Whitney U test finds a significant difference between primed and unprimed items on both the extreme estimate ($p < 0.05$) and the most likely estimate ($p < 0.001$).

Turning to the results for the two quantifiers individually, in the case of 'fewer than n' a significant effect of granularity is found in both the extreme and most likely estimates ($p < 0.001$ for both); a significant effect of priming is likewise found in the extreme estimate ($p < 0.05$) and the most likely estimate ($p < 0.001$). For the quantifier 'more than n', a significant effect of granularity is found in the extreme and most likely estimates ($p < 0.001$ for both). However, the effect of priming falls short of significance ($p = 0.18$ for the extreme estimate; $p = 0.12$ for the most likely estimate).

To investigate the relationship of priming and granularity, the results for unprimed and primed test items were analysed separately. For the unprimed items, a Kruskal–Wallis test shows a significant effect of granularity on both the most likely and extreme estimates ($p < 0.001$ for both). A significant effect of granularity is also found for the primed items, in both the most likely and extreme estimates

Table 3 Most likely number (relative to n)—median (MAD)

	Granularity			
	Fine	Medium	Coarse	Total
Total				
Unprimed	3 (3.0)	10 (7.4)	20 (14.8)	10 (10.4)
Primed	7 (7.4)	15 (7.4)	20 (14.8)	15 (11.9)
Fewer than n				
Unprimed	3 (3.0)	10 (7.4)	25 (22.2)	10 (10.4)
Primed	12 (8.9)	15 (10.4)	20 (14.8)	17 (13.3)
More than n				
Unprimed	2.5 (2.2)	10 (7.4)	20 (14.8)	10 (8.9)
Primed	4 (4.4)	15 (7.4)	20 (14.8)	10 (10.4)
Total	4.5 (5.2)	10 (7.4)	20 (14.8)	

Table 4 Extreme end of range (relative to n)—median (MAD)

	Granularity			
	Fine	Medium	Coarse	Total
Total				
Unprimed	12 (13.3)	20 (14.8)	50 (42.3)	20 (19.3)
Primed	23 (23.7)	30 (19.3)	50 (44.5)	27 (25.2)
Fewer than n				
Unprimed	18 (16.3)	20 (14.8)	50 (44.5)	27 (29.6)
Primed	27 (25.2)	40 (29.7)	50 (44.5)	33 (29.6)
More than n				
Unprimed	7 (5.9)	20 (14.8)	50 (33.4)	20 (19.3)
Primed	13 (14.8)	20 (14.8)	40 (29.7)	20 (19.3)
Total	17 (14.8)	20 (14.8)	50 (44.5)	

(again $p < 0.001$ for both). In each case, all three pairwise comparisons (coarse/ fine, coarse/medium, medium/fine) are significant at the $p < 0.01$ level or stronger.

4.2.3 Discussion

These data replicate our findings from Experiment 1 concerning our first prediction —once again, there is evidence of granularity-based implicature. The range of values ascribed to the expressions does appear typically to be limited by the locations of numerals that match the utterance in granularity but which would be more informative if used; the result is that subjects give estimates more distant from n for rounder values of n.

In addition to validating our first experiment's finding with a more controlled pool of participants, we also extend it, by showing that granularity-based inferences
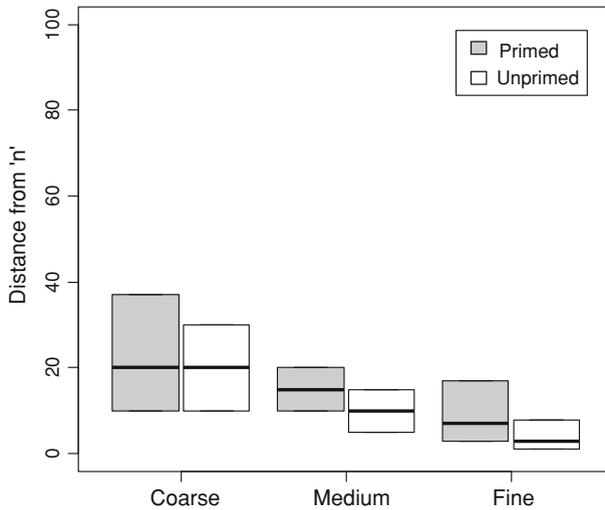
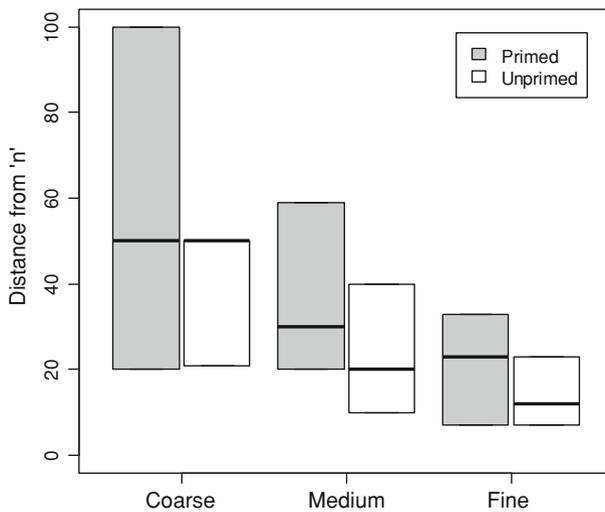**Fig. 2** Most likely number (relative to *n*)—median and quartiles

Fig. 3 Extreme end of range (relative to *n*)—median and quartiles

are available not only for monotone increasing quantifiers such as 'more than *n*' and 'at least *n*', but also for the monotone decreasing 'fewer than *n*'.

We also find support for our second prediction, namely that the inference should be attenuated by contextual salience of the numeral. When the numeral *n* has been previously mentioned in the discourse (primed condition), participants gave estimates significantly more distant from *n* than when it has not been not previously mentioned (unprimed condition). We interpret this pattern as indicating that, when

there is a contextually determined reason for the speaker to use the numeral, hearers are less likely to draw inferences based on the non-use of the next highest value on the scale of the appropriate granularity level.

Two aspects of our results are, however, unexpected in light of our predictions. First, granularity effects are observed not only in the unprimed condition—where we predicted them to occur—but also in the primed condition. Thus when the numeral n was made salient, participants were less likely to draw inferences based on the next higher value on a scale n occurs on, but they nonetheless appeared to derive pragmatically enriched interpretations which were in some way affected by granularity considerations. Thus the effect of contextual salience on granularity-based inferencing is a graded rather than absolute one. We return to this point in the conclusions, where we discuss more broadly the interaction of granularity and salience with other determiners of expression choice such as informativeness.

Second, the effects observed for priming are somewhat less robust than might have been expected, and in particular reach significance for 'fewer than n' but not 'more than n'. We hypothesise that task-related effects may have played a role here. The design of this experiment, which was intended to explore the interacting effects of priming and granularity, required that participants see both primed and unprimed test items. It is possible that the exposure to priming influenced participants to infer some relevance for the numeral stated even in the unprimed items where it was not mentioned, or conversely that the contextual occurrence of the numeral was sometimes overlooked in the primed items. Either could result in a blurring of the distinction between the two conditions. Relatedly, the various conditions were represented by 12 distinct dialogues, and in particular *more* and *fewer* were represented by different dialogues (see Appendix 2). It is thus conceivable that item effects contributed to the pattern of results observed. We investigate these two possibilities in Experiment 3.

4.3 Experiment 3: Effect of priming

In Experiment 2, we found evidence that subjects assigned wider ranges of interpretation to numerical quantifiers when the numeral was primed in the context than when it was not. We also found a difference between the behaviour of the quantifiers 'more than' and 'fewer than' with respect to priming. In Experiment 3, we investigate these effects more directly with a partial replication of Experiment 2, which was designed to address two potential issues with that study. Specifically: (a) participants saw only a single test item, in either primed or unprimed version; (b) the same scenario (dialogue) was used for both quantifiers tested.

4.3.1 Method

4.3.1.1 Participants   A total of 400 participants (100 per condition) were recruited via Amazon MTurk, using the same screening criteria and payment as in Experiment 1.

4.3.1.2 Materials and procedure Participants saw either the primed or the unprimed version of the following dialogue, featuring either 'more than' or 'fewer than', and were asked to estimate the number in question (range and most likely value):

Primed:
Salesperson: This storage unit holds 60 CDs. How many CDs do you own?
Customer: I have more than/fewer than 60 CDs

Unprimed:
Salesperson: This storage unit holds CDs. How many CDs do you own?
Customer: I have more than/fewer than 60 CDs

Participants also had the opportunity to comment on their answers.

The numeral 60 was chosen to represent an intermediate level of granularity (corresponding to a scale whose points are multiples of 10 or 20), at which level priming effects were clearly visible in Experiment 2.

The two 'more than' conditions were fielded in June 2010; the 'fewer than' conditions were fielded in December 2010. In both cases, fielding of the primed and unprimed conditions was separated by at least a day to minimise overlap between the samples.

4.3.1.3 Predictions We predict that for both 'fewer than 60' and 'more than 60', participants' estimates will be more distant from 60 (i.e. lower in the case of 'fewer', higher in the case of 'more') in the primed condition than in the unprimed condition.

4.3.2 Results

As in previous experiments, data were cleaned by the removal of incomplete and non-numerical responses. In addition, a number of responses (between 4 and 10 per condition) were removed where the respondent seemingly misinterpreted the instructions as calling for a single digit to be entered in each blank (resulting for example in a range of 'between 1 and 9' given for a 'more than' condition). This resulted in usable data from 92 subjects in the 'more than' primed condition, 91 in the 'more than' unprimed condition, 90 in the 'fewer than' primed condition, and 89 in the 'fewer than' unprimed condition.

Table 5 displays median values for subjects' estimates, expressed as both raw values and in terms of distance from the value 60 in the quantifier. Once again we observe that subjects' estimates are more distant from the value in the quantifier (i.e. higher for 'more than', lower for 'fewer than') in the primed than unprimed conditions. For example, for 'more than' the median estimate for the high end of the range condition was 100 in the primed condition (a distance of 40 from 60), compared to 80 in the unprimed condition (a distance of 20 from 60). Similarly, for 'fewer than' the median estimate for the low end of the range was 27.5 in the primed

Table 5 Number estimated—median (MAD)

	Raw values		Distance from n	
	Extreme	Most likely	Extreme	Most likely
More than n				
Unprimed	80 (23.7)	70 (8.9)	20 (23.7)	10 (8.9)
Primed	100 (29.7)	80 (23.0)	40 (29.7)	20 (23.0)
Fewer than n				
Unprimed	40 (17.8)	55 (4.4)	20 (17.8)	5 (4.4)
Primed	27.5 (25.9)	45 (14.8)	32.5 (25.9)	15 (14.8)

condition (a distance of 32.5 from 60), compared to 40 in the unprimed condition (a distance of 20 from 60).

A Mann–Whitney U test, with the distance of the extreme estimate from n as the independent variable, finds a significant difference between primed and unprimed conditions ($p < 0.001$); the corresponding analysis using the most likely estimate also shows a difference between primed and unprimed conditions ($p < 0.001$). Additionally, a significant difference between the quantifiers 'more than' and 'fewer than' is found with respect to the most likely estimate ($p < 0.001$), though not for the extreme estimate ($p = 0.15$).

Examining the two quantifiers separately, in the case of 'more than' a significant difference is found between primed and unprimed conditions with respect to both the extreme estimate and the most likely estimate ($p < 0.001$ for both). In the case of 'fewer than', a significant difference between conditions is also found for the extreme estimate ($p < 0.05$) and the most likely estimate ($p < 0.05$).

Further insight into participants' behaviour is obtained by examining the distribution of individual responses. Figures 4 and 5 show the distribution of the values given as the extreme end of the range.

Looking first at the results for 'more than' in Fig. 4, the pattern of responses is markedly different between the two conditions. In the unprimed condition, there are three main peaks, at 70, 80 and 100. The first two of these values (70, 80) correspond to the immediately higher values on two scales on which 60 occurs: the scale whose units are multiples of 10, and the one whose units are multiples of 20. The third of these values (100) corresponds to a point on a scale of higher granularity (i.e. multiples of 50 or 100). In the primed condition, by contrast, there is a single primary peak at 100, representing a higher level of granularity than the numeral in the quantifier; responses of 70 and 80 are much less frequent, while higher responses (200, 1000) also occur. A χ^2 test, comparing responses of 70, 80, 100, 200, and 1000 with all other values, shows the difference in distribution is significant ($\chi^2 = 18.7$, df $= 5$, $p < 0.01$).

The corresponding distributions for 'fewer than' are shown in Fig. 5. In both primed and unprimed conditions, an appreciable proportion of subjects gave a response of 0 or 1 for the lower end of the range. We might term these "truth conditional" responses, in that the range of interpretation is not affected by any sort
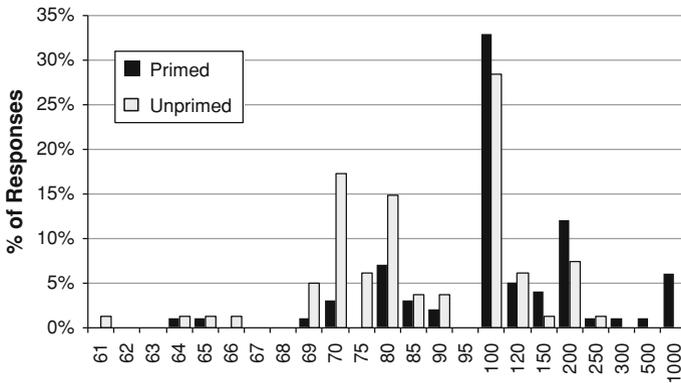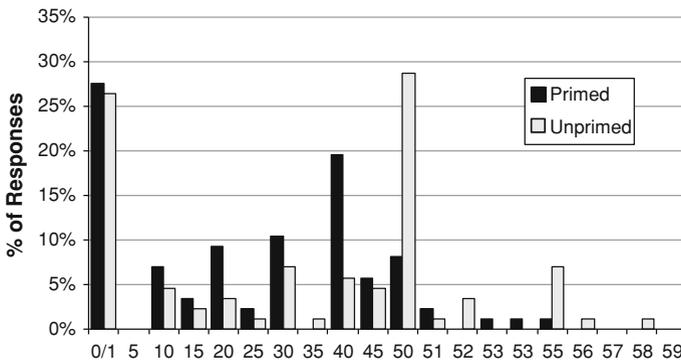
Fig. 4 High end of range—*more than n*

Fig. 5 Low end of range—*fewer than n*

of pragmatic strengthening. Other than these responses, we again observe a difference in the pattern of responses. In the unprimed condition, the primary peak is at 50, corresponding to the immediately lower point on two scales: that whose units are multiples of 10 (on which it occurs), and the coarser grained scale whose units are multiples of 50. Inferences relative to this scale point thus seem to be particularly robust. In the primed condition, in comparison, the pattern of responses is more diffuse, with 40 being the most commonly given value, followed by 30, 20, 50 and 10. A χ^2 test, comparing responses of 0/1, 10, 20, 30, 40, and 50 with all other values, shows the difference in distribution is significant ($\chi^2 = 16.92$, df $= 6$, $p < 0.01$).

4.3.3 Discussion

Corroborating the findings of Experiment 2, we find here further evidence for the role of contextual salience, this time for both quantifiers investigated. The interpretation of 'more than n' and 'fewer than n' is shaped by the contextual status

of the numeral n. When n is not in any way salient in the context, the interpretation assigned to 'more than n' is constrained by implicature to 'not more than m', where m corresponds to the next-highest point on a scale on which n participates, or a coarser-grained scale. Similarly, the interpretation of 'fewer than n' is constrained by implicature to 'not fewer than m' for some appropriate m. Round numbers typically occur on multiple scales, which differ in their granularity: 60, for example, seems to be a scale point on scales with granularity levels 10 and 20 (and perhaps others). As such, there are multiple options for the value m, and correspondingly variation in the range speakers assign to 'more than n' and 'fewer than n'. But notwithstanding this, granularity effects are observed.

On the other hand, when n is contextually salient, giving the speaker a reason to (re)use it, granularity effects are attenuated: hearers allow an overall wider range of interpretations for 'more than n' and 'fewer than n', and in particular show less of a tendency to draw inferences relating to the immediately higher or lower scale point. That is, alternative-based implicatures are less likely to be generated in the case where the numeral is salient. This accords with the second of our two main hypotheses, namely that prior mention of the numeral would attenuate scalar inferences of this type. Note however that typical responses in the primed condition (e.g. 100, 200, 1000) nevertheless correspond to points on scales of the same or coarser granularity than those on which 60 occurs. This sheds light on the finding in Experiment 2 that granularity effects are observed in the primed condition as well.

5 Conclusion

From our experiments, we conclude that hearers are able to enrich numerically-quantified expressions such as 'more than' and 'fewer than' with scalar implicatures, even in unembedded contexts, contrary to the claims prevalent in the literature (Krifka 1999; Fox and Hackl 2006). This suggests that no additional theoretical machinery need be posited to account for the seemingly deviant pragmatic behaviour of these quantifiers. Instead we showed that two established factors of pragmatic reasoning are important for the implicatures triggered by modified numerals. Specifically, we show that granularity, construed in terms of the density of measure points on the number scale, and contextual salience, specifically with regard to whether the numeral used has previously been mentioned in the context, exert significant influence on these inferences.

Our results also provide some insight into the relative strength of the two factors granularity and salience, showing specifically that salience can attenuate the effect of granularity. It remains to be established precisely how these factors interact. One possible account of this is offered by Cummins (2011), which posits a constraint-based model for the use of numerical quantifiers by a speaker. In this model, the speaker selects the utterance that best satisfies a ranked set of constraints, while maintaining truthfulness. Hearers then recover meanings by drawing inferences about the situation that must prevail, given the fact that the speaker's utterance was optimal. With respect to our experiments, the relevant constraints in Cummins's model are *Numeral Priming*, which is violated when a contextually activated

number is not used, *Numeral Salience*, which is violated when the number used is not intrinsically salient (i.e. round),[5] and *Informativeness*, which is violated when the speaker makes a less informative statement than they could have.

This model's treatment of priming effects can be sketched as follows. In the unprimed case, the Numeral Priming constraint exerts no effect. The speaker's choice is conditioned by Numeral Salience and Informativeness. Thus, the decision to use a particular numeral implicates that no alternative was available that was at least as good with respect to both constraints: that is, there was no more informative expression available to the speaker that used at least as round a number. Hence the hearer is predicted to draw the inference discussed earlier (for instance, that "more than 90" implicates "not more than 100"). By contrast, in the primed case, Numeral Priming becomes relevant as well. The decision to use a particular numeral now implicates only that no alternative was available that was at least as good with respect to all three constraints. If the primed numeral is reused, then this implicature is vacuous: no other numeral could be as good with respect to Numeral Priming. Hence, reuse of a numeral is predicted to block implicatures about informativeness. In this way, the model formalises the notion that the reuse of a numeral might merely reflect a decision to save effort, in which case it should not convey any pragmatic effects.

The above model requires further refinement to accommodate the observation that granularity effects are still observed in cases where the numeral is primed, as shown in experiments 2 and 3. One possibility is to posit gradient effects in constraint violation, which would imply that adherence to Numeral Priming becomes increasingly disfavoured as the loss of Informativeness increases. For example, in the case where 60 is primed, 'more than 60' might be favoured by Numeral Priming up to some ceiling value, at which point a more informative expression becomes preferred, whereupon Numeral Salience will favour the use of round numbers (thus preserving granularity effects). Alternatively, numerals such as 100 might be categorically more prominent than those at lower levels, in which case they may be preferred on the grounds of Numeral Salience, even in the face of a competing preference for Numeral Priming. This remains a matter for future study, so we shall not speculate here about the merits and demerits of such approaches. In general, though, it appears that a constraint-based account such as that proposed by Cummins (2011) has potential utility in explaining data such as those we present here.

A further point requiring discussion is the curious behaviour of quantified expressions at the finest level of granularity. As discussed above, their implicatures seem frequently to arise with reference to a higher level of granularity: for instance, "more than 93" implicating "not more than 100" rather than "not more than 94". As this is crucial to the analysis of quantifiers with small numerals, which have been adduced as evidence that modified numerals do not convey implicatures, it merits further discussion.

---

[5] Note that the term "salience" is used differently in Cummins (2011) than in the present paper. The Numeral Salience constraint relates to roundness, in our terms, while the Numeral Priming constraint relates to prior mention of the numeral.

Our account for the data at the finest level of granularity proceeds in two steps. In the first step, we observe that use of "more than n" is expected to be blocked at the finest level of granularity except in certain contexts or situations. Secondly, we propose that hearers therefore infer from a speaker's use of "more than n", where n is not round, that such a context must be present. However, as a consequence of this, the hearers cannot then draw the typical granularity-based inference from the speaker's utterance. In the following, we show that this account predicts the experimental data and also the speakers' intuitions concerning quantifiers with small numerals, such as 'more than three'.

The first step of our account has already been argued in the literature: Fox and Hackl ([2006]) observe that the lexical content combined with the implicature arising from 'more than three'—that is, 'exactly four'—could be communicated more efficiently by the use of '(exactly) four'. A knowledgeable speaker who is aware that 'exactly four' is the case should therefore say 'four' rather than 'more than three' (and similarly, *mutatis mutandis*, for the case of 'more than 93' versus '94').

This argument dispenses with the unwanted implicature, but at the cost of excluding the possibility of 'more than three' (or 'more than 93') being uttered by a knowledgeable speaker: given the perfectly good examples of this that have been discussed, it seems we are throwing the baby out with the bathwater. At this point, the second step of our account becomes important: we argue that 'more than n' can be used with non-round n, but only in certain contexts. One such case is if the speaker is not epistemically committed (that is, the case where the speaker's knowledge might be characterised as "$>n$"). The second is the one we have investigated in this research, namely when the numeral n is contextually salient. For this explanation to work, we would have to be able to argue that expressions such as 'more than three' actually surface only in such contexts. We would contend that the examples discussed in the literature do indeed meet this condition. For instance, (25) naturally supposes a context in which John's having three children is somehow critical, perhaps because it is a threshold for benefits, it is as many as will fit in the back of the car, or because it means he has more children than some other individual who is known to have exactly three. (Alternatively, it is compatible with a situation in which the extent of the speaker's knowledge is merely that the number of John's children is greater than three.)

(25)   John has more than three children.

This seems to apply not only for small n but for fine-grained values of n in general. Consider for instance (26), based on an item from Experiment 1.

(26)   More than 93 people attended the meeting.

Intuitively, it only seems natural for this to surface in a context in which 93 is a critical level for attendance; for example, if it is the quorum for the meeting, or the typical attendance, or a guess someone has made about the attendance. In fact, presented with this sentence without context, some participants in Experiment 1 commented on its oddness. There are good reasons for this. If 'more than 93' does

not reuse a number and does not directly address a question under discussion, both the speaker and the hearer are being forced to process a numeral that is not highly available. Furthermore, this numeral does not even correspond to the precise number of people at the meeting. All this does is to generate additional processing requirements without achieving much in the way of additional cognitive effects: i.e. it is less relevant than some alternative.

Furthermore, when one searches for corpus examples of modified numerals containing non-round numerals, the examples one finds are precisely those where the numeral has some contextual salience or relevance. For example, (27) refers to the number of seats required to command a majority in the UK House of Commons, (28) to the 'perfect' score for an over in cricket, and (29) to the size of a standard pack of cards.

(27) Maybe fewer than 25 % think the unthinkable—that the Tories will obtain fewer than 326 seats.[6]

(28) It is also possible to get more than 36 runs in an over, but it has never happened before.[7]

(29) [M]ost decks used in casinos for poker have more than 52 cards so people don't cheat.[8]

We propose that hearers, when confronted with a quantifier such as 'more than 93', actually posit that there is some reason for this particular expression to be chosen—either because the numeral itself is in some way salient, or because the speaker's knowledge does not allow a more informative expression to be used. Once it is supposed that there is some such reason, there will be no implicature, because the speaker's choice of utterance is to some extent forced.[9] Or perhaps we should more accurately say that, to the extent that the hearer supposes that there is a reason for using that number, the hearer does not draw an inference.

This effect should also influence hearers' interpretations at a coarser granularity level, albeit to a lesser extent. When a hearer interprets 'more than 100', they are entitled to question whether there is a specific reason for 100 to be referred to; and if so, they may not draw the inference. This might further account for the variability exhibited by participants in both our experiments: a competent user of language should interpret 'more than 100' without an implicature if the speaker's decision to mention 100 is contextually forced. In this way we, as hearers, can 'repair' infelicitous utterances such as (22) by positing a context against which they can be interpreted without implicature.

This proposal appears to account for the distribution of responses to quantifiers at the finest granularity level, and in doing so to account for the lack of implicatures

---

[6] http://www2.politicalbetting.com/index.php/archives/2010/01/17/do-these-1992-approval-ratings-hold-the-key/, retrieved 2nd August 2010.

[7] http://www.sports1234.com/cricket/1652-2-cricket.html, retrieved 2nd August 2010.

[8] http://answers.google.com/answers/threadview/id/100828.html, retrieved 2nd August 2010.

[9] Here it is not crucial whether the speaker is epistemically committed but choosing to reuse a contextually-activated number, or merely not epistemically committed. In either case, the implicature should not be available.

from expressions such as 'more than 3 children' that has been observed in the literature. However, clearly further experimental investigation is required to ascertain whether hearers are indeed reasoning in the way we hypothesise. If so, it remains to be shown how our participants' responses are derived: for instance, are they acting in accordance with some specific imagined context, and if so, can we establish how they select this context?

We turn finally to the question of how our results can be unified with the findings of Fox and Hackl (2006). Fox and Hackl observe that while comparative quantifiers do not seem to give rise to implicatures (and do not associate with 'only') in unembedded contexts, they do so when embedded under certain, though not all, modals. They exemplify this by citing the following examples (p. 544; their (13) and (14)).

(30) a. You're required to read more than 30 books.
          Implicature: There is no degree greater than 30, d, s.t.
          you are required to read more than d books.
      b. You're only required to read more than 30 books.

(31) a. You're allowed to smoke more than 30 cigarettes.
          *Implicature: There is no degree greater than 30, d, s.t.
          you are allowed to smoke more than d cigarettes.
      b. *You're only allowed to smoke more than 30 cigarettes.

Our results show that, contrary to Fox and Hackl's assumption, modified cardinals can yield implicatures even in unembedded contexts. However, this does not explain the difference between the behaviour of (30) and (31). If Fox and Hackl's analysis of (30) is correct, it does appear to constitute strong evidence in favour of their proposal. It might nevertheless be interesting to investigate whether the interpretation of (30a) involves the implicature that they posit or a weaker one conditioned by granularity or salience, such as appears to arise from a dialogue such as (32).

(32) A. Can I apply for an international business development grant?
          I have ten employees.
      B. You need more than ten employees to count as an 'international business'.

If however the Fox and Hackl (2006) type implicature is not subject to these considerations, it suggests that the two types of implicature require different accounts. In fact, at least the grammatical theory of scalar implicature allows for two kinds of implicatures and other theories of implicature may do so too. The grammatical theory of implicature proposes that scalar implicatures in most cases arise from a silent semantic exhaustification operator Exh (Chierchia et al. 2008). One version of this operator, as defined in (33), negates all alternatives to the actual utterance that are contextually salient—i.e. members of the set C.

(33) Exh(C)(p) = p & $\forall p' \in C: \neg p'$

However, the grammatical account does not attribute all implicatures to the silent exhaustification operator, and claims that pragmatic reasoning accounts for the

remaining implicatures. The pragmatic reasoning assumed by the grammatical account is weaker than the analysis of scalar implicatures we introduced in Sect. 2. Consider the case of a speaker uttering S that has a logically stronger scalar alternative S*. Above we assumed that hearers conclude in this case that the speaker thinks that S* is false. However, another version of the pragmatic reasoning assumes that hearers conclude that the speaker is not certain that S* is true.[10] More formally, if we use Bel_σ as an operator for speaker belief, the two candidate implicatures are represented as $Bel_\sigma(\neg S^*)$ for the first version of pragmatic reasoning and $\neg Bel_\sigma(S^*)$ for the latter. From these two representations, it is transparent that the former entails the latter—i.e. the latter form of pragmatic reasoning yields strictly weaker results. The grammatical account assumes only the weaker form of pragmatic reasoning. This predicts that in many examples the effect of pragmatic reasoning is impossible to detect, because exhaustification yields stronger inferences. Disjunction is the major example where both mechanisms can be seen at work. Use of a disjunctive statement of the form 'A or B' implicates (1) that the speaker believes that 'A and B' is false and (2) that the speaker is not certain that A is true and is also not certain that B is true. The Exh operator accounts only for the former implicature, and the later is taken to derive from pragmatic reasoning.

The implicatures triggered by modified numerals can receive a similar account to disjunction. On the one hand, as Fox and Hackl (2006) show, the implicatures from modified numerals embedded below a modal can be attributed to Exh, while similar implicatures are blocked when no modal is present. On the other hand, the granularity-based implicatures from modified numerals can be explained by pragmatic reasoning and are predicted to be independent of the presence of a modal, since modals are upward-entailing. For example, the use of an expression "more than 100" in a sentence S leads by pragmatic reasoning to an implicature that for any numeral $m > 100$ and at the same granularity level as 100, it is not certain that S would hold if "more than 100" were replaced by "more than m". The account predicts that these pragmatic implicatures should have the weaker epistemic force noted above: the speaker is uncertain that a stronger claim holds. Our experiments were not specifically designed to distinguish between epistemically strong and weak implicatures, so we are not yet able to test this prediction. We see this as a direction for future research.

---

[10] The difference between the two versions can be seen as relating to Grice's two maxims of quality.

Appendix 1: Excluded and flagged responses (Experiment 1)

	Range condition			Single number condition		
	Incomplete/not single numeral (excluded from analysis)	Outlier (high estimate)	Inconsistent w/truth conditions	Incomplete/not single numeral (excluded from analysis)	Outlier	Inconsistent w/truth conditions
More than n						
$n = 100$	3	6	15	6	1	19
$n = 110$	2	3	16	5	2	28
$n = 93$	3	3	31	1	1	34
At least n						
$n = 100$	3	2	25	0	1	11
$n = 110$	1	1	36	0	1	22
$n = 93$	2	1	33	1	2	15

Appendix 2: Experimental materials (Experiment 2)

For each of the following items, A's utterance contains a numeral in the primed condition. In the unprimed condition, the numeral is omitted. Participants saw either a version of the questionnaire in which items 1, 3, 4, 8, 9, 10, 12 and 16 were primed, or one in which items 2, 5, 6, 7, 11, 13, 14 and 15 were primed.

Please read the following short dialogues, and answer the questions by filling in a value for each blank space, according to your opinion. Consider each dialogue separately. Assume that participant B is well-informed, telling the truth, and being co-operative in each case.

1. A: We need to sell [60] tickets to cover our costs. How are the ticket sales going?
   B: So far, we've sold fewer than 60 tickets.

   How many tickets have been sold? From …… to ……, most likely …….

2. A: To win the election, the candidate had to convince [77] people to vote for him.
   How many votes did he get?
   B: He got more than 77 votes.

   How many votes did the candidate get? From …… to ……, most likely …….

3. A: This photo album has space for [150] 4 × 6 photos.
   How many photos of that size do you have?
   B: I have about 150 photos.

How many photos does B have? From …… to ……, most likely ……

4. A: We've invited [80] members and non-members to the reception.
    How much room is there in the hall?
    B: There's room for more than 80 people.

    How many people is there room for in the hall? From …… to ……, most likely …….

5. A: There are still [200] copies of the agenda left. How many delegates
    are we waiting for?
    B: We're waiting for fewer than 200 delegates.

    How many delegates are they waiting for? From …… to ……, most likely …….

6. A: This display case holds [80] CDs. How many CDs do you own?
    B: I own fewer than 80 CDs.

    How many CDs does B own? From …… to ……, most likely …….

7. A: The lecture hall can accommodate [130] students taking an exam.
    How many students will be taking tomorrow's exam?
    B: About 130 students will be taking tomorrow's exam.

    How many students will be taking the exam? From …… to ……, most likely …….

8. A: I expect [93] cars to arrive between now and three o'clock.
    How many spaces are left in the car park?
    B: There are more than 93 spaces left in the car park.

    How many spaces are left in the car park? From …… to ……, most likely …….

9. A: We wanted [200] new signatures on the petition. How many new people
    signed it?
    B: Fewer than 200 new people signed it.

    How many new people signed the petition? From …… to ……, most likely …….

10. A: The new prison still has space for [110] prisoners. How many prisoners
    need to be transferred out of the old facility?
    B: About 110 prisoners need to be transferred.

How many prisoners need to be transferred? From …… to ……, most likely …….

11. A: [60] [P]eople have already paid their deposit for this holiday. How many seats are there on the plane?
     B: There are more than 60 seats on the plane.

How many seats are there on the plane? From …… to ……, most likely …….

12. A: We can hire a bus with [77] reclining seats for the excursion. How many people will be coming?
     B: Fewer than 77 people are coming.

How many people are coming on the excursion? From …… to ……, most likely …….

13. A: We have space in the garden to plant [50] tulips. How many tulip bulbs are in the bag?
     B: There are about 50 tulip bulbs in the bag.

How many tulip bulbs are in the bag? From …… to ……, most likely …….

14. A: There are [100] free tickets available for people who call the station today. How many people have called so far?
     B: Fewer than 100 people have called so far.

How many people have called the station so far today? From …… to ……, most likely …….

15. A: We have enough budget left to print [93] colour copies of the brochure. How many people have asked for one?
     B: Fewer than 93 people have asked for a copy.

How many people have asked for a copy of the brochure? From …… to ……, most likely …….

16. A: We can provide [100] lunches for the guests. How many of them will stay for lunch?
     B: More than 100 of them will stay for lunch.

How many of the guests will stay for lunch? From …… to ……, most likely …….

References

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy, 4*, 159–219.

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

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). Oxford: Oxford University Press.

Chierchia, G., Fox, D., & Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. Ms.

Cummins, C. (2011). *The interpretation and use of numerically-quantified expressions*. PhD thesis, University of Cambridge.

Cummins, C., & Katsos, N. (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics, 27*, 271–305.

Curtin, P. (1995). *Prolegomena to a theory of granularity*. MA thesis, University of Texas at Austin.

Fox, D., & Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy, 29*, 537–586.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes, 25*(1), 130–148.

Geurts, B., & Nouwen, R. (2007). "At least" et al.: The semantics of scalar modifiers. *Language, 83*, 533–559.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*, 383–393.

Horn, L. (1972). *On the semantic properties of logical operators in English*. UCLA dissertation, distributed by Indiana University Linguistics Club, 1976.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin (Ed.), *Meaning, form, and use in context* (pp. 11–42). Washington: Georgetown University Press.

Jansen, C. J. M., & Pollmann, M. M. W. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics, 8*, 187–201.

Krifka, M. (1999). At least some determiners aren't determiners. In K. Turner (Ed.), *The semantics/pragmatics interface from different points of view*. Current research in the semantics/pragmatics interface (Vol. 1, pp. 257–292). Oxford: Elsevier.

Krifka, M. (2002). Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision. In D. Restle & D. Zaefferer (Eds.), *Sounds and systems. Studies in structure and change: A festschrift for Theo Vennemann* (pp. 439–458). Berlin: Mouton de Gruyter.

Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and evidence in semantics* (pp. 109–132). Stanford: CSLI Publications.

Mason, J. D., Healy, A. F., & Marmie, W. R. (1996). The effects of rounding on memory for numbers in addition problems. *Canadian Journal of Experimental Psychology, 50*(3), 320–323.

Nouwen, R. (2008). Upper-bounded *no more*: The exhaustive interpretation of non-strict comparison. *Natural Language Semantics, 16*(4), 271–295.

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass, 6*, 36–49. doi:10.1002/lnc3.321.

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

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43*, 155–167.

Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind, 111*, 583–632.