# Modelling implicatures from modified numerals

**Citation for published version:**
Cummins, C 2013, 'Modelling implicatures from modified numerals', *Lingua*, vol. 132, pp. 103-114.
https://doi.org/10.1016/j.lingua.2012.09.006

**Digital Object Identifier (DOI):**
10.1016/j.lingua.2012.09.006

**Abstract**

It has been argued that comparative and superlative quantifiers (such as “more than” and “at least”) fail to yield scalar implicatures in unembedded declarative contexts (Krifka 1999, Fox and Hackl 2006). However, recent experimental work has shown that such implicatures are available, and that these are constrained by considerations of granularity or numeral salience (Cummins, Sauerland and Solt, submitted). That is, “more than *n*” triggers a pragmatic upper-bound, the value of which depends on the numeral *n*. However, Cummins, Sauerland and Solt further show that this effect is weakened by prior mention of the numeral *n*, a finding that they interpret in terms of priming effects. In this paper I discuss a recent theoretical proposal that accommodates these findings by arguing for a model of numerical quantifier usage based on multiple constraint satisfaction. This approach provides a means of accounting for the influence of contextual factors on the speaker’s choice of utterance. It also makes predictions as to how a rational hearer should use context in their interpretation of utterances. Here I explore how this can yield the diversity of interpretations exhibited by participants in Cummins, Sauerland and Solt (submitted). More generally I consider how such an account predicts that a hearer will infer context on the basis of an utterance, and examine how this offers a potential explanation for the previously observed failure of implicature in the domain of modified numerals. I discuss the implications of this for experimental semantic and pragmatic methodologies, with particular reference to the numeral domain.

**Keywords:** Scalar implicature, numerals, quantifiers, priming, context

## 1. Introduction

In classical scalar implicatures, the use of a weak term on an informational scale is taken to convey information about the speaker's inability to use a stronger term on that scale. Under certain assumptions about the speaker's informational state and the appropriateness of the strong scalar term, the hearer is entitled to draw the further inference that the stronger statement is false (see Geurts 2010: 27-30). An example of this for the scale <some, all> is the exchange (1).

(1) A. Did your students pass the exam?

B. Some of them did.

In (1), B might have said "All of them did". This would have conveyed extra relevant information at no apparent cost, and would not obviously have been impolite or face-threatening. Furthermore, B can be presumed to know whether the stronger statement would hold. Under all these conditions, B's utterance appears to convey the implicature that not all of B's students passed the exam.

A curious observation concerning numerical expressions is that certain classes of these seem systematically to fail to convey implicatures. Consider the putative scale <more than three, more than four, …> This appears to be a legitimate scale in principle, its terms meeting the criteria proposed by Horn (1984): they are equally lexicalised and vary in strength (the latter entailing the former) while conveying information of the same semantic type. However, an utterance using a weaker term fails to give rise to the implicature that the strong term does not hold, as in (2).

(2) A. How many children does John have?

B. More than three.

By a similar argument to that proposed above, B's utterance should implicate the falsity of the stronger statement "More than four" if B is knowledgeable, cooperative and so on. This in turn would entail that John had exactly four children. However, this is clearly not an appropriate interpretation of B's utterance (a point supported empirically by Geurts et al. 2010). On the basis of examples of this kind, Fox and Hackl (2006) argue that comparative quantifiers such as "more than" fail to participate in informational scales, a point made for superlative quantifiers ("at least", "at most") by Krifka (1999). In both cases, this observation serves as a partial motivation for a specific semantic account that renders the scale inoperative.

In recent experimental work, Cummins, Sauerland and Solt (submitted) challenge this observation, arguing that implicatures are admitted by expressions such as "more than *n*", but that these are restricted by properties of the numeral in question. They elicit judgements as to the values conveyed by “more than *n*” and similar expressions, in unembedded declarative contexts, for various values of *n*. Their results indicate that “more than *n*” does attract a pragmatic enrichment, not to the effect of “not more than *n*+1” but to the effect of “not more than *m*”, where *m* is some specific numeral greater than *n* in each case. For instance, “more than 80” is reliably interpreted to mean “not more than 100”.

Specifically, Cummins et al. (submitted) argue that “more than *n*” is taken to implicate the falsity of any “more than *m*” for $m > n$ that could have been asserted at no greater cognitive cost. They argue, following the literature on numerical cognition (Dehaene 1997 i.a.) that numerals vary in their inherent accessibility, or in the cognitive cost associated with their use. On this account, the use of a round number (in the sense of Jansen and Pollmann 2001) is argued to be more costly than the use of a non-round number. Similarly, they appeal to the notion of granularity, and argue that the use of a value corresponding to the scale point of a coarse-grained scale (in the sense of Krifka 2009, such as “24 hours”) is less costly than the use of a value that corresponds only to a scale point of a fine-grained scale (such as “22 hours”). Consequently, on this account, even a fully-informed speaker typically has a potential reason to make a less informative statement than they might otherwise have done. The hearer is implicitly aware of this and moderates their inference accordingly.

This argument addresses the curious fact that examples such as (3) are widely judged to be pragmatically anomalous (in the absence of a special supporting context), despite being true in semantic terms and apparently immune from implicature. According to this account, (3) fails to implicate the falsity of (4) merely because (4) uses a non-round number and would therefore be more costly to utter. (3) does however implicate the falsity of (5), which uses a highly round number, and it is the clash between this implicature and encyclopaedic knowledge that renders (3) infelicitous.

(3) More than 1000 people live in Paris.

(4) More than 1001 people live in Paris.

(5) More than a million people live in Paris.

Cummins, Sauerland and Solt (submitted) demonstrate that hearers interpret expressions with comparative quantifiers as having pragmatic bounds of this type. For “more than 100”, a typical pragmatic upper-bound is 150 or 200; for “more than 110”, it is 120 or so. Notably, “more than 100” can be used felicitously (according to the intuitions of their experimental participants) to refer to higher values that can be felicitously referred to by “more than 110”, which is apparently inexplicable unless attention is paid to the psychological attributes of the numeral itself.

Intuitively it also appears that utterances such as (3) can be salvaged by being furnished with a context in which that precise numeral (“1000”) is relevant (e.g. by asking “in which cities do more than 1000 people live?”) Cummins, Sauerland and Solt (submitted) hypothesise that the prior mention of a numeral in the preceding context renders it more highly salient in the mind of the speaker, thus favouring its re-use in a wider range of situations than would otherwise be the case. They also hypothesise that hearers should be aware of this and adjust their interpretation accordingly. For instance, the meaning of “more than 60” should receive a tighter pragmatic bound if “60” is chosen out of the blue than if “60” is already activated in the discourse. They test this prediction by presenting items in which a numeral is present or absent from the immediately preceding conversational turn, and asking participants to judge the meaning of the subsequent quantifier, as in (6). They show that, in accordance with their hypothesis, cases in which the numeral is mentioned give rise to weaker pragmatic bounds (that is, they admit a greater range of interpretation) than those in which the numeral is not mentioned.

(6) A. This case holds (60) CDs. How many CDs do you own?

B. I own more than 60 CDs.

Given the contexts tested, it appears debatable whether the effects observed are due to low-level priming considerations or relate to higher-order considerations such as question under discussion (QUD; Roberts 1996). It is possible that the re-use of a numeral merely serves to anchor the response to the question, and that it is preferred for this reason. In either case, the observation remains that the implicatures arising from modified numerals are potentially curtailed by the preceding context.

In this paper, I discuss a recent proposal for modelling the speaker's choice of utterance that treats this as a problem of multiple constraint satisfaction. I examine how this model accounts for the critical data discussed above, and briefly consider how alternative accounts of discourse context could be framed within this model. Building upon this account, I then consider the reasons why implicatures might be blocked within such a model, with particular reference to examples such as (2). I identify a set of possible circumstances under which cases such as (2) should resist implicature, and suggest that the apparent immunity from implicature is in fact attributable either to the presence of such circumstances, or the strong tendency of the hearer to infer their presence. In defence of this latter point, I consider evidence for hearers drawing inferences about aspects of context to which they were not privy, and explore how the model accommodates this, and how this tendency has implications for experimental work on the semantics-pragmatics interface.

## 2. Constraints referring to structures and to context

Cummins (2011) proposes a constraint-based model of numerical quantifier usage and interpretation, using a speaker-referring unidirectional Optimality Theory (OT) formalism (Prince and Smolensky 1993). On this account, usage is governed by a set of violable ("soft") constraints, which each speaker individually ranks. In any given situation, the speaker then produces the optimal output given their constraint ranking. The model differs from bidirectional OT accounts (e.g. Blutner 2006) in its focus on processing rather than the optimisation of the entire system, and consequently imputes different roles to the speaker and hearer. Under this account, the hearer computes the speaker's intention by attempting to reconstruct it from the utterance, given the knowledge that the utterance is optimal[1]. This will be discussed further in what follows.

OT posits two kinds of constraint: markedness and faithfulness constraints. Markedness constraints are those that govern the output, and are violated by the use of certain forms (those which are defined to be marked, in effect) in any circumstances. Faithfulness constraints are those that govern the relation between the input and output levels, and are violated when there is a mismatch of the appropriate type between these levels. In the case of phonology, the input level is that of underlying forms and the output level is that of surface forms. However, in the model proposed by Cummins (2011), the input level is interpreted as consisting of (the speaker's representation of) the situation in which the

[1] In principle this does not rely upon any scalar inferencing.

utterance is made, while the output level is that at which the utterance is selected. Thus, contextual factors are treated using faithfulness constraints in this model[2].

Constraints are proposed based on individual experimental and philosophical justification (following McCarthy 2002: 41f). Cummins (2011) adopts six such constraints, the definitions of which are sketched here. Two are markedness constraints: numeral salience (requiring the use of a round number, in the sense of Jansen and Pollmann 2001) and quantifier simplicity (requiring the use of a simple quantifier, as measured by the cognitive cost of its usage). Four are faithfulness constraints: informativeness (make the utterance maximally informative given the knowledge state of the speaker), granularity (respond at the appropriate granularity level, where the context specifies this), numeral priming (if a numeral is contextually activated, re-use it) and quantifier priming (if a quantifier is contextually activated, re-use it). Of these, the last three refer to external aspects of the situation, which we might consider to belong to the discourse context. In addition, Cummins (2011) states an overarching requirement that the utterance should be truthful, in accordance with the Gricean maxim of quality (Grice 1975), which disqualifies semantically false outputs from consideration[3].

The operation of the OT model can be sketched as follows. Given a situation, the speaker proceeds from the highest-ranked constraint, excluding all utterances that incur more than the minimum number of violations of that constraint. The speaker then considers the

[2] Blutner (2006: 13) refers to 'linking' constraints rather than faithfulness constraints, but no significance appears to attach to the choice of terminology.

[3] This could also be treated as a violable constraint: the uncertainty as to its proper classification echoes Grice's uncertainty as to whether quality should be a maxim or a more general principle (Grice 1975: 46).

next highest-ranked constraint, again excluding all utterances that violate it more than minimally. When only one possible utterance remains, this is selected. It should be stressed that this is an idealisation, in that the speaker evaluates infinitely many possible utterances under this account: in tractable implementations of this the search space must necessarily be restricted. In the examples given here I restrict my attention to a few possibilities in each case, noting that the others are generally excluded by extensive constraint violations.

Notably, the only information that is relevant to the speaker's choice of utterance, in this model, is that concerning the evaluation of constraint violations. What constitutes a violation is specified precisely in the full definition of the constraints; for reasons of space I only consider the relevant aspects of their definitions here (but see Cummins 2011 for a specific proposal). In particular, with regard to context, only considerations that are referred to by faithfulness constraints are potentially relevant to the choice of utterance: the selection procedure is indifferent to other factors. So a descriptively adequate model in this framework would encompass a definition of relevant context (although this definition might be too broad, as the model would not necessarily be minimal).

Now, from the hearer's perspective, the task is different. The hearer is assumed to have implicit knowledge of the constraints (some participants in Cummins, Sauerland and Solt's study in fact exhibited explicit knowledge of at least some of them), but is not privy to the speaker's constraint ranking. However, the hearer is aware that the utterance is presumably optimal given that constraint ranking and the situation. The hearer may therefore infer the falsity of stronger statements to the extent that those statements could have been made by the speaker. As speakers are assumed to vary in their constraint rankings, it may not be possible to know for certain whether a given speaker would have made a given stronger statement. However, if a particular stronger statement would have been preferable with respect to all relevant criteria, then any knowledgeable speaker would preferentially have made that statement. In this case, the speaker's failure to utter it should reliably implicate its falsity (under the standard licensing conditions for scalar implicature, such as the speaker's epistemic commitment). If the stronger statement would be dispreferred on the basis of incurring additional violations of certain constraints, then the implicature of its falsity is weakened. This echoes standard Gricean reasoning (Geurts 2010: 27-30): implicatures concerning the falsity of stronger statements should not be available if the speaker is blocked from making those statements by other considerations, such as ignorance, politeness, or the irrelevance of the stronger proposition to the discourse purpose. Experimental studies show that implicatures are indeed blocked under these conditions (e.g. Breheny, Katsos and Williams 2006, Bonnefon, Feeney and Villejoubert 2009).

In the following section I exemplify the task of the hearer with respect to the cases discussed earlier, before considering how context may be used as a clue to utterance interpretation, and how the utterance may be used as a clue to context.

**3. Pragmatic enrichments without preceding context**

Within the constraint-based model, the pragmatic bounds demonstrated by Cummins, Sauerland and Solt (submitted) for “more than *n*” and “fewer than *n*” in the absence of prior context can be accounted for in terms of informativeness (INFO) and numeral salience (NSAL). INFO is violated by the use of an expression that fails to exclude possibilities that are known to be false by the speaker; NSAL is violated by the use of a numeral that fails to exhibit roundness (on the grounds that such a number is not psychologically salient to the speaker). Cummins (2011) adopts a notion of roundness based on Jansen and Pollmann (2001), but for these examples an intuitive approach will suffice.

Taking the speaker's point of view, we can consider how the various different expressions compete in given situations. In this section, we first assume that there is no relevant material in the preceding context. Now for instance suppose that the speaker knows the value under discussion to be in the range 101-109. Semantically truthful options include (7a-c) and so on[4].

(7a) more than 100

(7b) more than 90

(7c) more than 80

According to Jansen and Pollmann's (2001) classification, 100 is the roundest of these numbers, followed by 80 and then 90. Thus, 100 does not violate NSAL. Moreover, it incurs fewer violations of INFO than its competitors. In OT parlance, it harmonically bounds the alternatives, and will be selected under any constraint ranking, as shown in

[4] In particular, a set of alternatives with "at least" should be considered. In this partial example I ignore these on the basis that they are argued to violate quantifier simplicity (following Cummins and Katsos 2010, Geurts and Nouwen 2007). This effectively means that expressions with "more than" should not reliably implicate the falsity of expressions with "at least" in this model. For a detailed discussion, see Cummins (2011).

OT tableau form in Table 1. In this tableau, * signifies a constraint violation, and the use of a dotted line signifies that the constraints are not assumed to be ranked.

**Table 1: Constraint violations of (7a-c) with respect to INFO and NSAL for quantity 101-109**

| | INFO | NSAL |
|---|---|---|
| more than 100 | | |
| more than 90 | * | ** |
| more than 80 | ** | * |

Now suppose the value is in the range 91-99. (7a) is ruled out on semantic grounds, leaving (7b) and (7c) as potential utterances (among others). Here, (7b) uses a less round number, but (7c) is less informative. Hence a speaker who ranks INFO above NSAL (INFO > NSAL) will prefer (7b), while one who ranks NSAL above INFO will prefer (7c), as shown in Table 2.

**Table 2: Constraint violations of (7a-c) with respect to INFO and NSAL for quantity 101-109**

| | INFO | NSAL |
|---|---|---|
| more than 90 | | ** |
| more than 80 | * | * |

Generally speaking, (7a) will out-compete (7b) and (7c) for all situations in which the value is known to exceed 100 (although for higher values there may be still better options), for all speakers irrespective of their constraint ranking. A hearer, then, may correctly infer from the utterance of (7b) or (7c) that (7a) does not hold: that is, "more than 80” and “more than 90” both reliably implicate “not more than 100”. However, a hearer may not infer with certainty that the utterance of (7c) indicates that (7b) does not hold: this may merely reflect a preference for the rounder number. That is, “more than 80” does not reliably implicate “more than 90”, although a hearer should be aware that this might be the correct inference to draw, depending on the disposition of the speaker. All the options that incur more extensive constraint violations (“more than 50”, for instance) are irrelevant for the purposes of implicature, as they could not be uttered by a speaker in this scenario.

If we consider a situation in which the value is in the range 151-199, options include (8a) and (8b).

(8a) more than 150

(8b) more than 100

If we assume that 150 is not as round as 100, there is no globally optimal candidate: speakers who rank NSAL > INFO are predicted to prefer (8b), while those who rank INFO > NSAL should prefer (8a), analogously to Table 2. Consequently, "more than 100" is predicted to be the preferred utterance of its type for describing values in the range 101-109, but also to be a possible utterance for describing values in the range 151-199. Hence "more than 100" should, on this account, correspond to a wide range of possible interpretations, while "more than 80" and "more than 90" are narrowly restricted in their interpretation by the presence of their highly round upper near-neighbour 100.

These observations are borne out by the experimental data of Cummins, Sauerland and Solt (submitted). In the absence of numerical prior context, round numbers such as 100 act as watersheds for implicature: “more than *n*” for $n < 100$ is reliably taken to implicate “not more than 100”. By contrast, the implicatures licensed by the utterance “more than 100” vary considerably, with participants typically accepting it as an expression of values up to 150 and sometimes up to 200. Cummins et al. observe that participants were, in some cases, explicitly aware of the role of numeral salience in shaping the appropriate pragmatic interpretation, and in some cases remarked that a particular higher round number would have been used if it had been factually correct to do so. It appears, then, that appeal to considerations of numeral salience and informativeness provides a reasonable way of characterising hearers' interpretations of expressions such as "more than *n*".

**4. Effect of expressions in the preceding context**

By appeal to priming constraints, the model also makes predictions about the effect of prior context on the hearer's interpretation of numerically-quantified expressions. In particular, by appeal to the notion of numeral priming (NPRI), it is possible to spell out how the prior use of a numeral is predicted to influence the pragmatic enrichments that the speaker draws.

Revisiting the example of a situation in which the relevant quantity is known to be 101-109, let us further suppose that the numeral 80 is activated in the preceding context (e.g. "We need to sell 80 tickets"). Semantically possible expressions of the relevant quantity ("We have sold Q tickets") again include (9a-c).

(9a) more than 100

(9b) more than 90

(9c) more than 80

With respect to the INFO and NSAL constraints, the situation remains as described earlier. However, the NPRI constraint is now also relevant: (9a) and (9b) each violate this, while (9c) does not. The resulting situation is shown in Table 3.

**Table 3: Constraint violations of (9a-c) with respect to INFO, NSAL and NPRI for quantity 101-109**

| | INFO | NSAL | NPRI |
|---|---|---|---|
| more than 100 | | | * |
| more than 90 | * | * | * |
| more than 80 | ** | * | |

It is no longer the case that (9a) harmonically bounds the alternatives. Speakers who rank NPRI above both NSAL and INFO are predicted to prefer (9c) to (9a). From the hearer's point of view, it follows that the utterance of (9c) should no longer predict the falsity of (9a): "more than 80" might be said of a situation in which "more than 100" is known also to be true. Consequently, the utterance of "more than 80" should not implicate the falsity of "more than 100".

This also coheres with the findings of Cummins, Sauerland and Solt (submitted). They observed that the prior mention of a numeral gives rise to a weaker pragmatic bound for the subsequent comparative quantifier: that is, the latter is judged to admit a wider range of interpretation than would be the case in the absence of this prior mention. This appears to indicate that expressions that involve numeral re-use are considered pragmatically appropriate in certain situations for which they would be pragmatically inappropriate in the absence of prior context. For instance, "more than 80" becomes appropriate for the expression of values above 100, and hearers acknowledge this by admitting that it might convey this meaning. Within the constraint-based model, the numeral priming constraint encompasses this finding. However, as discussed earlier, this might also be attributed to considerations such as QUD, in which case alternative constraints would have to be posited within this model.

## 5. The interplay of context and speaker's intention

In the account described here, the asymmetry between speaker and hearer lies in their treatment of contextual factors. The discourse context, along with the speaker's intention, feeds into the decision procedure which selects the optimal utterance. However, the role of the hearer is not, canonically, to reverse this procedure in order to calculate the speaker's intention plus the discourse context. Indeed, this is not necessary, because the hearer is already party to shared aspects of the discourse context. Instead, hearers are argued to exploit this existing contextual knowledge in order to infer the speaker's intention. In the example dialogue (6), repeated below as (10), it is clear that both interlocutors know whether or not the numeral 60 is contextually salient, or primed, and therefore this information is available both when coding and decoding B's utterance.

(10) A. This case holds (60) CDs. How many CDs do you own?

B. I own more than 60 CDs.

However, the hearer cannot be presumed to have full awareness of all the relevant aspects of context. For instance, the speaker may introduce a number into the discourse because it is contextually activated for them at a psychological level, for instance in the following explanation of the British political situation.

(11) Because the Tories won fewer than 326 seats, they had to form a coalition.

The correct inference for the hearer of (11) to draw is that 326 is a critical value, and that if the Tories had won 326 seats they would not have needed to form a coalition. This does not involve prior knowledge that the numeral 326 is a salient one in British politics. Instead, the hearer can infer the relevance of the numeral from the fact of its use. Similar examples from the domain of sports might include (12) and (13).

(12) Jamie Burnett was the first snooker player to record a break of more than 147.

(13) It is theoretically possible to score more than 36 runs in an over of cricket.

As discussed above, there is evidence that the prior mention of a numeral yields a broader range of interpretation. However, our intuitions about examples such as (11), (12) and (13) suggest that the characterisation of numerals as “primed” or “unprimed” based on their presence or absence in the preceding linguistic context is only part of the story. The seemingly natural interpretation of (11), (12) and (13) corresponds closely to the model's predictions for primed numerals, namely that there is a pragmatic upper or lower bound on the interpretation, but that this is not as tight a bound as would be predicted on the basis merely of numeral salience and informativeness. Yet this arises without any indications of priming in the preceding context – indeed, in these examples, even without any preceding context.

This observation implies that the hearer's role is not restricted to the recovery of speaker meaning given shared contextual information, but extends to the positing of contextual information in order to account for otherwise inexplicable or improbable usage choices on the part of the speaker. This also fits naturally with the structure of the constraint-based model. Recall that, in this model, informativeness is treated as a constraint, along with numeral priming, granularity and so on. Within this model, an implicature is an inference driven by the informativeness constraint, in that it arises from the awareness that - all things being equal - the speaker will attempt to use a maximally informative utterance. Parallel inferences should in principle be available driven by these other constraints: in the case of numeral priming, this is the inference that the numeral used is salient, and arises from the observation that - all things being equal - the speaker will attempt to use a salient numeral. The relatively marginal nature of the inference derived from numeral priming, compared to that from informativeness, may reflect the importance of informativeness to cooperative interaction, but is not an architectural feature of this model.

The putative availability of this form of inference has two potential consequences, one of which relates to the licensing conditions of the implicature-blocking use of comparative quantifiers, and one of which relates to the interpretation of experimental findings involving the interpretation of apparently de-contextualised utterances. I discuss these in the following sections.

## 6. "More than *n*" in implicature-free contexts?

Following the definition of Jansen and Pollmann (2001), numerals from 1-9 are all salient. Therefore, if we are to account for the lack of implicature from examples such as (14), we cannot appeal to considerations of numeral salience. In this example, the stronger alternative (15) would involve the use of at least as round a number ("four" being at least as round as "3" in Jansen and Pollmann's terms), and therefore there should be no obstacle to the standard implicature process.

(14) John has more than three children.

(15) John has more than four children.

Examples such as these are the paradigm cases of the apparent failure of implicature from comparative quantifiers. Therefore, if the account offered here is to be adequate, it must be able to explain why implicature fails in these examples.

In this section I discuss three possible categories of context in which an utterance such as (14) could arise. The hypothesis I put forward is that every implicature-free utterance of a comparative quantifier can be identified as arising from a context belonging to one (or more) of these three categories. For each category of context I show how the implicature "John has not more than four children" fails to arise on standard Gricean principles.

### *6.1 Speaker's uncertainty as to the precise value*

In the standard account of implicature derivation from (14), it is assumed that the speaker has full information, i.e. that they are knowledgeable about the topic under discussion, and that their failure to make a more informative statement is not a direct consequence of their epistemic inability to do so.

Nevertheless, it appears quite possible for a speaker not to have full information when formulating an utterance such as (14). The precise extent of their knowledge will then become relevant in the interpretation of their utterance. Recall that in discussing the use of "more than 100", "more than 80" etc., I referred to the speaker knowing that a value was in a certain range, and did not consider whether they possessed more precise information. However, in cases such as (14), this becomes significant, because the immediate neighbours of the numeral *n* are also highly salient and participate in competing alternative utterances.

To exemplify this, suppose that a speaker wishes to express a quantity in the range 104-109[5]. They could say "more than 103", respecting informativeness, or "more than 100", respecting numeral salience. If they wished to express a quantity in the range 105-108, the corresponding choice would be between "more than 104" and "more than 100". In both cases, the option that satisfies numeral salience remains the same: “more than 100”. By contrast, if the speaker is talking about a range 4-9, "more than three" respects both informativeness and numeral salience. If the range is 5-8, "more than four" accomplishes both these goals, and harmonically bounds "more than three". So, unless other contextual factors are at play, the use of "more than three" gives rise to the weak implicature that the speaker is not in a position to assert "more than four".

---

[5] Once again, other quantifiers could be used, but these would arguably violate quantifier salience.

Moreover, in the absence of prior context, a speaker who possesses precise information ("John has (exactly) four children") is predicted to make a precise statement[6]. I assume that the use of a bare numeral is less costly than that of a comparative quantifier (in the model, it is favoured by the quantifier simplicity constraint), and that its precision is also favoured by the informativeness constraint. Therefore, the use of "more than three" apparently out of the blue should signal that the speaker either does not possess precise information, or is responding to contextual cues of which the hearer is not aware.

If no contextual factors of any type are intruding, the utterance of "more than three" is thus predicted to give rise to the weak implicature that the speaker is not in a position to assert "more than four". However, it should also signal that the speaker is not in possession of more precise information. This blocks the further stages of inference that would be required to draw the strong implicature that "more than four" does not hold. The predicted inference from (14) in this case is that the speaker considers it possible that John has exactly four children.

This aspect of the proposal is similar to the accounts of "at least (four)" proposed by Geurts and Nouwen (2007) and Cummins and Katsos (2010). Indeed, on this account, "more than three" behaves just like "at least four" in such cases. This is, however, not predicted to apply when the use of "more than three" reflects the action of numeral priming, as discussed in the following subsections.

### *6.2 Low-level numeral/quantifier priming*

[6] I assume here that bare numerals convey punctual meanings, following the analysis of Krifka (2009) and the argumentation of Breheny (2008).

According to the constraint-based model, the use of "more than three" might also reflect the action of priming effects, both of the numeral and the quantifier. An artificial example of this is the dialogue (16).

(16) A. Does John have more than three children?

B. I think he does have more than three children.

Here, B's use of "more than three" could be regarded as the consequence of an automatic alignment process, in the sense of Pickering and Garrod (2004) – that is, the use of a particular chunk by A renders it especially accessible for B. The constraint-based account would consider this as a case of the preferential use of a form that satisfies numeral and quantifier priming. Still another approach might be to regard B's choice of expression as the most direct possible answer to the question under discussion (Roberts 1996), which would be compatible with a relevance-theoretic account (and could also be modelled by the addition of a QUD constraint, as touched upon earlier)[7].

All the above accounts share the prediction that B's choice of utterance is conditioned by the prior context rather than the need to make a maximally informative statement. In such cases, the hearer A is clearly aware of the relevant preceding context (A's own utterance), and can use this to modulate the pragmatic inferences arising from B's choice of utterance. As stronger alternative statements that B might have made are disfavoured with respect to this prior context, no conclusions about their falsity can be drawn, even if

[7] Note however that Roberts (1996) does not appear to commit to the view that the most direct answer to the QUD is preferable.

B is presumed to be knowledgeable on the topic under discussion. Consequently, in such cases, no inference is predicted to arise.

A related class of usages are those in which an individual quotes a quantified expression that they have heard previously (e.g. "B said that John has more than three children"). These can be analysed either as a case of explicit priming, based on the fact of the reported earlier utterance having been made in the speaker's presence, or as a case of limited knowledge on the part of the current speaker. In either case, no inference is predicted to be available.

### *6.3 High-level numeral priming*

In the above two subsections, I discuss cases where the speaker's knowledge is incomplete, and in which the speaker's choice of utterance is influenced by low-level priming constraints. However, a separate class of examples includes cases such as (17). In these, no explicit context is mentioned and the speaker's knowledge is shown to be perfect on the topic under discussion, yet the use of an imprecise expression involving a comparative quantifier is judged acceptable.

(17) John has more than three children; in fact, he has five.

The issue here is that the first clause cannot give rise to an implicature (to the effect that John has not more than four children), because the second clause would contradict that, rendering the utterance incoherent. Moreover, the second clause demonstrates that the speaker knows the precise quantity under discussion, so the implicature cannot fail on epistemic grounds. So how can the first clause of (17) be acceptable? To answer this, we note that, insofar as this utterance is naturalistic, it does intuitively appear to require some kind of licensing context. According to the constraint-based model, such a context would be any one that caused the first clause of (17) to be optimal with respect to a priming constraint. This might include cases in which the numeral is previously explicitly mentioned, as in (16). Other possible motivations include the existence of circumstances under which the numeral mentioned is known to be a critical value for some purpose. For instance, perhaps having three or more children entitles an individual to extra benefits, or maybe three children is as many as will fit in John's car.

If a context of this kind is present, the hearer is aware that the choice of utterance may reflect the need to satisfy a priming constraint, and thus that it does not reflect the speaker's lack of access to stronger statements. It is thus predicted that no implicature will result. Consequently, no pragmatic contradiction arises from the subsequently manifest fact that the speaker has more precise knowledge. The apparent puzzle of why B risks giving rise to inappropriate implicatures by using this weaker expression (e.g. the implicature that it is possible that John has exactly four children) is resolved: the hearer is immediately aware that "more than three" is not a trigger for implicature processing in such a case, so there is no risk of misapprehension.

The problem that remains is how a participant in an experimental situation, presented with (17) and no context, could also arrive at the conclusion that the utterance is coherent (as they do; see Cummins and Katsos 2010: 292), implicitly failing to draw an implicature from the first clause. As discussed earlier in this paper, I argue that this arises because the participant infers a context for the utterance in which this apparent circumlocution makes sense: that is, one in which ‘three children’ constitutes some kind of critical value. From the point of view of the constraint-based approach, this arises because the hearer is logically able to draw inferences about the prior context from the utterance. In a more theory-neutral setting, we could see this as a natural attempt to accommodate an utterance by making reparatory inferences, which appears to be characteristic of natural language usage. After all, outside a laboratory setting, it is not usually a communicatively useful response to a linguistic stimulus merely to judge it as ill-formed. In the following section of this paper I discuss the implications of this for experimental work in this area.

### *6.4 Summary*

On the constraint-based model, the use of an expression such as "more than three" in a cardinal quantifying context might arise for three reasons. One, because this reflects the speaker's best knowledge but this knowledge is incomplete (in which case the implicature should be available that the speaker considers "exactly four" to be possible). Two, because this reflects low-level priming from a prior mention of the numeral in the preceding discourse. Three, because this reflects higher-level awareness of the status of the numeral as a critical value for the topic under discussion. In cases two and three, the absence of a strong implicature ("not more than four") is predictable for the standard pragmatic reason that the stronger alternative is not (necessarily) an admissible utterance in the context. The model claims, in effect, that all usages of "more than three" in unembedded declarative contexts can be identified as belonging to one of these categories, and thus that the strong implicature is never available. This appears to cover all the cases discussed in the literature, and I am not aware of any counterexamples, although the question of whether it genuinely encompasses all such cases in real-life usage will be a more laborious one to settle.

**7. Inferring context based on the utterance in experimental settings**

In their experiments, Cummins, Sauerland and Solt (submitted) demonstrate that the range of interpretation of numerically quantified expressions such as "more than 100", "fewer than 60" etc. is conditioned by numeral salience. However, they note that the observed behaviour differs from this prediction for the least salient numerals. For instance, "more than 93" is interpreted as meaning "not more than 100", rather than "not more than 94". This coheres once again with the observation that this implicature would render "more than 93" equivalent in meaning to "exactly 94", violating our intuitions.

We can account for this by generalising the argument put forward in the previous section. Although $n$ = 93 is not salient or round by any reasonable definition, it is no less salient than $n$+1: the local landscape of numeral salience is just as flat for $n$ = 93 as for $n$ = 3. It should therefore follow that the use of "more than 93" either implicates the possibility (as far as the speaker is concerned) that "exactly 94" holds, or indicates that 93 is itself a salient numeral.

This raises difficulties in connection with the experimental investigation of the pragmatics of these expressions. For instance, Cummins, Sauerland and Solt (submitted) use tasks in which numerically-quantified statements are presented and participants are asked to interpret the quantities being referred to. In their experiment that demonstrates the effect of prior mention of the numeral on interpretation, the preceding context is explicitly manipulated by including or excluding the explicit mention of the numeral, as shown in example (6) here, repeated below as (18).

(18) A. This case holds (60) CDs. How many CDs do you own?

B. I own more than 60 CDs.

In interpreting the resulting data, it is assumed that the explicit previous mention of the numeral leads to priming, and in the absence of this the numeral is unprimed. However, given the discussion above, this assumption seems likely to be an oversimplification. Instead, given a usage that appears infelicitous, for instance in which a non-round number is used without prior motivation, the participant seems likely to posit a licensing context in order to justify the use of the utterance, and is therefore liable to interpret the quantifier against the background of this assumed context.

As applied to this paradigm, it seems likely that the postulation of context will exert a damping effect on the results. Firstly, it blurs the distinction between primed and unprimed contexts, which suggests that the effect size of priming as measured by this experiment is likely to be an underestimate. Secondly, the propensity of participants to furnish contexts in this way is presumably going to be greater in cases involving non-round numbers, which are less likely than round numbers to surface in general contexts (Jansen and Pollmann 2001), and which therefore should trigger a strong expectation that their use is somehow motivated on contextual grounds. Non-round numbers are exactly the numbers for which the predicted implicatures are strongest (that is, they should give rise to a narrow range of interpretation). So the greatest effect of implicature attenuation is predicted to apply to the strongest implicatures. This suggests that the effect of manipulating numeral salience is also likely to be understated in these experimental results. That is to say, although Cummins, Sauerland and Solt (submitted) already obtain significant results for priming and numeral salience, contextual considerations suggest that the true effect of these may be greater than they have measured.

Generally the possibility that participants are likely to invent contexts would also give rise to the expectation that participants will vary considerably in their behaviour, not just on the basis of cognitive preferences about quantifier usage and interpretation (as modelled by the constraint-based account) but also according to the extent to which they do imagine contexts of this kind, and the types of context that they are disposed to imagine. This suggests that an accurate model of usage and interpretation would nevertheless tend to under-predict the variability elicited experimentally, as this would depend in part upon considerations idiosyncratic to the hearer that could not readily be treated by such a model.

This line of argument has points in common with the remarks of Geurts and Pouscoulous (2009: 14-15) concerning the doubtful validity of implication judgement tasks. They observe that "the very question whether [a consequent] might be implied changes the context in which [its antecedent] is interpreted: the question makes it relevant to decide whether or not the speaker believes that [the consequent] might be true". While the presentation of the premises in the paradigm they discuss is entirely decontextualised, the very fact of this presentation nevertheless invites pragmatic intrusion. Similarly, I argue here that the presumed fact of an utterance being made has the potential to change the context against which it is interpreted. If the utterance requires a special context in order to surface at all, that is necessarily the context against which it is liable to be interpreted. Under this account, it could be argued that there is no such thing as a neutral context for certain numerically quantified expressions, potentially even those as simple as “more than three”. It could also be argued that there are more and less typical contexts for all sorts of expressions, and the precise interpretation arrived at may depend on which one is postulated by the participant.

However, this observation does point in a different direction to that of Geurts and Pouscoulous (2009) with respect to some experimental work. For instance, in the implication judgements tested by Geurts et al. (2010), “at most two” reliably fails to entail “at most three”, in the judgement of participants, even though this is logically sound under the classical meaning of the superlative quantifier (merely “less than or equal to”). These experiments involved the presentation of pairs of written sentences. According to Geurts and Pouscoulous, this juxtaposition of sentences renders participants more likely to look for an entailment relation than they necessarily would if confronted with the antecedent ‘in the wild’. However, if the presentation of sentences causes the participant spontaneously to look for contexts of utterance for these forms, then the participants’ failure to endorse the entailment relation might readily be explained, because there is (on the constraint-based account) no context of which “at most two” and “at most three” may at once be felicitously uttered, even by different speakers. (“At most three” is harmonically bounded by “fewer than three” in any situation of which “at most two” is true.) Indeed, Cummins and Katsos (2010: 288-91) conduct a similar experiment in which “at most two” is adjudged to imply “at most three” by a majority of participants, which merely involves the embedding of the superlative quantifier into a conjunctive sentence that justifies its use.

More generally this observation appears to raise questions about the logical possibility of obtaining experimental data that bears unambiguously upon the semantics of expressions. If we do not provide a specific context against which utterances can be interpreted, it seems possible that participants will furnish their own contexts and interpret the utterances against these. On a reductionist view of semantics this may be wholly appropriate – we could argue that the tendency of an expression to arise in a particular context is an indivisible part of its meaning. However, for most purposes it seems desirable to exert tighter control than this.

The difficulty of constructing neutral contexts for the interpretation of utterances is a known problem in the experimental pragmatics literature: for instance, Breheny, Katsos and Williams (2006: 445) observe that "even single sentence utterances can create their own context through a variety of presupposition triggers and information-structure triggers". Under the account discussed here, the problem is even more widespread, extending to ostensibly unmarked declaratives and expressions as simple as "more than *n*", merely on the basis that their utterance gives rise to a set of presuppositions as to the content of the preceding discourse. The results of Cummins, Sauerland and Solt's study suggest that this is not only a theoretical concern but also an observable phenomenon for certain numerically quantified expressions. Ultimately, the proper experimental semantic analysis of such an expression might require it to be presented not only in a plausible sentence, but within a plausible discourse, in which the contextual factors can be controlled and analysed.

## 8. Conclusion

The constraint-based model discussed in this paper appears to make the correct predictions about the availability of implicatures from comparative quantifiers, and their tendency to be attenuated by prior mention of the numeral. Moreover, appeal to the priming constraints of the model, as well as traditional pragmatic principles, enables us to account for the systematic lack of implicatures from expressions such as “more than three” and those involving low-salience numerals, in cardinal contexts. However, this model also makes predictions about the possible availability of inferences concerning the nature of prior context, thus suggesting that the interplay between context, intention, utterance and interpretation is more complex than we might initially suppose. Participants’ behaviour in experiments, and our own intuitions about expressions that introduce salient numerals, tend to support the idea that hearers will supply contexts in a *post hoc* attempt to justify the usage of otherwise inexplicable choices of expression. This raises interesting questions about the interpretation of experimental findings in this domain, and suggests that it may be difficult to design experiments that unambiguously resolve aspects of the ongoing demarcation dispute between semantics and pragmatics.

## Acknowledgements

This research was partially funded by a University of Cambridge Domestic Research Studentship. The author also acknowledges the support of the EURO-XPRAG and

XPrag-UK networks.  No funding sources played any role in determining any aspect of this research.

## References

Blutner, R. (2006). Embedded implicatures and optimality theoretic pragmatics. In T. Solstad, A. Grønn and D. Haug (eds.), *A Festschrift for Kjell Johan Sæbø: in partial fulfilment of the requirements for the celebration of his 50th birthday*. Oslo: University of Oslo.

Bonnefon, J.-F., Feeney, A. and Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112: 249-258.

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

Breheny, R., Katsos, N. and Williams, J. (2006). Are scalar implicatures generated by default? *Cognition*, 100: 434-63.

Cummins, C. (2011). The interpretation and use of numerically-quantified expressions. PhD thesis available from http://www.dspace.cam.ac.uk/handle/1810/241034.

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

Cummins, C., Sauerland, U. and Solt, S. (submitted). Granularity and scalar implicature in numerical expressions.

Dehaene, S. (1997). *The Number Sense*. New York: Oxford University Press.

Fox, D. and Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, 29: 537-86.

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

Geurts, B., Katsos, N., Cummins, C., Moons, J. and Noordman, L. (2010). Scalar quantifiers: logic, acquisition, and processing. *Language and Cognitive Processes*, 25: 130-48.

Geurts, B. and Nouwen, R. (2007). 'At least' et al.: the semantics of scalar modifiers. *Language*, 83: 533-59.

Geurts, B. and Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and Pragmatics*, 2(4): 1-34.

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Schiffrin, D. (ed.), *Meaning, Form and Use in Context (GURT '84)*. Washington DC: Georgetown University Press. 11-42.

Jansen, C. J. M. and 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. Oxford: Elsevier. 257-92.

Krifka, M. (2009). Approximate interpretations of number words: a case for strategic communication. In Hinrichs, E. and Nerbonne, J. (eds.), *Theory and Evidence in Semantics*. Stanford: CSLI Publications. 109-132.

McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

Pickering, M. J., and Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27: 169-226.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. In Yoon, J.-H. and Kathol, A. (eds.), *OSUWPL Volume 49: Papers in Semantics*. Columbus, OH: Ohio State University Department of Linguistics.