***Between 3 and 5* sometimes means *at least 3* – new ways to detect a new ambiguity.**
Bare numerals, when used as quantifiers, are known to be ambiguous between an 'exact', upper-bounded reading, and an 'at-least' reading. As explained below, under certain natural assumptions, the mechanisms responsible for this ambiguity are expected to yield a similar ambiguity for modified numerals of the form *between n and m.* While this ambiguity seems to be at odds with naïve intuitions, we provide experimental evidence for its existence. Our contribution is thus both theoretical and experimental. On the theoretical side, we show that some abstract semantic mechanisms which might be thought to overgenerate in fact make correct predictions. On the experimental side, we present two different experimental designs which we argue are able to detect ambiguities.

Theoretical Background. There is no agreement in the literature regarding the *source* of the ambiguity of bare numerals. The traditional, neo-Gricean view, takes the 'at-least' reading to be the basic, literal reading of bare numerals, and the 'exact' reading to be derived as a scalar implicature. According to some other views, numerals are intrinsically ambiguous between the two readings (e.g., Geurts 2006) or only have an exact reading as far as their literal meaning is concerned (Breheny 2008). We assume that both readings are derived from a more basic, predicative reading (cf. (1)), by means of two general type shifting operations which turn it into a determiner (of type $\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$), *existentialisation* and *maximization* (cf. (2) – this is similar in spirit but not in letter to Geurts' proposal; while the specifics of our implementation do not really matter given our main goal in this paper, we provide them for the sake of explicitness).

(1) Basic predicative meaning: $[\![\text{three}]\!] = \lambda x.\#x = 3$ (where $\#x$ denotes the number of atoms which are part of the (possibly) plural individual x, given standard mereological assumptions)

(2) Type shifting operations:
a. Existentialization: $\mathbf{E}\,(P_{\langle e,t\rangle}) = \lambda Q.\lambda R.\exists x(P(x) \wedge Q(x) \wedge R(x))$ $\quad_{\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle}$
b. Maximization: $\mathbf{M}(P_{\langle e,t\rangle}) = \lambda Q.\lambda R.P(\max\{y : Q(y) \wedge R(y)\})$ $\quad_{\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle}$
(where max, when applied to a set of plural individuals, returns the unique member of that set, if there is one, such that every member of the set is part of it, e.g., $\max\{x, y, x \oplus y\} = x \oplus y$).

The two readings of *Three linguists are bald* are derived as follows:

(3) a. Applying existentialization gives rise to the at-least reading:
$[\![[(\mathbf{E}(\text{three}))\text{ linguists}]\text{ [are bald]}]\!] = \exists x(\#x = 3 \wedge [\![\text{linguists}]\!]\,(x) \wedge [\![\text{bald}]\!]\,(x))$
⇝ True iff at least three linguists are bald, assuming linguists and bald are distributive predicate. (If 5 linguists are bald, one can find a plural individual x made up of 3 of them making the above formula true).
b. Applying maximization gives rise to the exact reading:
$[\![[(\mathbf{M}(\text{three}))\text{ linguists}]\text{ [are bald]}]\!] = [\![\text{three}]\!]\,(max\{y : [\![\text{linguists}]\!]\,(y) \wedge [\![\text{bald}]\!]\,(y)\})$
⇝ The maximal plural individual made up of linguists who are bald contains 3 atomic parts, i.e. exactly 3 linguists are bald.

Now, let us assume the following basic predicative meaning for *between n and m*:

(4) $[\![\text{between 3 and 5}]\!] = \lambda x.3 \leq \#x \leq 5$

Applying existentialization and maximization to *between 3 and 5* yield the following results for *Between 3 and 5 linguists are bald*:

(5) a. Existentialization makes the sentence equivalent to *At least 3 linguists are bald*:
$[\![[(\mathbf{E}(\text{Between 3 and 5}))\text{ linguists}]\text{ [are bald]}]\!] = \exists x(3 \leq \#x \leq 5 \wedge [\![\text{linguists}]\!]\,(x) \wedge [\![\text{bald}]\!])$
⇝ True iff three or more linguists are bald. Suppose for instance that 7 linguists are bald. Then there is a plurality made up of bald linguists whose cardinality is between 3 and 5.
b. Maximization gives rise to the 'standard', expected reading:
$[\![[(\mathbf{M}(\text{Between 3 and 5}))\text{ linguists}]\text{ [are bald]}]\!] = [\![\text{betw 3 and 5}]\!]\,(max\{y : [\![\text{ling.}]\!]\,(y) \wedge [\![\text{bald}]\!]\,(y)\})$
⇝ The maximal plural individual made up of linguists who are bald satisfies the property *between 3 and 5*, i.e the number of bald linguists is at least 3 and at most 5.

The two sets of experiments below provide evidence for this predicted (but intuitively surprising) ambiguity.

Experiments 1a, 1b, 1c: picture-matching task using graded judgements. These experiments were sentence-picture matching tasks, in which subjects had to tell us, using a continuous scale, the extent to which the sentence was a correct description of the picture. Answers were coded as the position of the response on the scale, from 0% for a rejection and 100% for acceptance of the sentence as a correct description. The pictures were arrays of dots with different colors. All three experiments contained control sentences of the form *At least x/At most y dots are red*, which are not ambiguous. Judgments for these control sentences were as expected: there were small discrepancies for *at most* sentences that are in line with previous results (e.g., Katsos & Cummins 2010), but the overall mean accuracy is 81% (if accuracy = 'raw score' for expected true answers, and '100 - raw score' for expected false responses).

In **Exp. 1a** (16 subjects), the target sentences were of the form *Between n and (n+2) dots are red*, with $n = 3$ or $n = 4$. For the target sentences, we distinguished 3 types of conditions as follows. **(a) False**: the sentence is false on both readings in (5), i.e. the picture contains fewer than n red dots. **(b) True**: the sentence is unambiguously true (n, $n + 1$ or $n + 2$ red dots in the picture). **(c) Target**: (5-a) is true but

(5-b) is false (more than $n+2$ red dots). In the **target** condition, we obtained a rating (33%) intermediate between the **false** (5%) and **true** (85%) conditions (pairwise comparisons: all $ts > 3.2$, $ps < .01$). Such a result is expected if the sentence is indeed ambiguous between two readings, only one of which being true in the **target** condition (see also Chemla & Spector 2011).

In order to confirm our interpretation, we ran a control **experiment 1b** (11 new subjects). We used exactly the same design, except that we replaced the previous target sentences with new ones of the form *The number of red dots is comprised between n and* $n+2$, which can only be given the upper-bounded reading, even with the above analysis. We obtain the following ratings: **false** = 3%, **true** = 86%, **target** = 10%, with no significant difference between **false** and **target** Furthermore, a 3×2 ANOVA that compared mean responses in Exp. 1a and 1b revealed a significant interaction between condition (true/false/target) and experiment/type of target sentence: $F(2,75) = 5.96, p < .01$. This result suggests that the two constructions are different, and confirm that the first kind of *between* sentences, but not second kind, is ambiguous.

Finally, in **Exp. 1c** (11 new subjects), we kept exactly the same design but replaced the target sentences with sentences containing bare numerals, of the form *n dots are red*. As expected, we obtained an ambiguity type of pattern **false** = 4%, **true** = 99%, **target** = 78% (the target condition now corresponds to pictures with more than n red dots, i.e. pictures making the at-least reading true but the exact reading false). All pairwise comparisons yield significant differences (all $ts > 3.7$, $ps < .01$).

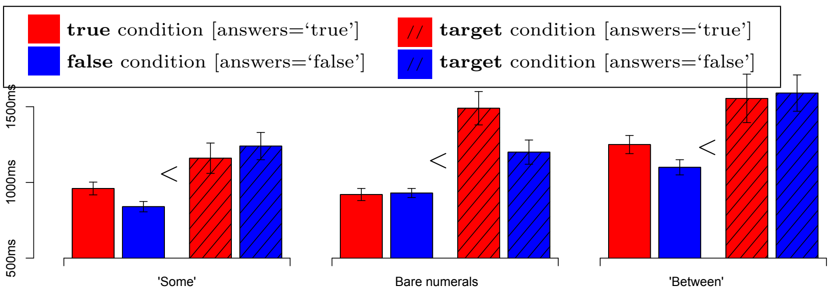
Our results also show that the lower-bounded reading is less salient in the case of *between n and m*-sentences than in the case of sentences involving bare numerals, since the **target** conditions got a significantly lower rating in the former case than in the latter case. This result is in line with introspective judgments: (5-a) had never been observed before. This difference should receive an independent explanation, e.g., a pragmatic one: using 'between n and m', instead of a more simple expression (such as a bare numeral or an 'at least'-numeral), in order to convey the 'at least' reading might be a violation of Grice's maxim of manner.

Experiment 2: response time study. Our second study was also based on a sentence-picture matching task, but aimed at gathering response time data. Instead of using a continuous scale, we asked 33 subjects to provide binary answers (*true* or *false*). Our experimental hypothesis was the following: when a sentence is ambiguous between two readings $R1$ and $R2$, response times will be greater (everything else being equal) when the picture makes $R1$ true and $R2$ false than in cases where the picture makes either both $R1$ and $R2$ true or both $R1$ and $R2$ false. Quite generally, if several responses are in principle acceptable, participants will hesitate between them and get slowed down, no matter which response they eventually choose. Notice that this hypothesis does not require awareness of the ambiguity, but merely requires that different aspects of the stimulus push participants in different directions, making the decision process harder to terminate.

For our target sentences, we thus expected response times to be higher for the **target** condition than for the other two unambiguously true/false conditions. In order to motivate this interpretation of our results, we also tested other cases which are known to lead to similar ambiguities, namely sentences involving scalar implicatures (*some dots are red*) and sentences involving bare numerals (*n dots are red*). We constructed similar **false**, **true** and **target** conditions by varying the number or proportion of red dots in the picture.

<u>Results:</u> The response times given in the figure below confirm our expectations. For the three types of ambiguities, we found that correct responses to the **true** (■) and **false** (■) unambiguous conditions were faster than the 'true' (//) and 'false' (//) responses to the **target** condition ($ts > 2.4$, $ps < .01$): the first two bars are always shorter than the last two.

Two comments are in order. **(i)** These results are orthogonal to differences that have already been noticed. For instance, in the **target** conditions for scalar implicatures, we also detect a difference between the RTs for true and false answers (readings with scalar implicatures are slower, as in Bott & Noveck 2004 and subsequent studies). **(ii)** These differences could not be accounted for only in terms of the properties of the relevant pictures, independently of the sentence they are paired with (for instance, the fact that in **target** conditions, the picture contained more dots that in the **true** and **false** conditions), since all the pictures were also tested with clearly unambiguous sentences (*At least/At most*) and no similar difference was found.



Response times for correct answers in the unambiguous **true** and **false** conditions, and for both 'true' and 'false' answers in the **target** condition.

Main result:
answers in **target** condition
are slower than in
the two other conditions.

Conclusion. Plausible formal semantic approaches to bare and modified numerals predict the possibility of a lower-bounded reading for 'between n and m'-sentences, despite the presence of an explicit upper-bound. We offered two types of experimental evidence which confirm this surprising prediction.