An experimental investigation of superlative modifiers and modals*

Yaron McNabb
Universiteit Utrecht

Doris Penka
Universität Konstanz

Draft, November 24, 2014
Please do not cite without permission

Abstract

There has been much debate recently about the meaning of superlative modifiers like *at least* and *at most*. The main challenge analyses of superlative modifiers face is accounting for the ignorance implication they give rise to, whereby the speaker holds higher (in *at least*) or lower (in *at most*) numbers as possible. In this study, we present results from four experiments that test the interpretation of superlative modifiers, with special focus on the readings the give rise to when occurring in sentences with deontic modals. We show that the results of the experiments are only partially predicted by the various competing and incompatible analyses in the literature. The analysis that best accords with the data we present here is the neo-Gricean account put forth by Büring (2008); Schwarz (2011, 2013), and Kennedy (2013), according to which ignorance inferences arise as quantity implicature inferences, and fine-tuned by Penka (2014) to account for the interpretational difference between *at least* and *at most*. In addition to adjudicating between various analyses of superlative modifiers, our experimental investigation contributes to the methodological efforts to develop an experimental paradigm that explicitly examines ignorance inferences, tests subtle readings in complex constructions, and sheds light on the semantic-pragmatic divide.

1 Introduction

There has been much debate recently about the meaning of superlative modifiers like *at least* and *at most* (Geurts and Nouwen, 2007; Büring, 2008; Cummins and Katsos, 2010; Nouwen,

*We thank audiences at the universities of Konstanz, Tübingen, Chicago, Radboud Nijmegen, and ZAS Berlin, and at the conferences *XPRAG 2013, Sinn und Bedeutung 16, Linguistic Evidence 2014, CUNY 2014*, and the *Formal and Experimental Pragmatics Workshop* at *ESSLLI 2014*. We have especially benefited from conversations with Oliver Bott, Bart Geurts, Irene Heim, Itamar Francez, Chris Kennedy, Manfred Krifka, Rick Nouwen, Fabian Schlotterbeck, and Maribel Romero. Ali Eglimez, Vasiliki Erotokritou, Simone Reinke, and Bianca Veser were very helpful with with creating, editing and translating stimuli and running experiments. This research was supported by DFG grant PE 1765/2-1 and the Zukunftskolleg at the University of Konstanz.

2010; Schwarz, 2011, 2013; Cohen and Krifka, 2011; Coppock and Brochhagen, 2013b; Kennedy, 2013). As Geurts and Nouwen (2007) observe, the superlative modifiers *at least* and *at most* give rise to ignorance inferences: By using a superlative modifier, a speaker generally conveys that she is unsure about the precise value, e.g. *at least 15 pages* in (1a) implies that the speaker considers both exactly 15 and higher values possible and *at most 15 pages* in (1b) implies that the speaker considers 15 and lower values possible, but not values higher than 15.[1]

(1) a. The paper is **at least** 15 pages long.
b. The paper is **at most** 15 pages long.

Geurts and Nouwen (2007) and Büring (2008) also observe that when *at least* is embedded under a necessity modal like *have to* in (2), ignorance inferences can be suppressed. In these cases, a reading that Büring calls authoritative reading emerges: Under this reading, (2) does not convey speaker ignorance and rather specifies 15 pages as the minimally required paper length.

(2) Your term paper **has to** be **at least** 15 pages long.

In contrast, when *at least* is embedded under a possibility modal like *can* in (3), an authoritative reading does not seem to be available. The only possible reading of (3) seems to be one conveying speaker ignorance. Under this reading, which Büring calls speaker insecurity reading, the speaker is unsure about the maximally allowed paper length, and for all she knows the upper bound of permissible paper lengths might be 15 pages or more.

(3) Your term paper **can** be **at least** 15 pages long.

The main challenge analyses of superlative modifiers face is accounting for the ignorance implication they give rise to and the mechanisms that lead to the suppression of these inferences as well as to the availability of the authoritative reading in certain combinations with modals. We show in Section 2 that the various analyses that have been proposed for superlative modifiers differ greatly regarding the predictions they make about the derivation and suppression of ignorance inferences under modals.

Following the discussion of previous analyses, we present results from four experiments that test the interpretation of superlative modifiers, the first three in English and the fourth in German. In Experiment 1, we establish a methodology with which we experimentally explore the meaning of superlative modifiers. In Experiment 2, we test which combinations of modals and superlative modifiers can obviate speaker ignorance. In Experiment 3, we investigated which readings are available for each one of the superlative modifier-modal combinations in terms of lower and upper bound of permissible values. In Experiment 4 we replicate the task used in Experiment 3, this time using an incremental self-paced reading methodology in order to examine the time-course of processing these combinations. We then adjudicate between the competing analyses given the

[1]We assume that there is no distinction between numerals like *fifteen* and measure phrases like *fifteen metres* in terms of how they combine with superlative modifiers. This will become apparent in our choice of experimental stimuli, which include superlative modifiers modifying bare numerals as well as measure phrases, either in attributive or predicative position.

results of the four experiments and show that the analysis that best accords with the data we present here is the neo-Gricean account put forth by Büring (2008); Schwarz (2011, 2013), and Kennedy (2013), according to which ignorance inferences arise as quantity implicature inferences, and fine-tuned by Penka (2014) to account for the interpretational difference between *at least* and *at most*.

2 Analyses of speaker ignorance in superlative modifiers

This section provides an overview over existing accounts of the ignorance inferences arising with superlative modifiers.[2] Because the interaction of superlative modifiers and modals is a crucial test case for any theoretical analysis, we will pay particular attention to the predictions the different analyses make in this respect and largely disregard in the discussion additional problems of conceptional nature that some of the analyses might face (for a a discussion of these see in particular Coppock and Brochhagen (2013b)).

2.1 Geurts and Nouwen (2007)

Geurts and Nouwen (2007) account for the fact that utterances with *at least* or *at most* generally convey speaker ignorance by incorporating epistemic modality into the lexical entries of superlative modifiers. The truth conditions their analysis assigns to sentence (4a) involving *at least* are shown in (4b), where $\boxdot$ and $\diamonddot$ symbolize epistemic necessity and possibility, respectively. The sentence in (4a) means that the speaker is certain that the paper is (at least)[3] 15 pages long and considers it possible that the paper is longer than 15 pages.

(4) a. The paper is **at least** 15 pages long.
  b. $\boxdot\ [\textsc{length}(p) \geq 15pp] \wedge \diamonddot\ [\textsc{length}(p) > 15pp]$

The sentence in (5a) with *at most* means that the speaker considers it possible that the paper is (at least) 15 pages long and does not consider it possible that the paper is longer than 15 pages, i.e. the speaker is certain that the paper is not longer than 15 pages.

(5) a. The paper is **at most** 15 pages long.
  b. $\diamonddot\ [\textsc{length}(p) \geq 15pp] \wedge \neg\diamonddot\ [\textsc{length}(p) > 15pp]$

Regarding the interaction with modals, Geurts and Nouwen assume that superlative modifiers obligatorily take wide scope over deontic modals. This follows from their analysis of superlative modifiers as epistemic operators and the observation that epistemic operators generally outscope deontic ones. They also assume a rule of modal concord, which strips off the layer of epistemic modality in case the primary epistemic operator in the lexical entry of the superlative modifier

[2] We leave out the account of Cohen and Krifka (2011), who analyze superlative modifiers as illocutionary operators, as it is not clear to us how their account applies to cases where superlative modifiers interact with deontic modals. For a detailed discussion of Cohen and Krifka (2011), see Coppock and Brochhagen (2013b).

[3] Throughout the paper and unless otherwise noted we assume a one-sided semantics of numerals and a downward monotone semantics of gradable adjectives.

matches the modal force of the deontic modal. This predicts that an authoritative reading not conveying speaker ignorance arises if *at least* is combined with a necessity modal and if *at most* is combines with a possibility modal. The first case is illustrated in (6). The reading resulting from modal concord in (7a) says that it is deontically necessary that the paper be (at least) 15 pages long and lengths above 15 pages are deontically possible. This is the authoritative reading, according to which 15 pages is the minimally required length of the paper. The truth conditions for this reading are schematically illustrated in (7b), where the straight line signifies the range of permissible paper lengths, which we will also call the deontic range.

(6) The paper **has to** be **at least** 15 pages long.

(7) Authoritative reading resulting from modal concord:

a. $\Box\ [\textsc{length}(p) \geq 15pp] \wedge \Diamond\ [\textsc{length}(p) > 15pp]$

b.
```
                 |——— . . .
13          15          17
```

As Geurts and Nouwen assume that modal concord is optional, a compositional construal is also possible yielding the speaker insecurity reading. Under this reading, the sentence in (6) conveys that the speaker is unsure about the minimally-required length of the paper. The speaker is only certain that the lower bound of the deontic range is not below 15 pages, but for all she knows it might be higher than 15 pages. The truth conditions for this reading are given in (8a) and illustrated in (8b), where the shaded area (marked with forward slashes) signifies the epistemic range, i.e. the range of values that for all the speaker knows might or might nor be permissible.

(8) Speaker insecurity reading resulting from compositional interpretation:

a. $\boxdot\Box\ [\textsc{length}(p) \geq 15pp] \wedge \text{⟐}\Box\ [\textsc{length}(p) > 15pp]$

b.
```
            [///////////////——— . . .
13          15          17
```

For *at most*, modal concord is possible when combined with a possibility modal. Under the modal concord reading shown in (10), sentence (9) says that paper lengths of up to 15 pages are deontically possible while higher lengths are not possible; that is, 15 pages is the maximally allowed paper length or the upper bound of the deontic range. Under the compositional reading in (11) the sentence says that the speaker is unsure about the maximally allowed paper length. While she is certain that the upper bound of the deontic range is not above 15 pages, for all she knows, it might be 15 pages or less.

(9) The paper **can** be **at most** 15 pages long.

(10) Authoritative reading resulting from modal concord:

a. $\Diamond\ [\textsc{length}(p) \geq 15pp] \wedge \neg\Diamond\ [\textsc{length}(p) > 15pp]$

b.
```
 . . . ———|
13          15          17
```

(11) Speaker insecurity reading resulting from compositional interpretation:

a. ⟐◇ [LENGTH(p) ≥ 15pp] ∧ ¬⟐◇ [LENGTH(p) > 15pp]

b. . . . ——|//////////////]
13 15 17

In the remaining two combinations of superlative modifiers and modals, the epistemic modal in the superlative modifier and the deontic modal do not correspond in their modal force, and therefore modal concord is not possible and only a compositionally-derived reading conveying speaker ignorance is available. *At least n* combined with a possibility modal is predicted to mean that the speaker is unsure about the maximally-allowed number and thinks that the upper bound of the deontic range might be *n* or more.

(12) The paper **can** be **at least** 15 pages long.

(13) Speaker insecurity reading resulting from compositional interpretation:

a. ⊡◇ [LENGTH(p) ≥ 15pp] ∧ ⟐ ◇ [LENGTH(p) > 15pp]

b. . . . ——[////////////
13 15 17

At most n plus necessity modal says that the speaker is unsure about the minimally required number. While she is sure that the lower bound of the deontic range is not more than *n*, it might be *n* or less.

(14) The paper **has to** be **at most** 15 pages long.

(15) Speaker insecurity reading resulting from compositional interpretation:

a. ⟐□ [LENGTH(p) ≥ 15pp] ∧ ¬⟐□ [LENGTH(p) > 15pp]

b. ////////////]—— . . .
13 15 17

2.2 Nouwen (2010)

Rather than hard-wiring speaker ignorance into the lexical meaning of superlative modifiers as in Geurts and Nouwen (2007), Nouwen (2010) derives ignorance inferences from a covert epistemic possibility modal embedded under the superlative modifier. He proposes that superlative modifiers are degree operators indicating minima (for *at least*) or maxima (for *at most*):

(16) a. $[\![\text{at least}]\!] = \lambda n.\lambda D_{(dt)}.min\{d|D(d)\} = n$

b. $[\![\text{at most}]\!] = \lambda n.\lambda D_{(dt)}.max\{d|D(d)\} = n$

As degree operators, superlative modifiers are expected to scopally interact with modals (Heim, 2001). Nouwen's proposal builds on two additional assumptions. The first is that numerals are generally ambiguous between a lower- and a double-bounded meaning. Similarly, we can assume that for gradable predicates, both a functional (in terms of =) and a relational (in terms of ≥) meaning is generally available. Nouwen's second assumption is that linguistic expressions compete: If a certain meaning can be expressed by two or more expressions differing in their complexity, the simpler expression is preferred and more complex expressions are blocked. The components of the analysis, taken together, predict that in many cases superlative modifiers cannot be used because the resulting sentences either express a contradiction or a meaning that is equivalent to the sentence with a bare numeral and thus blocked. Consider for instance sentence (17a). Under Nouwen's assumptions, two different meanings are available depending on whether the weak (relational) or the strong (functional) meaning of the adjective *long* is used, yielding (17b) and (17c), respectively.

(17) a. The paper is **at least** 15 pages long.
 b. $min\{d|\textsc{length}(p) \geq d)\} = 15\text{pp}$
 c. $min\{d|\textsc{length}(p) = d)\} = 15\text{pp}$

The truth conditions in (17b) are contradictory, because the minimum of the set of degrees *d* such that the paper is at least *d*-long is always 1 (or less if fractions of pages are taken into account). This is avoided by using the strong meaning of the adjective in (17c), which refers to the minimum of a singleton set whose only element is the paper length. But (17c) is equivalent to the strong meaning of the sentence with the bare numeral in (18c) and should thus be blocked.

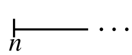
(18) a. The paper is 15 pages long.
 b. $\textsc{length}(p) \geq 15\text{pp}$
 c. $\textsc{length}(p) = 15\text{pp}$

The sentence with *at least* in (17a) is thus predicted not to express any sensible meaning. To rescue the sentence, Nouwen (2010) argues that a covert epistemic possibility modal can be inserted in the scope of the superlative modifier as shown in (19a). If the speaker is unsure about the length of the paper, i.e. the paper length varies across the worlds epistemically accessible to the speaker, the superlative modifier applies to a degree property denoting a range of values. Under the strong meaning of *long* this results in the non-contradictory truth-conditions (19b), saying that in the epistemically accessible world where the paper is shortest, it is exactly 15 pages long. Since this is a meaning not expressed by the bare numeral it is not blocked.

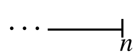
(19) a. at least 15pp $\lambda d[\Diamond$ the paper is d-long]
 b. $min\{d|\Diamond[\textsc{length}(p) = d]\} = 15\text{pp}$

As the possibility modal rescues a sentence with a superlative modifier, there is no need to insert an additional covert epistemic modal in cases with an overt possibility modal. Nouwen's account thus predicts that authoritative readings always arise if *at least* and *at most* take scope over a deontic possibility modal. In combination with *at least*, the modified numeral or measure phrase specifies the lower bound of the deontic range, and for *at most n*, *n* corresponds to the upper bound of the deontic range, as shown in (20).

(20) a. *at least n* $\succ \Diamond$:



 b. *at most n* $\succ \Diamond$:



Since the readings where the superlative modifiers take narrow scope are either contradictory or blocked by the bare numeral, the authoritative reading is the only one predicted to be available when superlative modifiers are combined with possibility modals.

Regarding the interaction with necessity modals, Nouwen's analysis predicts that neither *at least* nor *at most* expresses sensible truth-conditions, because the narrow as well as the wide scope readings are either contradictory or blocked. But we can assume that these combinations too can be rescued by inserting a covert epistemic possibility modal in the scope of the superlative modifier and above the deontic necessity modal. *At most* combined with a necessity modal will then convey speaker ignorance regarding the lower bound of the deontic range, as illustrated in (21).

(21) a. *at most* $n \succ \Diamond \succ \Box$:

////////////////|——— ···
n

Nouwen (2010) moreover proposes that a necessity modal is interpreted as a possibility modal when minimality is at stake, such that *at least* plus necessity modal comes out equivalent to *at least* plus possibility modal and thus has the authoritative reading shown in (20a).

2.3 Ignorance implications as quantity implicatures (Büring 2008, Schwarz 2011, 2013, Kennedy 2013)

Another line of research, pioneered by Büring (2008), analyzes ignorance implications of superlative modifiers as pragmatic inferences, more precisely as quantity implicatures. Büring's proposal sets out from the intuitive equivalence of *at least n* with *n or more* and builds on the observation that ignorance inferences also arise from disjunction, as illustrated in (22).

(22) Ernie or Bert called.
$\rightsquigarrow$ The speaker is not certain but considers it possible that Ernie called.
$\rightsquigarrow$ The speaker is not certain but considers it possible that Bert called.

Büring proposes that the lexical semantics of *at least n* corresponds to *n or more*. He thus assumes that the standard neo-Gricean account of ignorance inferences arising with disjunction as quantity implicatures carries over to superlative modifiers.

While the idea of deriving ignorance inferences as quantity implicatures is both conceptually attractive and empirically well-supported, Büring (2008) does not yet over a full fledged account. Semantic equivalence with an expression that is known to trigger implicatures is in fact not sufficient for the generation of implicatures (see Coppock and Brochhagen 2013b for discussion). Büring's idea has been taken up by Schwarz (2011, 2013) and Kennedy (2013), who spell out the parallel with disjunction within a neo-Gricean account of ignorance implications (Sauerland, 2004). The basic ingredients are the same in the analyses of Schwarz and Kennedy. Superlative modifiers are degree operators expressing non-strict comparison of the maximum:

(23) a. $[\![\text{at least}]\!] = \lambda n.\lambda D_{(dt)}.max\{d|D(d)\} \geq n$
b. $[\![\text{at most}]\!] = \lambda n.\lambda D_{(dt)}.max\{d|D(d)\} \leq n$

They moreover argue that utterances with superlative modifiers are obligatorily considered against alternative, more informative utterances. While Schwarz and Kennedy differ in the details of how alternatives are computed, the result is the same:[4] For a sentence with *at least n* the more informative alternatives are the ones that express strict comparison with *n* or indicate *n* as the exact maximum, as illustrated in (25) for sentence (24).[5]

(24) a. The paper is **at least** 15 pages long.
b. $max\{d|\text{LENGTH}(p) \geq d)\} \geq 15pp$

(25) Stronger scalar alternatives:
a. $max\{d|\text{LENGTH}(p) \geq d)\} > 15pp$
b. $max\{d|\text{LENGTH}(p) \geq d)\} = 15pp$

These scalar alternatives are symmetric, i.e. they cannot simultaneously be false while the assertion is true. Therefore, ignorance implications rather than scalar implicatures are generated (see Sauerland 2004). Thus (24) implicates that the speaker is not certain that the paper is more than 15 pages long and she is not certain that the paper is exactly 15 pages long. In other words, for all the speaker knows the paper might be exactly 15 pages long or longer than 15 pages. Similarly for *at most n* where the more informative alternatives are the ones involving '$< n$' or '$= n$' in the truth conditions.

These pragmatic accounts thus explain why unembedded occurrences of superlative modifiers convey speaker ignorance. They also predict that ignorance inferences can be obviated when superlative modifiers are combined with necessity modals. When *at least* and *at most* are interpreted in the scope of a necessity modal, the stronger scalar alternatives are not symmetric, and consequently scalar implicatures rather than ignorance implications are generated. In the case of (26) for instance, the stronger scalar alternatives in (27) can be simultaneously false while the assertion is true. This is the case if permissible paper lengths correspond to a range of values whose lower bound is 15 pages. As a consequence, these alternatives lead to scalar implicatures according to which the speaker is sure that paper lengths of exactly 15 pages as well as more than 15 pages are permissible.

(26) a. The paper **has to** be **at least** 15 pages long.
b. $\Box\,[max\{d|\text{LENGTH}(p) \geq d)\} \geq 15pp]$

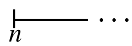
(27) Stronger scalar alternatives:
a. $\Box\,[max\{d|\text{LENGTH}(p) \geq d)\} > 15pp]$
b. $\Box\,[max\{d|\text{LENGTH}(p) \geq d\ )\} = 15pp]$

[4]Schwarz (2011, 2013) assumes the Horn set {*at least, exactly, at most*} of scalar modifiers in addition to the Horn set of numerals. Kennedy (2013) proposes that in the alternatives the numeral is kept constant while *at least n* is substituted by *more than n* (and *at most n* is substituted by *less than n*) and the bare numeral, for which he assumes a two-sided semantics.

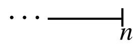
[5]See Schwarz (2013) for a discussion of the consequences of taking the full set of stronger alternatives into account.

Interpreting a superlative modifier in the scope of a necessity modal thus gives rise to the authoritative readings illustrated in (28), where *n* in *at least n* specifies the lower bound of the deontic range, and *n* in *at most n* specifies the upper bound of the deontic range.

(28) a. $\Box >$ *at least n*:

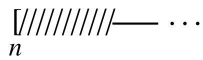


b. $\Box >$ *at most n*:

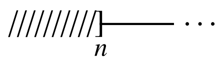


In addition, speaker insecurity readings result if *at least* and *at most* take wide scope over a necessity modal. In these cases the scalar alternatives are symmetric, just as we saw in the case of unembedded occurrences above. Hence ignorance implications are generated to the effect that the speaker is uncertain about the minimally required paper length, as shown in (29).

(29) a. *at least n* $> \Box$:

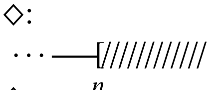


b. *at most n* $> \Box$:

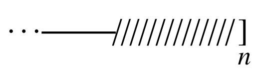


For combinations with possibility modals, the neo-Gricean approach predicts obligatory ignorance inferences for both *at least* and *at most*. In these cases the narrow as well as the wide scope readings lead to symmetric scalar alternatives and thus to ignorance implications.[6] When superlative modifiers take wide scope over the modal, the resulting truth conditions are about the upper bound of the deontic range. In the pragmatic component, ignorance implications are generated according to which the speaker is not sure about the upper bound of the deontic range, as shown in (30).[7]

(30) a. *at least n* $> \Diamond$:



b. *at most n* $> \Diamond$:



In sum, this family of neo-Gricean accounts predicts that both *at least* and *at most* are able to suppress ignorance inferences and give rise to authoritative readings when they are embedded under a necessity modal. In combination with possibility modals in contrast, neither *at least* nor *at most* are predicted to yield authoritative readings.

[6] An obvious way to extend the pragmatic account would be to build on the fact that disjunction in combination with possibility modals leads to free choice inferences, which would go beyond the neo-Gricean approach (see Fox 2007 among others). If we assume that the Büring-Schwarz-Kennedy-account can be extended along the line of free choice, we would expect that both *at least* and *at most* give rise to authoritative readings under possibility modals. The predictions of this extended version of the analysis would then be equivalent to the ones discussed for the account of Coppock and Brochhagen (2013b) albeit applying a different pragmatic analysis.

[7] Interpreting the superlative modifiers with narrow scope under the modal leads to a weak reading saying merely that there is a deontically accessible world in which the paper is 15 pages long or longer (less long in the case of *at most*), as well as ignorance inferences, which are too strong in the sense that the speaker is not even sure that *n* constitutes a permissible value.

2.4 Decomposing *at most*

Penka (2014) adopts the neo-Gricean account of ignorance inferences, but proposes a modification in the analysis of *at most*. Following the idea that negative antonyms are decomposed into an antonymizing operator and the corresponding positive antonym (Büring, 2007b,a), *at most* is analyzed as ANT + *at least*, where ANT is a scopally-mobile degree operator that denotes degree negation (Beck, 2012). When ANT takes scope immediately above *at least*, the resulting truth conditions and implicatures are the same as the ones resulting from the non-decomposed meaning of Schwarz and Kennedy for *at most* in (23b) above. Specifically, ignorance implicatures are generated for unembedded occurrences of *at most* and in cases where both ANT and *at least* take scope over a modal.

But in the interaction of *at most* and modals, the decompositional analysis makes available an additional scope order, where ANT takes wide and *at least* takes narrow scope with respect to the modal. In combination with a possibility modal, this results in the truth conditions shown in (31b), according to which the paper isn't allowed to be longer than 15 pages.

(31) a. The paper **can** be **at most** 15 pages long.
b. $\neg\Diamond\,[max\{d|\textsc{length}(p) \geq d)\} > 15\text{pp}]$ ANT $\succ \Diamond \succ$ *at least* 15

In addition, rather than formulating scalar alternatives in terms of Horn sets, Penka assumes they are restricted by structural complexity, whereby alternatives are generated by lexical substitution or deletion (Katzir, 2007). This means that in addition to the substitution of numerals and *at least* by *exactly*, ANT can be deleted in the scalar alternatives (Alxatib, 2013). Further adopting the common assumption that modals can be substituted by each other, leads to the following stronger scalar alternatives for (31):[8]

(32) Stronger scalar alternatives:
a. $\neg\Diamond\,[max\{d|\textsc{length}(p) \geq d)\} > 14\text{pp}]$
b. $\Box\,[max\{d|\textsc{length}(p) \geq d)\} = 15\text{pp}]$

Since these alternatives aren't symmetric, they don't lead to ignorance inferences. Instead, scalar implicatures are generated according to which the paper can be longer than 14 pages, but doesn't have to be exactly 15 pages long. The scalar implicatures together with the the truth conditions in (31b) say that the permissible paper lengths correspond to a range with 15 pages as upper bound. In other words, the scope configuration in (33) yields the authoritative reading.

(33) ANT $\succ \Diamond \succ$ *at least* n:
$\cdots$ ——| n

Penka (2014) further suggest that modals have certain scope preferences with respect to ANT which are related to the scope preferences these modals show vis-à-vis negation. For the combination of *at most* with the possibility modal *can*, which strongly prefers narrow scope with respect

[8] Note that substitution of modals by itself is not sufficient to prevent symmetric alternatives. The decomposition into two components that take scope independently is crucial.

to negation, this assumption allows the derivation of the authoritative reading in (33) as well as the speaker ignorance reading in (34), which both are derived from LFs where *can* is interpreted in the scope of ANT.

(34) ANT $\succ$ *at least n* $\succ$ $\Diamond$:

··· ——|//////////////]
n

Since the derivation of the authoritative reading in combination with necessity modals requires wide scope of the modal, only the speaker ignorance reading is predicted to be available for sentence 35 with the modal *have to*, which also strongly prefers narrow scope with respect to negation and arguably ANT.[9]

(35) a. The paper **has to** be **at most** 15 pages long.

b. ANT $\succ$ *at least n* $\succ$ $\Box$:

//////////]|—— ···
n

In sum, Penka's analysis predicts that *at most* gives rise to an authoritative reading in combination with possibility, but not necessity, modals. The predictions for *at least* are inherited from the analyses of Büring, Schwarz and Kennedy, i.e. *at least* gives rise to an authoritative reading in combination with necessity, but not possibility, modals.

2.5 Coppock & Brochhagen (2013)

Coppock and Brochhagen (2013b) also take a pragmatic approach to the ignorance inferences arising with superlative modifiers, but cast their analysis in a different framework, namely Inquisitive Semantics. They analyze superlative modifiers as expressions denoting sets of alternatives ("possibilities" in Inquisitive Semantics parlance) that are ranked as least as high (for *at least*) or as most as high (for *at most*) according to some pragmatic ranking. Speaker ignorance is attributed to a Maxim of Interactive Sincerity, according to which a speaker should only utter a sentence denoting a set of alternatives if her information state is consistent with those alternatives. For illustration, let us consider our usual example (36a).

(36) a. The paper is **at least** 15 pages long.

b. *The paper is 15 pages long, or the paper is 16 pages long, or the paper is 17 pages long ...*

Under Coppock and Brochhagen's analysis, (36a) denotes a set of possibilities and is equivalent to the disjunction (36b). According to the Maxim of Interactive Sincerity, this can only be uttered if the speaker considers each of the disjuncts an epistemic alternative, that is if the speaker is unsure about the exact length of the paper.

[9] The other scope order where ANT takes wide scope, ANT $\succ$ $\Box$ $\succ$ *at least n*, leads to a reading with overly strong ignorance implicatures, which is probably not detactable.

Regarding the interaction with modals, we again have to consider the wide and narrow scope configurations. The configuration where a superlative modifier takes wide scope over a modal denotes a set of alternatives, just as unembedded cases of superlative modifiers, and thus gives rise to ignorance implicatures. If *at least 15 pages* in (40a) is interpreted with wide scope over the necessity modal, the resulting set of possibilities is equivalent to the disjunction in (38a). According to the Maxim of Interactive Sincerity, this can only be uttered if the speaker considers each of the disjuncts an epistemic alternative, that is if she is unsure about the lower bound of the deontic range, as illustrated in (38b).

(37) The paper **has to** be **at least 15** pages long.

(38) *at least* $\succ \Box$:

a. *The paper has to be 15 pages long, or the paper has to be 16 pages long, or the paper has to be 17 pages long ....*

b. [//////////////—— ···
13 15 17

Similarly for the other combinations, the speaker insecurity reading results if the superlative modifier is interpreted with wide scope over the modal.

(39) a. *at least* $\succ \Diamond$:

···——[//////////////////
n

b. *at most* $n \succ \Box$:

///////////////]—— ···
n

c. *at most* $n \succ \Diamond$:

···——///////////////]
n

If superlative modifiers take narrow scope under a modal, no ignorance inferences arise due to Existential Closure, which applies in the scope of modals and whose function is to gather all the alternatives into a single proposition corresponding to the disjunction of all these alternatives. Coppock and Brochhagen (2013b) therefore predict that for each superlative modifier-modal combination, both a reading with and without speaker ignorance is possible. In the scope of a necessity modal, *at least n* and *at most n* specify the lower and upper bound of the deontic range, respectively. For *at most n* in the scope of a possibility modal, Coppock and Brochhagen (2013b) argue that the resulting reading, which says that values up to n are permissible, is strengthened by an exhaustivity implicature, according to which values higher than n are not permissible. The same reasoning should apply to *at least n* in the scope of a possibility modal: The truth conditions derived from this structure specify that n and higher numbers are permissible and are subsequently strengthened by an implicature to the effect that lower numbers are not permissible. Note that this strengthening by implicature effectively makes $\Diamond$ + *at least* equivalent to $\Box$ + *at least*, and $\Diamond$ + *at most* equivalent to $\Box$ + *at most*, as shown in (40).

(40) a. $\Box \succ$ *at least n*, $\Diamond \succ$ *at least n*:

⊢—— ···
n

b. $\Box >$ *at most n*, $\Diamond >$ *at most n*:

$\cdots$ ——⊣
n

In sum, Coppock and Brochhagen (2013b) predict that for each combination of superlative modifiers and deontic modals, both the authoritative and the speaker insecurity reading are available.

2.6 Summary of previous analyses

All the analyses discussed here make clear predictions regarding (i) which combinations of superlative modifiers and modals can suppress ignorance inferences and give rise to the authoritative reading and which only have a speaker insecurity reading, and (ii) whether the respective reading is in terms of upper or lower bound of permissible values. As the discussion in this section made clear, the different analyses vary considerably regarding their predictions. Before we present an experimental paradigm that is designed to test which of the predictions are borne out, we summarize all the predictions of the different analyses in (41).

(41) a. $\Box$ + *at least n*: n
  i. ⊢—— $\cdots$ G&N, N, B/S/K, P, C&B
  ii. [//////////////—— $\cdots$ G&N, B/S/K, P, C&B
 b. $\Diamond$ + *at least n*:
  i. ⊢—— $\cdots$ N, C&B
  ii. $\cdots$ ——⊣////////////////// G&N, B/S/K, P, C&B
 c. $\Box$ + *at most n*:
  i. $\cdots$ ——⊣ B/S/K, C&B
  ii. ///////////////⊢—— $\cdots$ G&N, N, B/S/K, P, C&B
 d. $\Diamond$ + *at most n*:
  i. $\cdots$ ——⊣ G&N, N, P, C&B
  ii. $\cdots$ ——⊣//////////////] G&N, B/S/K, P, C&B

3 Experiment 1: Drawing the boundary between semantics and pragmatics experimentally

3.1 Research question

As discussed in detail in Section 2, there is disagreement in the literature about the status of the ignorance inferences that superlative modifiers give rise to. While Geurts and Nouwen (2007) locate them in the lexical semantics, Büring (2008); Nouwen (2010); Cummins and Katsos (2010); Schwarz (2011); Kennedy (2013); Coppock and Brochhagen (2013b), and Penka (2014) argue that they arise as pragmatic inferences. Both views have been supported by experimental investigations (see Geurts et al. (2010) for the former approach and Cummins and Katsos (2010) and Coppock and Brochhagen (2013a) for the latter).

In order to develop an experimental methodology that would help with determining whether ignorance inferences arise semantically or pragmatically, we present here Experiment 1, in which we adapt, and improve on, methodology used in Cummins and Katsos (2010: §9). In Experiment 1 we address methodological problems of previous studies and show that the adapted methodology is successful at detecting what participants identify as semantic or pragmatic inconsistency.

3.2 Previous experimental studies and methodological forethought

Cummins and Katsos (2010) utilize an experimental methodology for capturing the difference between semantic contradiction and pragmatic infelicity using the notion of 'coherence'. In their study, they asked participants to evaluate the coherence of two juxtaposed statements as in (42) on a Likert scale ranging from 5 ('coherent') to −5 ('incoherent').

(42) $$\text{Jean has } \left\{\begin{matrix}\text{at least}\\ \text{at most}\end{matrix}\right\} \text{ 3 houses. } \left\{\begin{matrix}\text{Specifically}\\ \text{In fact}\end{matrix}\right\}\text{, she has exactly } \left\{\begin{matrix}2\\ 3\\ 4\end{matrix}\right\}.$$

(adapted from (Cummins and Katsos, 2010, 291, ex. 9))

What Cummins and Katsos (and Katsos and Smith (2010); Katsos and Bishop (2011) before them) found was that participants divided a Likert scale into three regions corresponding to three types of semantic-pragmatic proposition-utterance status: semantically true and pragmatically felicitous, semantically true but pragmatically infelicitous, and semantically false and (as a consequence) pragmatically infelicitous. In the case of a Likert scale ranging from −5 to 5, the positive region reaching 5 was associated with semantically true and pragmatically felicitous utterances, the region encircling 0 was associated with semantically true but pragmatically infelicitous utterances, and the negative region reaching −5 was associated with semantically false and (as a consequence) pragmatically infelicitous utterances.

While we find this methodology developed in Cummins and Katsos (2010) ingenuous and suitable for our research question, we take issue with two main methodological problems in their study. The first methodological bit we take issue with is the type of statement sequences participants were asked to assess. The first statement involved a numeral modified by *at least* or *at most*, and the second the same numeral or a numeral smaller or larger by one modified by *exactly*, as in (42). Although Cummins and Katsos interpret the results of their study as indicative of participants' distinguishing between the semantic and pragmatic components of superlative modifiers, this type of discourse, which includes an explicit revision of the first, less certain statement, seems marked and unnatural. Specifically, what seems marked and unnatural is the abrupt transition from an epistemic state in which the speaker is uncertain about the exact number to one in which she is certain about it, without any sentential connectives that signal the change in epistemic state, e.g. *actually, oh wait*. Instead, the second statement involved *specifically* or *in fact*. *Specifically* prepares the interlocutor for a more precise and detailed elaboration and *in fact* does the same but in addition buttresses the rhetorical strength or introduces a proposition to which the speaker has a stronger epistemic commitment (Traugott and Dasher, 2002). That is, both expressions communicate that speaker's certainty is given in the discourse, and not that the speaker is shifting from an uncertain to a certain epistemic state. In addition, since the information with which participants assessed the statements was given entirely and only by the statements, it raises the question whether a different scenario in which participants had more information about the correct number under discussion and the speaker's epistemic state in advance would result in a different response pattern. We therefore chose a setting where a statement with a superlative modifier was evaluated after a context had been revealed rather than vice versa, similarly to Geurts et al. (2010), who did so with text as well, and to Coppock and Brochhagen (2013a), who did so with pictures (see Section 3.3 for example stimuli.)

Another methodological issue emerges in the tasks used by Geurts et al. (2010) and Cummins and Katsos (2010). In their first experiment, Geurts et al. used sentences involving bare numerals and asked whether they entailed or were entailed by statements with *at least* or *at most*. This raises the question whether bare numerals were interpreted under their one- or two-sided meaning (e.g. *3* interpreted as *at least 3* or *exactly 3*). Cummins and Katsos took a different approach and used the expression *exactly 3* to contrast it with *at least/at most 3*. Recall that in the pragmatic analysis in Schwarz (2011, 2013), the precisifier *exactly* is assumed to be in the same Horn set as the superlative modifiers. Since we wanted to remain agnostic as to whether such a Horn set is being evoked, we avoided the potential confound that competing expressions in the discourse might introduce scalar alternatives by steering clear of the expression *exactly* altogether. In order to ensure an *exactly* reading of the numeral, we introduced instead the numeral-containing NP with a definite description or a quantified NP, e.g. *the eight balls, all (of the) eight ducks* (see Section 3.3 for example stimuli).

3.3 Methods

In Experiment 1, 45 participants (26 Female; Mean Age: 35.90; Age range: 20-56) were asked to rate the coherence of a speaker uttering a sentence with a superlative modifier in a given context. The task was conducted on Amazon's Mechanical Turk. Participants received $3.24 for answering 108 experimental items at a rate of $0.03 per item. The task took between 20 and 40 minutes to complete.

Participants were given a context that included a number description that received an *exactly* reading, facilitated by a definite description or a quantified NP (e.g. *the three ducks, all (of the) three ducks*). The utterances differed in whether the numerals were unmodified (i.e. bare), modified by either *at least* or *at most* and whether the number in the utterance matched the one in the context (N condition henceforth), was smaller in one integer (N-1 condition henceforth) or was greater in one integer (N+1 condition henceforth), as illustrated in (43) below. Following the presentation of the context and utterance, participants were asked to evaluated the speaker's coherence on a scale from -5 to +5, where -5 is *definitely* **not** *coherent* and +5 is *definitely coherent*.[10]

[10]In contrast with Cummins and Katsos, who asked participants to rate the coherence of the utterance, we directed participants' attention to the *speaker's* coherence instead.

(43) After the soccer practice, Cassandra sent Eduard to collect all of the soccer balls lying around in the court. He collected the eight balls that were scattered around the court. When he was done, he went back to Cassandra and said:

$$\text{“I collected} \begin{Bmatrix} \emptyset \\ \text{at least} \\ \text{at most} \end{Bmatrix} \begin{Bmatrix} \text{seven} \\ \text{eight} \\ \text{nine} \end{Bmatrix} \text{balls.”}$$

In light of the context given above, how coherent do you think the speaker is on a scale ranging from -5 to +5, where -5 is *definitely* **not** *coherent* and +5 is *definitely coherent*?

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
definitely **not** coherent										definitely coherent

The experimental design was therefore 3 (modifier) × 3 (number discrepancy) conditions. There were six observations for each one of the conditions, totalling in 54 target items. Nine lists were created using a latin square design so that each participant saw each context with only one modifier-number discrepancy combination.

In addition to the 54 target items, there were also 18 control items that consisted of 9 contradictions and 9 entailments, as well as 36 scalar items. Many of the control and scalar items were adapted from Doran et al. (2012).

The control items included trials in which the utterance was entailed by the context and items in which the utterance contradicted the context. The purpose of the control items was to introduce utterance types that would be mapped onto the regions approaching the extreme poles of the coherence range. We expect entailments as in (44) to be given coherence rates in the region approaching the +5 end of the scale, and contradictions as in (45) to be given coherence rates in the region approaching the −5 end of the scale.

(44) *Example control item: Entailement*
Luke's wife is a very important and sought-after pediatric doctor at Vincentius Hospital. Dominique asked Irene to describe Luke's family to her. She talked about Luke's parents and then said:
"Luke's wife is a doctor at a hospital."

(45) *Example control item: Contradiction*
Claire had her chemistry final yesterday. The professor posted the answers to the exam questions and Claire found out she didn't make any mistakes on her chemistry final. When Jon asked Claire why she was upset she said in response:
"I made a big mistake on my chemistry final."

The scalar items all involved a context that was pitted against scalar expressions, e.g. *some/many/all, good-looking/pretty/gorgeous*. In light of the context, we expected each expression on a posited scale to be interpreted as logically-weaker, -similar, or -stronger than the context, as illustrated by *some of the, most of the*, and *the whole*, respectively in (46). We expect the logically-weaker expressions, as *some of the* in (46), to be judged as less coherent in comparison with the entailment and the logically-consistent items, as they are underinformative and therefore pragmatically degraded. We expect the logically-stronger scalars to be judged as the contradiction items, as they oftentimes would introduce additional information that doesn't correspond to the information given in the preceding context. For example, given that Esther invited more than half of the class to the party in (46), uttering a statement with *the whole class* would be degraded, as it is logically-stronger than the information participants obtained from the preceding context.

(46) *Example scalar item:*
It's the end of the school year, and Esther wanted to celebrate this important landmark with the rest of the class, so she organized a party and invited more than half of the class to it.
Esther said to her friend:
"I've invited $\left\{\begin{array}{c}\text{some of the}\\ \text{most of the}\\ \text{the whole}\end{array}\right\}$ class to my party."

Since the scalar items involve utterances that could potentially be regarded as infelicitous—being underinformative, for example—they will be used as an additional yardstick for the ternary categorization of the semantic-pragmatic status of utterances.

3.4 Results

Participants judged the contradiction items as incoherent (median: -5; mean: -3.75; SD: 0.83) and entailment items as coherent (median: 5; mean: 3.63; SD: 0.73), as expected.

Recall that the scalar items were divided into three utterance-context discrepancy conditions: in the 'weaker' condition, the expression used in the utterance was informationally weaker (read: under-informative) than the one stated in the context; in the 'similar' condition, the expression in the utterance was roughly synonymous with the expression in the context; and in the 'stronger' condition, the expression in the utterance was informationally (and logically) stronger than the one in the context—that is, the former asymmetrically entailed the latter. We ran a Cumulative Link Mixed Model fitted with the Laplace approximation (Christensen, 2012) with condition as a fixed effect and subject and experimental item as random effects, and found that each one of the scalar types was different from the other two. When we compared each of the scalar types to entailments, we found that the similar condition was only marginally significantly different than entailments (p=0.094), unsurprisingly. (See results of statistical tests in Table A.1 in the Appendix A.)

As Figure 1 illustrates, utterances with under-informative, weaker (and thus pragmatically infelicitous) expressions were rated as reliably less coherent (median: 3; mean: 1.65; SD: 2.99) than the similar expressions that fit the description in the context (median: 4; mean: 3.71; SD: 1.89), and stronger expressions (median: 1; mean: 0.5; SD: 3.46) were rated even less coherent than the weaker utterances but surprisingly more coherent than contradictions.

We turn now to the bare numeral conditions. As shown in Figure 2 and further supported by the results of a Cumulative Link Mixed Model (see Table A.2 in Appendix A), utterances with bare numerals were judged as completely coherent in the N condition, i.e. when the number

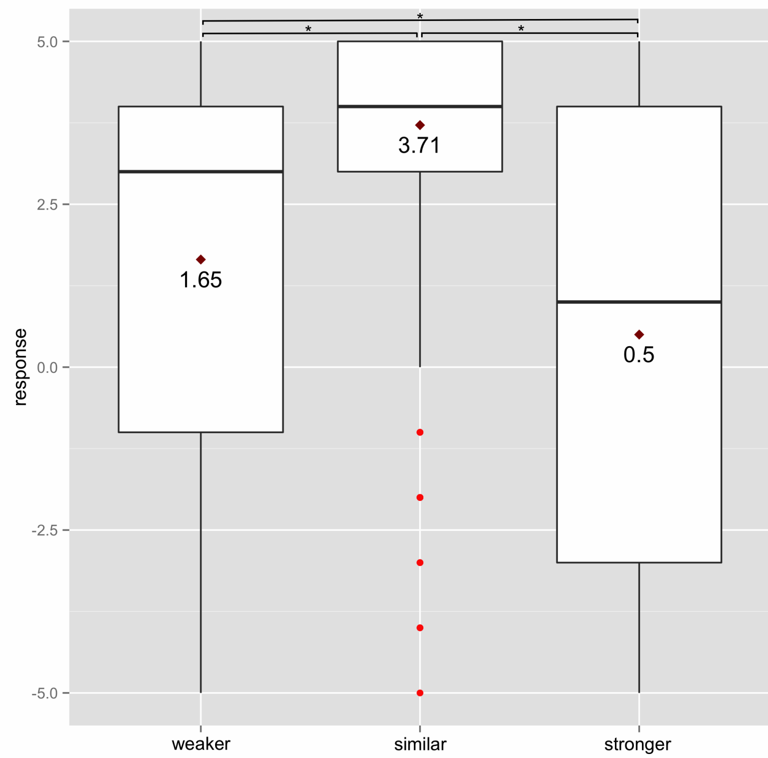


Figure 1: Experiment 1: Coherence Rates in the Scalar Items

in the utterance matched the one in the context (median: 5; mean: 4.75, SD: 0.14), as expected. Moreover, the coherence rates in this condition were reliably higher than the ones for the entailment items.

Utterances in the N−1 condition, i.e. the condition in which the number in the utterance was smaller than the number in the context (median: -3; mean: -1.55, SD: 0.33) were reliably less coherent than the ones in the N condition but reliably more coherent than the ones in the N+1 condition, i.e., the condition in which the number in the utterance was greater than the number in the context (median: -3; mean: -2.82, SD: 0.29). In addition to being significantly different than each other, the three bare numeral conditions were also reliably different than entailments and contradiction.

As shown in Figure 3.4, utterances with numerals modified by *at least* were judged as reliably more coherent in the N condition, i.e., when the number in the utterance matched the one in the context (median: 4; mean: 2.68, SD: 2.96) than in the N−1 condition, i.e., when the number in the utterance was smaller than the number in the context (median: 3; mean: 2.13, SD: 2.82). In the N+1 condition, i.e., when the numeral was greater in the utterance than in the context, participants judged the utterance as the least coherent from among the three *at least* conditions (median: -3; mean: -2.35, SD: 2.82). All three *at least* conditions were significantly different than entailments

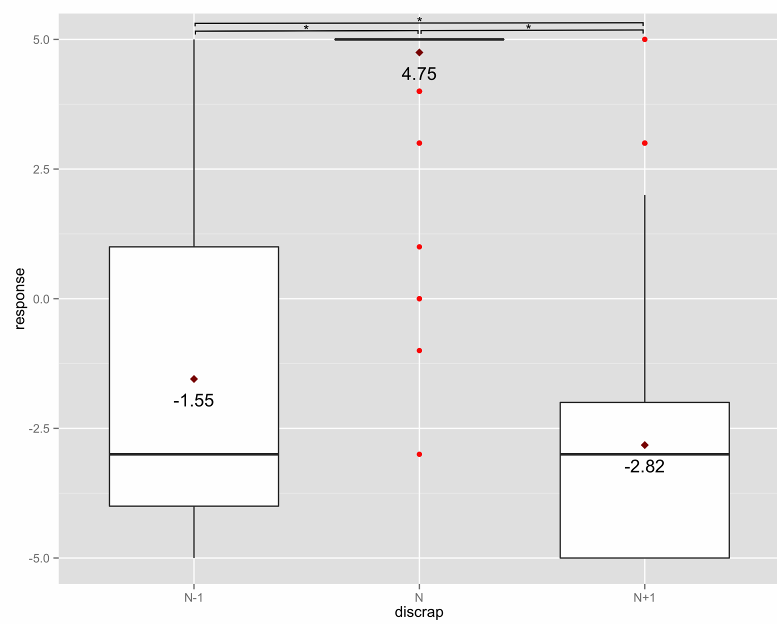


Figure 2: Experiment 1: Coherence Rates in the Bare Numeral Condition

and contradictions (see Table A.3 in Appendix A).

Statistical tests have revealed some similarities (or rather, no significant differences) between some *at least* conditions and other stimuli types. Coherence rates for the *at least* N−1 condition, for example, weren't different than those for the weaker scalar condition. The coherence rates for Esther's saying she's invited some of the class while she's invited more than half of it were similar to Eduard's saying he's collected at least seven balls when he's in fact collected eight balls. In addition, coherence rates for the *at least* N+1 condition weren't different than those for the bare numeral N+1 conditions. That is, the coherence rates were similar for Eduard's saying that he's collected nine balls or that he's collected at least nine balls when he's in fact collected eight balls.

As shown in Figure 4, utterances with numerals modified by *at most* were judged as reliably more coherent in condition N, i.e. when the number in the utterance matched the one in the context (median: 4; mean: 2.47, SD: 3.17), than in condition N+1, i.e. when the number in the utterance was greater than the one in the context (median: -2; mean: -1.30, SD: 2.99). Utterances in condition N−1, when the number in the utterance was smaller than the number in the context, were judged as even less coherent than in the other two *at most* conditions (median: -3; mean: -1.83, 3.33). All three conditions were reliably different from each other as well as from entailments and contradictions (see Table A.4 in Appendix A).

Some similarities in coherence rates between the *at most* conditions and other stimulus types should be noted. The coherence rates for the N-1 *at most* condition didn't differ from those in the N+1 *at least* condition. That is the coherence rates for Eduard's saying he collected at most seven balls were similar (i.e. equally low) to those when he said he had collected at least nine balls when

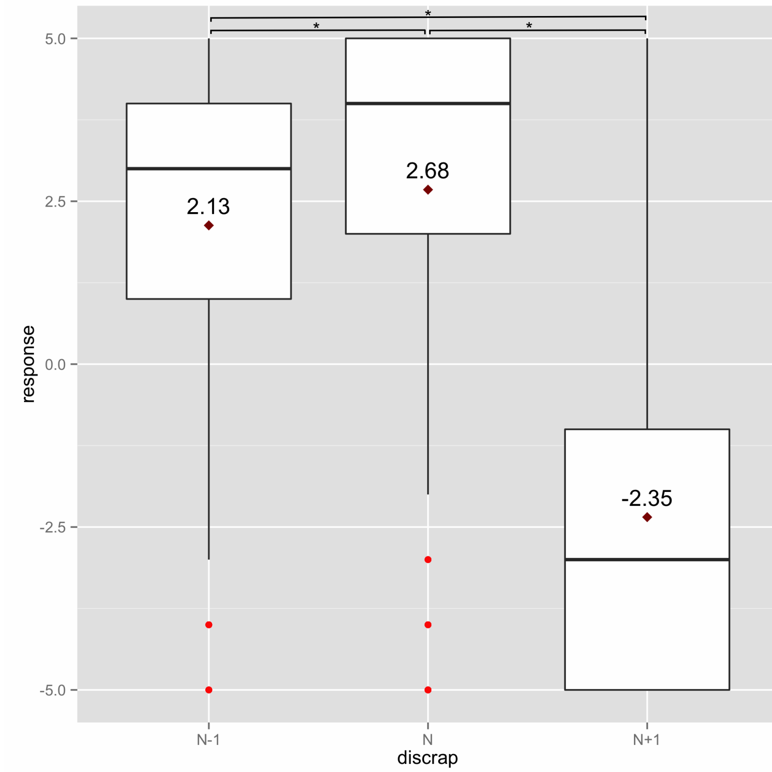


Figure 3: Experiment 1: Coherence Rates in the *At Least* Condition

he collected eight. The coherence rates for the N+1 *at most* condition didn't differ from those in the N-1 bare numeral condition. That is, the coherence rates for Eduard's saying he's collected seven balls weren't different than those for Eduard's saying he collected at most nine balls when he collected eight. And finally, the coherence rates for the N *at most* condition didn't differ from those in the N *at least* condition. That is, the coherence rates for Eduard's saying he's collected at most eight balls weren't different from the coherence rates for his saying he's collected at least eight balls when he collected eight balls.

3.5 Discussion

The results of Experiment 1 show that there is a difference in the response patterns in the task between three types of utterances: (1) utterances that are logically and pragmatically true; (2) utterances that are logically false; and (3) utterances that are logically true but pragmatically degraded. The first category included entailments and the 'similar' condition in the scalar inference target type, which weren't statistically significant from each other. A third stimulus type one would expect to pattern with entailments and 'same' scalar inferences would be the N condition of bare numerals. This condition, however, received reliably higher coherence rates than the previously-

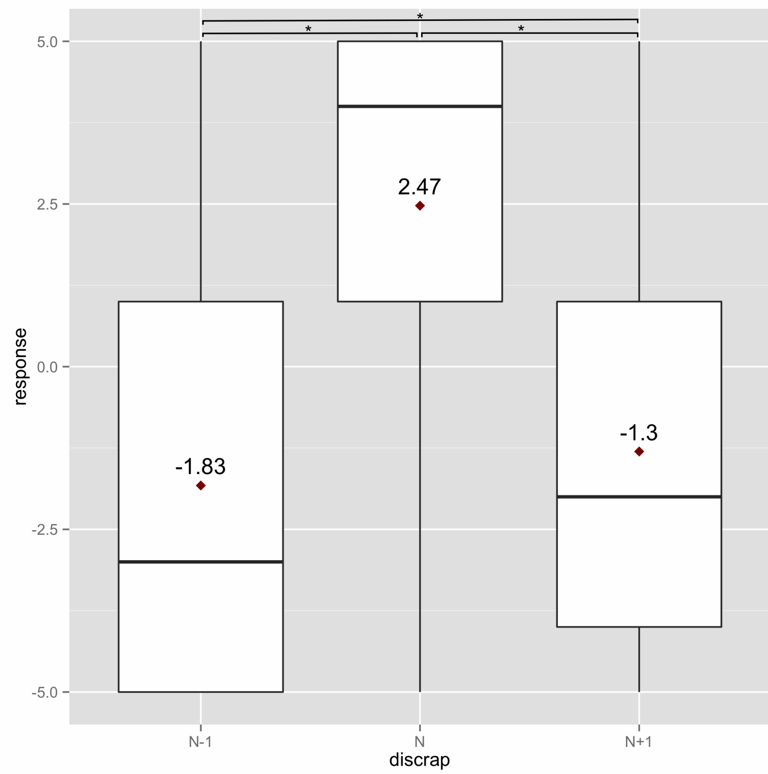


Figure 4: Experiment 1: Coherence Rates in the *At Most* Condition

mentioned stimulus types. We attribute this difference to the fact that in the N condition of the bare numeral stimulus type, the same number was given in the context and the utterance, leaving no doubt as to whether the utterance described the state of affairs as given in the context. In the entailment items and the 'similar' scalar inferences items, participants may have had reasons to believe that the utterance didn't merit the maximal coherence score, leading the rates to be reliably lower than in the other two stimulus types. The second category of utterances is represented by the contradiction items, which were significantly different than all other stimulus types. The third utterance type includes the 'weaker' and 'stronger'[11] scalar inference conditions and all the target conditions except for the N condition in bare numerals. This group of utterance types have proven to be diverse with respect to the coherence rates they exhibited.

Recall that in this experiment we've adopted and adapted the task used in Cummins and Kat-

[11] Recall that items in the 'stronger' condition were predicted to pattern with contradictions. They were indeed less coherent than the items in the 'weaker' condition but were different than the contradiction items. Discussing the source of this difference with respect to our experimental task is beyond the scope of this paper, but it may be that since the utterance in the 'stronger' condition entailed the information in the context, this was enough not to rate it as completely incoherent but still led to lower coherence rates, whereas the utterance in the contradiction items was completely false given the context. In any case, what is important is that the methodology used in this experiment clearly helped differentiate between three types of semantic-pragmatic proposition-utterance status.

sos's Experiment 2. Although our experiment is quite different than Cummins and Katsos's—namely with respect to the discourse context against which participants had to evaluate the target utterance and in general the experimental items used—we would like to highlight a few differences between our findings and theirs. Table 3.5 includes the means and standard deviations for coherence rates in our experiment and in Cummins and Katsos's and Coppock and Brochhagen's 2013a experiments.[12]

Table 3.5 shows many similarities between our and their results, but a few differences are noteworthy. Cummins and Katsos did not find any difference between *at least* and *at most*: When the number introduced in the correction/clarification sentences was previously denied by using *at least* N+1 or *at most* N−1, the sequence was judged as not coherent (mean of less than −4). But when the sequence wasn't logically contradictory but rather communicated uncertainty about the exact number introduced by the correction/clarification sentence, the sequences were given ratings that were in the middle of the −5 to 5 coherence scale (means around 1-2).

Superlative Modifier	Condition	Exp.	1	Cummins and Katsos		Coppock and Brochhagen
		Mean	SD	Mean	SD	Percent true
At least	N−1	2.13	2.82	1.95/2.55	2.53/2.16	98.3%
	N	2.68	2.96	1.28/1.33	2.50/2.56	100%
	N+1	-2.35	2.82	-4.48/-4.27	1.50/1.88	1.0%
At most	N−1	-1.83	3.33	-4.08/-4.05	2.34/2.10	3.3%
	N	2.47	3.17	1.90/1.25	2.31/2.60	97.8%
	N+1	-1.30	2.99	1.58/1.87	2.57/2.53	76.1%

Table 1: The results in Experiment 1 vs. in Cummins and Katsos (2010) and Coppock and Brochhagen (2013a)

The coherence rates in our experiment, on the the other hand, suggest that participants were more charitable by judging *at least* N+1 and *at most* N−1 as more coherent than participants in Cummins and Katsos's study and, more importantly, significantly more coherent than contradiction. An additional difference between our and Cummins and Katsos's results is that utterances involving *at most* in the N+1 condition (i.e., with a higher number than in the context (e.g., saying *at most six ducks* having seen exactly five ducks), were penalized more than utterances that should have the same degree of semantic and pragmatic compatibility (e.g, the N conditions of *at least*

[12]We made the following adjustments in the presentation of Cummins and Katsos's results: (i) The two means and standard deviations divided by forward slashes represent the rates for a correction sentence prefaced with 'specifically' and 'in fact,' respectively; and (ii) We translated their condition labels to match ours, where N represents the number in the context (in our experiment) or in the follow-up, correction/clarification sentence (in their experiment). Coppock and Brochhagen (2013a) used a picture verification task that didn't test for a contrast between semantic and pragmatic inconsistence. The rate of Yes responses should then be interpreted as truth-conditional compatibility (in spite of pragmatic infelicity), though see discussion of the response pattern for the *at most* N+1 condition. The results from their studies are taken from their experiments 1 and 2.

and *at most*). The asymmetry found in this experiment between *at least* and *at most* as well as the general response pattern is similar to Coppock and Brochhagen's 2013a study, who used a picture verification task.

The finding that participants rate utterances involving *at most* in the N+1 condition incoherent seems to speak in favour of Geurts and Nouwen's analysis, according to which the inference associated with *at most n*, namely that the speaker considers *n* possible, is part of the truth-conditional meaning. Our results, however, don't support Geurts and Nouwen's analysis of *at least n*: Although under their analysis *at least n* entails that the speaker considers higher numbers than *n* possible, we found that incompatibility of the context with higher numbers (as in the N condition) doesn't lead to incoherence. The results could then be interpreted as indicating an asymmetry between *at least* and *at most*: While the inferences about the epistemic state of the speaker seem to be more pragmatic in the case of *at least*, they seem to be more semantic with *at most*.

We believe, however, that a different conclusion is warranted. A closer look at our results reveals that only a third of the participants ruled out utterances involving *at most* with a higher number than the one mentioned in the context, whereas the remainder judged them as merely infelicitous, similarly to Cummins and Katsos (2010). The fact that a third of the participants found this case bad could be attributed to participants' inability to accommodate an infelicitous utterance that introduces a greater (as opposed to a smaller) number in the context. That is, participants found it difficult to formulate an explanation for why the speaker would say *at most six ducks* having seen exactly five ducks. Under this view, the different results for *at most* and *at least* reflect a difference in the ease with which an infelicitous utterance can be accommodated. Accommodation is easier in the case of *at least* where the speaker might be taken to make a cautious statement than it is for *at most* where the speaker should have sufficient evidence that *n* is possible and that *n*+1 isn't possible in order to say *at most n*. Coppock and Brochhagen make the same observation about the greater degree of infelicity when one modifies a discourse-new numeral with *at most* than when the numeral is already given and provide an account couched in Inquisitive Semantics. In this framework, stating a possibility explicitly is called HIGHLIGHTING. When uttering *at most* $n + 1$, the speaker highlights a false possibility, as the number in the context is *n*.

In sum, the results from our study are more compatible with the pragmatic account but point to a more nuanced analysis of superlative modifiers, in which a distinction is made between the number made prominent by mentioning it explicitly in the utterance and alternative numbers, such that some pragmatic inferences are more easily accommodated given certain conversational contexts.

Now that we've established that the methodology used in Experiment 1 is suitable in determining which components of meaning are are derived from the truth conditions and which from additional inferences, we proceed to discuss Experiment 2, in which we used this same task to study the interaction between the speaker's epistemic state as understood from the context and the speaker epistemic state as inferred from the use of superlative modifiers when embedded under deontic modals.

4 Experiment 2: Ignorance inferences of superlative modifier-modal combinations

4.1 Research question

In Experiment 2, we tested which combinations of superlative modifiers and modals can suppress ignorance inferences. As discussed in Section 2, various analyses of superlative modifiers derive the ignorance inference they give rise to either from their lexical semantics or as pragmatic inferences. Both views have been supported by experimental investigations (see Geurts et al. (2010) for the semantic view and Cummins and Katsos (2010) and Coppock and Brochhagen (2013a) for the pragmatic view). However, these studies as well as our Experiment 1 did not directly test for ignorance inferences, as participants had to judge the validity of a statement given a context that unequivocally settled the number under discussion, and lower acceptance rates were attributed to the ignorance inferences conveyed by the superlative modifiers, which were in conflict with the precise contextual information.

In our second experiment, we instead aimed to investigate speaker ignorance directly by examining the degree to which superlative modifiers embedded under modals are compatible with an explicitly-stated speaker epistemic state. In order to test this question, we pitted the speaker's epistemic state, i.e. whether a speaker is knowledgeable or ignorant with respect to the number under discussion, against the ignorance inferences conveyed by a superlative modifier-modal combination. If it is clear from the context that the speaker has the relevant information, then only utterances that can suppress ignorance inferences should be acceptable. If it is clear from the context that the speaker does not have the information, then utterances that do not convey speaker ignorance should be infelicitous. And finally, if it is not clear from the context whether the speaker has the knowledge or not, then either an utterance that leads to ignorance inference or not should be compatible with the speaker's epistemic state.

4.2 Methods

In Experiment 2, 40 participants (18 Female; Mean Age: 43.8; Age range: 26-62) were asked to rate the coherence of a speaker uttering a sentence with a modal and a superlative modifier. The task was conducted on Amazon's Mechanical Turk. Participants received \$3.24 for answering 80 experimental items at a rate of \$0.03 per item.

In this experiment, participants read scenarios like (47), in which Speaker A asks for information and Speaker B provides this information in the form of an utterance with one of the two types of deontic modals and one of the two types of superlative modifiers. Similarly to Experiment 1, following the presentation of the context and utterance, participants were asked to evaluate the speaker's coherence on a Likert scale of −5 to +5, where −5 is *definitely* **not** *coherent* and +5 is *definitely coherent*.

(47) Boris is applying for a graphic designer position at an ad company. He called the secretary of the company asking for the amount of art works in the portfolio. The secretary, who

was involved in the selection process, said:

"You are $\begin{Bmatrix} \textit{allowed} \\ \textit{required} \end{Bmatrix}$ to have $\begin{Bmatrix} \textit{at least} \\ \textit{at most} \end{Bmatrix}$ 3 works in the portfolio you send us."

In light of the context given above, how coherent do you think the speaker is on a scale ranging from -5 to +5, where -5 is *definitely* **not** *coherent* and +5 is *definitely coherent*?

	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
definitely **not** coherent											definitely coherent

In order to determine which superlative modifier-modal combinations can suppress ignorance inferences, we aimed to contrast the potential ignorance or authority reading that the utterances may generate with the knowledge of the speaker made explicit in the context. We expected that mismatches in the explicitly-stated speaker epistemic state and the one inferred by the utterance would have an effect on the coherence rates attributed to the speaker.[13] We created three different speaker epistemic state conditions (with illustrations from the example stimulus in (47)):

(48) a. ±knowledgeable: the speaker may or may not have the knowledge
"The secretary said:"

b. +knowledgeable: context makes it explicit that the speaker has the knowledge
"The secretary, who was involved in the selection process, said:"

c. −knowledgeable: context makes it explicit that the speaker does not have the knowledge
"The secretary apologized for not knowing the requirements for the application, and said:"

The experimental design was 2 (modals) × 2 (superlative modifiers) × 3 (speaker conditions). In addition to the 40 target items, there were also 20 control items that consisted of 10 contradictions and 10 entailments, which served as false and true (and pragmatically felicitous), respectively, baseline for comparison with the target items, as well as 20 scalar items as fillers.

4.3 Results

Participants rated entailments as coherent (median: 5; mean: 4.41; SD: 0.89) and contradictions as incoherent (median: -5; mean: -4.70; SD: 0.63), thereby establishing that the negative and positive ends are associated with false and true (and pragmatically felicitous) sentences, respectively.

[13]The reader may observe a potential confound in the design, noting that the speaker's epistemic state can be stated explicitly in the context or (automatically, as some may argue) inferred from the authority of the speaker. In the example stimulus (47), for instance, the secretary's knowledge regarding the job search information may be inferred by virtue of her being an employee in the company who holds a position that entails being familiar with the intricacies of, and the processes taking place in, the company. We show in Sections 4.3 and 4.4 that differences in the epistemic state of the speaker is apparent despite this potential confound. We thank Irene Heim and Ede Zimmerman for discussing this issue with us.

Before focussing on specific conditions, it's clear from the boxplots in Figure 5 that the overall rates were the highest for the +knowledgeable condition (mean: 3.02), a little lower for the ±knowledgeable condition (mean: 2.83) and the lowest for the −knowledgeable condition (mean: 0.93). An ANOVA shows a main effect of speaker condition, thereby supporting the difference observed in the coherence rates between the three speaker conditions ($F(2,1587) = 98.69$, $p<0.01$).

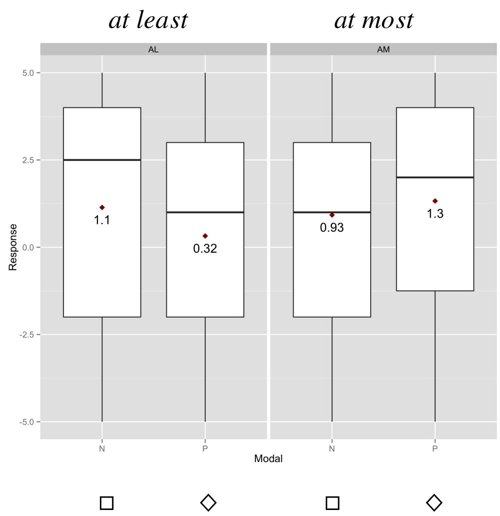
A cumulative link mixed model fitted with the laplace approximation revealed a few differences in the −knowledgeable condition: The coherence rates for the □ + *at least* condition (median: 2.5; mean: 1.14; SD: 3.33) were significantly higher than for ◇ + *at least* (median: 1: mean: 0.32; SD: 3.14). In addition, ◇ + *at most* (median: 2; mean: 1.32; SD: 2.86) was judged as significantly more coherent than ◇ + *at least* (median: 1; mean: 0.32; SD: 3.14), and □ + *at most* (median: 1; mean: 0.93; SD: 2.99) was judged as significantly more coherent than ◇ + *at least*. Lastly, coherence rates for ◇ + *at least* were significantly different than coherence rates for all of the other conditions. No other differences have been found. (See detailed results in Appendix B, Table B.1.)

In the ±knowledgeable speaker condition, we found a main effect of superlative modifier (ANOVA: $F(1,410) = 4.75$, $p<0.05$) and that superlative modifier was a significant predictor ($\beta = 0.25$, $p<0.01$). This manifested itself by *at most* receiving higher coherence rates (mean: 3.10; median: 4; SD: 2.45) in aggregate (i.e. regardless of which modal it's embedded under) than *at least* (mean: 2.56; median: 3; SD: 2.63). In order to compare the various superlative modifier-modal combinations, we submitted the data to a cumulative link mixed model fitted with the laplace approximation. This showed that □ + *at least* (median: 3; mean: 2.89; SD: 2.43) and ◇ + *at most* (median: 4; mean: 3.17; SD: 2.47) were significantly different than all other superlative modifier-modal combinations. In addition, □ + *at most* (median: 4; mean: 3.04; SD: 2.44) was significantly more coherent than ◇ + *at least* (median: 3, mean: 2.21; SD: 2.79). ◇ + *at least* was in turn significantly different than □ + *at least* and □ + *at most*. (See detailed results in Appendix B, Table B.2.)

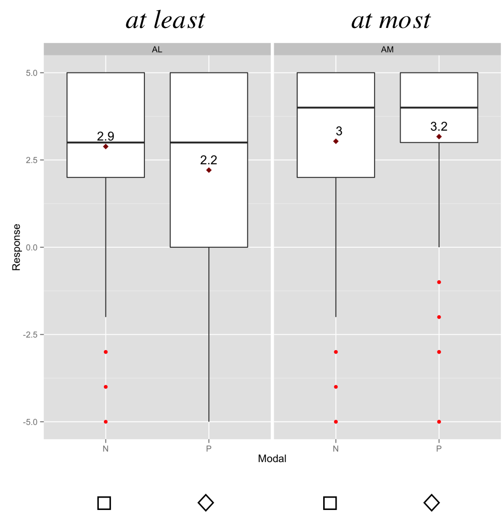
And finally, in the +knowledgeable speaker condition, the cumulative link mixed model revealed grouping between certain superlative modifier-modal combinations: The coherence rates for □ + *at least* (Median: 5; Mean: 3.59; SD: 2.50) were only marginally different than those for ◇ + *at most* (Median: 4; Mean: 3.44; SD: 2.16) and reliably different than the other two superlative modifiers-modal combinations. The coherence rates for ◇ + *at least* (Median = 3, Mean = 2.42, SD: 2.85) were not different than the ones for □ + *at most* (Median: 4; Mean: 2.60; SD: 2.99) but were reliably different than for the other two combination. □ + *at most* was only different than □ + *at least* but not than the other two combinations. And ◇ + *at most* was reliably more coherent than ◇ + *at least*.

4.4 Discussion

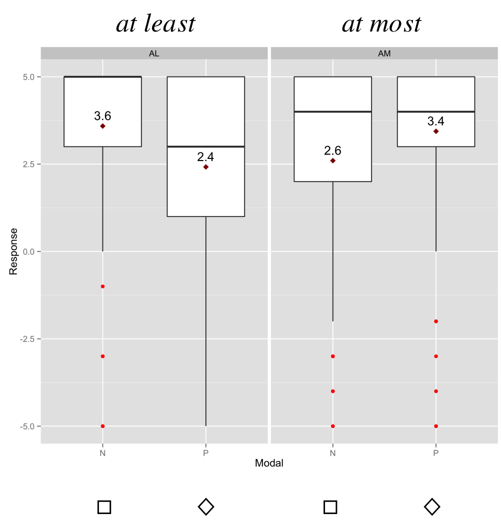
In the discussion, we start with the results from the +knowledgeable speaker condition, which show a robust contrast in coherence rates between two groups of superlative modifier-modal combinations. In this speaker condition, the combinations □ + *at least* and ◇ + *at most* are judged as more coherent than the other two combinations, ◇ + *at least* and □ + *at most*. This finding corresponds to the grouping delineated by the predictions of two of the analyses discussed in Section



(a) −Knowledgeable Condition



(b) ±Knowledgeable Condition



(c) +Knowledgeable Condition

Figure 5: Experiment 2: Boxplots of Coherence Rates in the Three Speaker Conditions

2. Under the analyses of Geurts and Nouwen (2007) and Penka (2014), □ + *at least* and ◇ + *at most* are able to suppress ignorance inferences and allow for an authoritative reading, whereas ◇ + *at least* and □ + *at most* have the speaker insecurity reading only. Under their accounts, the irrepressible, so to speak, ignorance inferences in the latter two combinations conflict with contexts in which the speaker has the knowledge and therefore should be certain about the information conveyed in the utterance. Under Nouwen's account, ◇ + *at least* would also be expected to suppress ignorance inferences, while under the neo-Gricean approaches □ + *at most* rather than ◇ + *at most* should allow for an authoritative reading. And finally, under the analysis of Coppock and Brochhagen (2013b), all four combinations should be able to suppress ignorance inferences and

thus be acceptable in the +knowledgeable speaker condition.

Moreover, since the coherence rates for the combinations that didn't exhibit ignorance inferences suppression in the +knowledgeable condition were low but distinct from those given to contradiction items—that is, since the mismatch between the speaker's certainty and the ignorance inferences generated by these combinations was not judged as a contradiction—we conclude that ignorance inferences are pragmatic rather than semantic (contra Geurts and Nouwen 2007 and in line with, e.g., Büring 2008; Cummins and Katsos 2010; Coppock and Brochhagen 2013b).

Before we discuss the results from the remaining two speaker epistemic state conditions, we would like to point out a potential problem of our experimental design. In most of our experimental stimuli the speaker in the contexts shown to participants could be claimed to be an authority on the topic to which the utterance pertains. For example, in the scenario (47), in which the secretary provides the number of works the applicant must send, her knowledge regarding the job search may be inferred by virtue of her being an employee in the company who holds a position that entails being familiar with the intricacies of, and the processes taking place in, the company. This is related to the concept of EPISTEMIC AUTHORITY in psychology and sociology, whereby individuals attribute high confidence to information provided by a source they identify as epistemic authority, consequently often assimilating it to the common ground as uncontested truth (Kruglanski, 1989). What is relevant to this study is that even in contexts in which the knowledge of the speaker was underdetermined, it could be that participants inferred that the speaker did in fact have sufficient information and thus participants would favour the authoritative reading. Note, however, that in this experiment, the coherence rates in the ±knowledgeable condition were not the same as in the +knowledgeable, in which the speaker clearly had the relevant information. This may be so because, lacking sufficient information in the ±knowledgeable condition to determine without doubt that the speaker was certain about the value in question, participants differed in how they evaluated the speaker's epistemic state, perhaps even changing this evaluation from stimulus to stimulus. In the −knowledgeable condition, contextually-inferred ignorance was at odds with participants' preference for an informed speaker. Thus utterances produced by a contextually-inferred unknowedgeable speaker were considered degraded even if this contextually-inferred ignorance matched the ignorance inferences to which the utterance itself gave rise. But even with the potential confound of inferred epistemic authority, the results from Experiment 2 shed light on the ignorance inferences of superlative modifiers if we focus on the +knowledgeable condition.

Let us summarize what the results of Experiment 2 can say about the predictions of different analyses of superlative modifiers. First, the accounts of Nouwen (2010) and Coppock and Brochhagen (2013b) as well as approaches in line with Büring (2008)/Schwarz (2011)/Kennedy (2013) seem untenable in light of the results from this experiment. Each analysis makes some predictions regarding the suppression of ignorance inferences that are not borne out. Regarding the predictions about which combinations of superlative modifier and deontic modal can suppress ignorance inferences and for which combinations ignorance inferences are obligatory, the analyses of Geurts and Nouwen (2007) and Penka (2014) fare best. Only these accounts predict that $\Box$ + *at least* and $\Diamond$ + *at most* can suppress ignorance inferences, while $\Box$ + *at most* and $\Diamond$ + *at least* cannot. But Geurts and Nouwen (2007) also assume that ignorance inferences are semantic, whereas our results suggests that they are pragmatic, which is in line with the neo-Gricean approaches (Büring, 2008; Schwarz, 2011; Kennedy, 2013; Penka, 2014) and Coppock and Brochhagen (2013b).

5 Experiment 3: Readings of superlative modifier-modal combinations

5.1 Research question

While Experiment 2 aimed at detecting ignorance inferences and their suppression, Experiment 3 tested which interpretations were available for certain combinations of deontic modals and superlative modifiers. As discussed in Section 2, the different analyses proposed in the literature make different predictions regarding the truth conditions and available inferences regarding the permitted values. The response pattern we detect in this experiment will help adjudicate between the different analysis in order to determine which of their predictions are borne out.

5.2 Methods

In Experiment 3, 40 participants (17 Female; Mean Age: 34.2; Age range: 23-59) read scenarios similar to the ones read in Experiment 2, but in this experiment the contexts were underspecified regarding the knowledge of the speaker, as in (49). The utterance was then followed by a description of an action or a state of affairs, in which the stated number was either lower (the UNDER CONDITION) or higher (the OVER CONDITION) than the one used with the superlative modifier in the utterance. Then, participants were asked to judge whether the description was in accordance with the utterance and had to reply with 'Yes' or 'No'.

(49) Professor Samsa is teaching an Introduction to Semiotics class. Jeremy, a student in his class, asked him about the length of the paper for the class, and Professor Samsa said:

"Your term paper $\left\{\begin{matrix} \textit{has to} \\ \textit{can} \end{matrix}\right\}$ be $\left\{\begin{matrix} \textit{at least} \\ \textit{at most} \end{matrix}\right\}$ 15 pages long."

Jeremy handed in a $\left\{\begin{matrix} 13 \\ 17 \end{matrix}\right\}$ page-long paper.

Did the length of Jeremy's term paper comply with Professor Samsa's specifications?

The experimental design was 2 (modal conditions) × 2 (superlative modifier conditions) × 2 (under/over conditions). There were five observations for each one of the conditions, totalling in 40 target items. Eight lists were created using a latin square design so that each participant saw each context with only one modal-superlative modifier-over/under combination. In addition to the 40 target times, there were also 20 control items and 20 scalar items. The control items consisted of 10 violation conditions, in which the description sentence violated the utterance, and 10 compliance conditions, in which the description sentence was in accordance with the utterance. The task was conducted on Amazon's Mechanical Turk. Participants received \$3.24 for answering 80 experimental items at a rate of \$0.03 per item.

Before we present the results, we would like to illustrate how the experimental paradigm we utilized is helpful in determining which of the analyses makes the correct predictions about which readings are available for the various superlative modifier-deontic modal combinations. For illustration, we focus here on the predictions the different analyses make for *at least*, but in Section 5.4 we also include *at most* in the discussion.

Consider the example stimulus in (49) and the readings the various analyses predict for the sentence in (50). As discussed in Section 2, the analyses vary whether they predict the authoritative reading (50a) to be available for the combination necessity modal plus *at least 15 pages*, or the speaker insecurity reading (50b), or both. Under the authoritative reading only higher values than 15 pages would be allowable. Therefore, we expect to get No responses in the Under condition and Yes responses in the Over condition. Under the speaker insecurity reading, the speaker is unsure about the minimally-required length and thinks that the lower bound of the deontic range might be 15 pages or more. Therefore, again, we expect to get No responses in the Under condition. The expectations for the Over condition under this reading are less clear. Since the speaker considers it possible, but is not certain, that 15 pages or 17 pages are permissible, neither replying "Yes" nor "No" in the Over condition would seem correct. Nevertheless, participants were forced to decide whether the number in the description sentence was in accordance with the utterance or not and didn't have the option to hedge their response.[14] Since we don't know how participants resolved this conflict, we take both Yes and No responses in the Over condition to be compatible with the speaker insecurity reading.

(50) $\Box$ + *at least*: The paper has to be at least 15 pages long.

a. ⊢—— . . .
13 15 17 — authoritative reading

b. [//////////////—— . . .
13 15 17 — speaker insecurity reading

It is important to point out that the theoretical accounts do not make any predictions about which one of the readings is preferred if both the authoritative and the speaker insecurity reading is possible. But the availability of one reading can make the other reading undetectable. As Meyer and Sauerland (2009) argue, an ambiguous sentence will be judged true in a situation if it is true in that situation under the most accessible reading, even if it would be false under another reading. Thus we do not expect to find experimental evidence for the availability of the weaker speaker insecurity reading if the stronger authoritative reading is available.

For the combination possibility modal plus *at least 15 pages* the two readings predicted by the various analyses are illustrated in (51). Under the authoritative reading (51a), the sentence is in fact equivalent to the authoritative reading of necessity modal plus *at least 15 pages* in (50a), and we again expect No responses in the Under condition and Yes responses in the Over condition. Under the speaker insecurity reading, the speaker is unsure about the maximally-allowed length. While

[14] A third response along the lines of "I don't know", which would have enabled participants to hedge at some stimuli, was ruled out in order to avoid the risk that participants would choose it instead of seriously evaluating the scenario described in the stimuli.

lower values than 15 pages are within the deontic range, for all the speaker knows higher values than 15 pages might or might not be permissible. Therefore we would expect Yes responses in the Under condition, while we take both Yes and No responses in the Over condition to be compatible with this reading.

(51) ◇ + *at least*: The paper can be at least 15 pages long.

a. ⊢—— . . .
13 15 17 authoritative reading

b. . . . ——⊣////////////////
13 15 17 speaker insecurity reading

5.3 Results

Table 5.3 shows the results in terms of percentages of "Yes" Responses in each of the eight conditions. For the □ + *at least* combination, the modified number was clearly interpreted as denoting the lower bound of the deontic range, shown by the fact that the vast majority of participants accepted the description in the Over Condition and rejected it in the Under Condition. We also found a clear pattern of response rates in the ◇ + *at most* combination, where the number was interpreted as the upper bound, shown by the fact that in almost all cases participants rejected the description in the Over Condition and accepted it in the Under Condition. The other two combinations, ◇ + *at least* and □ + *at most*, exhibited a less robust contrast but nevertheless showed a clear tendency of interpretation. In the ◇ + *at least* combination participants tended to interpret the number as specifying the lower bound, shown by the fact that they accepted it in 84% of the Over Condition and only in 16% of the Under Condition. The □ + *at most* combination was mainly interpreted as specifying the upper bound, shown by the lower Yes rates in the Over Condition.

Discrepancy Condition	□ + *at least*	◇ + *at least*	□ + *at most*	◇ + *at most*
Over	90%	84%	16.67%	1.05%
Under	4.7%	16%	77.65%	93.91%

Table 2: Results from Experiment 3: Means of Yes Responses

We converted the Yes responses to 1 and a No responses to 0 and fitted the data to a linear mixed model with subjects and items as random effects and eight grouped conditions (e.g. □ + *at least* + under as one condition) as fixed effects. We have found that □ + *at least* and ◇ + *at most* are inversely parallel: the response pattern for the former combination in the over condition parallels the response pattern for the latter combination in the under condition, and the same for the former combination in the under condition and the latter in the over condition. ◇ + *at least* and □ + *at most* were likewise inversely parallel: The response pattern for the former combination in the over condition parallels the response pattern in the latter combination in the under condition as vice versa. In addition, □ + *at most* (in the over condition) patterns together with □ + *at least* (with respect to the under condition) and ◇ + *at most* (in the over condition), and ◇ + *at least* (in the over condition) patterns together with □ + *at least* (with respect to the over condition) and ◇ + *at most* (in the under condition). The other combinations were different from each other. (See detailed results in Appendices C.1-C.8.)

Returning to Table 2, notable response patterns are the ones for ◇ + *at least* and □ + *at most*, as they were not as robust as for the other two combinations. A closer look at the response pattern of each participant revealed that participants were not consistent in their response strategies; that is, we did not find that there were participants who consistently responded to the stimuli with, e.g., □ + *at most* in the Over condition with Yes (which happened in 16.67% of the cases). We likewise did not find specific stimuli that had led to a consistent response pattern that matched those low-percentage response.

5.4 Discussion

Before we discuss our findings vis-à-vis the predictions the various analyses make, we would like to spell out the assumptions we make regarding how the response pattern should be interpreted. Recall from the discussion of Experiment 2 that speakers tend to assume that their interlocutor has the epistemic authority with respect to information exchanged in discourse and consequently prefer an interpretation of an utterance that is consistent with their assumptions. This tendency is further enhanced in our task, as participants could only choose between Yes and No. We assume therefore that participants would prefer the authoritative reading when an utterance could be interpreted both under an authoritative reading and an speaker insecurity reading. This preference also leads to the assumption that an insecurity reading can only be detected if the authoritative reading isn't available.

In order to compare our findings with the readings in terms of upper and lower bound of permissible values predicted by the various analyses, we again summarize the predicted readings in (52) to (55) and include our results from Table 5.3 above.

(52) □ + *at least n*:

a.	4.7% < 90%	
b.	⊢—— ···	G&N, N, B/S/K, P, C&B
c.	[//////////////—— ···	G&N, B/S/K, P, C&B

(53) ◇ + *at least n*:

a.	16% < 84%	
b.	⊢—— ···	N, C&B
c.	··· ——[//////////////////	G&N, B/S/K, P, C&B

(54) □ + *at most n*:

a.	77.65% < 16.67%	
b.	··· ——⊣	B/S/K, C&B
c.	///////////////]—— ···	G&N, B/S/K, P, C&B

(55) ◇ + *at most n*:

a. 93.91% < 1.05%

b. . . . ——┤ G&N, N, P, C&B

c. . . . ——//////////////] G&N, N, B/S/K, P, C&B

Starting with the two combinations for which we found a clear response pattern, □ + *at least* and ◇ + *at most*, it turns out that the predictions of all analyses are compatible with our results. For □ + *at least n* in (52), greater values than *n* were accepted in 90% of the cases, while lower values were rejected in 96.3% of the cases, cf. (52a). This is expected under the authoritative reading (52b), which all of the analyses predict to be available. The additional, weaker reading in (52c) may be available as well but, as discussed in Section 5.2, is masked by the prefered authoritative reading.

Turning next to ◇ + *at most n* in (55), we found that only smaller values than *n* were accepted, cf. (55a). This result too is compatible with the predictions of all of the analyses. Although the neo-Gricean approaches (Büring, 2008; Schwarz, 2011, 2013; Kennedy, 2013) only predict the speaker insecurity reading (55c) to be available for this combination, participants accepting lower values in the vast majority of the cases is compatible with this reading. As explained before, under this reading, the speaker is not sure whether values in the epistemic range are permissible or not. Since participants were forced to decide between Yes and No and we don't know how they resolved this conflict, the high rate for Yes responses might be due to participants interpreting values in the epistemic range as permitted values in spite of the possibility that the speaker may not be entirely certain whether these values are permitted.

For the remaining two combinations, ◇ + *at least* and □ + *at most*, the response patterns we found were not as uniform as for the other two combinations.

For ◇ + *at least n* we found that greater values than *n* were accepted in 84% of the cases, but smaller values were still accepted in 16% of the cases, cf. (53a). This pattern is in fact not compatible with the predictions of any of the analyses. If the authoritative reading (53b) had been consistently available, as predicted by Nouwen (2010) and Coppock and Brochhagen (2013b), participants should always have opted for this stronger reading and thus should have consistently accepted higher values and rejected lower values. But if only the speaker insecurity reading (53c) had been available, as predicted by Geurts and Nouwen (2007) and the neo-Gricean approaches, lower values should have been consistently accepted. So this mixed pattern suggests that participants got both readings: in about a sixth of the cases (16%), participants got the speaker insecurity reading and accepted lower values, while in the majority of cases they got the authoritative reading, resulting in the rejection of lower values and acceptance of higher values.

For □ + *at most n* in (54), smaller values than *n* were accepted in 77.65% of the cases but greater values were still accepted in 16.67% of the cases, cf. (54a). This pattern is similarly incompatible with the predictions of any of the analyses. If the authoritative reading (54b) had been consistently available, as predicted by the neo-Gricean approaches and Coppock and Brochhagen (2013b), lower values should have been consistently accepted and higher values ruled out. But if only the speaker insecurity reading (54c) had been available, as predicted by Geurts and Nouwen (2007) and Nouwen (2010), higher values should have been consistently accepted. So similarly to the ◇

+ *at least* combination, this mixed pattern suggests that in about a sixth of the cases, participants got the speaker insecurity reading, while in the majority of cases they got the authoritative reading.

So how can we make sense of the data in light of the available theoretical analyses? In terms of the response pattern, two groups of superlative modifier-modal combinations seem to emerge. The first two combinations, $\Box$ + *at least* and $\Diamond$ + *at most*, lead to a clear response pattern favouring the authoritative reading. We thus conclude that for these combinations the authoritative reading is in fact consistently available. The remaining two combinations, $\Diamond$ + *at least* and $\Box$ + *at most*, show a mixed response pattern, which isn't compatible with the authoritative reading being consistently available, because if it had been available it would have trumped out the speaker insecurity reading. We thus interpret the mixed response pattern to indicate that the authoritative reading isn't consistently available. But how can the authoritative reading be available at all if it isn't consistently available? The answer, we believe, lies in the bias towards authoritative readings, to which our task might have led, as discussed in Section 4.4. Although the contexts left the knowledge of the speaker underdetermined, it could be that participants inferred that the speaker did in fact have sufficient information and thus participants would favour the authoritative reading.

If for $\Diamond$ + *at least* and $\Box$ + *at most* only the speaker insecurity reading is available, the only possible reading is one in which all values are potentially allowed, some within the deontic range and some within the epistemic range (that is, the speaker cannot rule out any values, modulo pragmatic restrictions involving relevance). This unrestricted reading might be felt to be at odds with the task and the speaker's utterance, which included two expressions that normally communicate restriction, namely deontic modals and superlative modifiers. Participants might therefore have decided that the reading conveying speaker ignorance was not felicitous and tried somehow to arrive at an authoritative reading. We hypothesize that participants resorted to a reanalysis of the modal to arrive at a clearer reading that better matched the task, an interpretive strategy we dub MODAL REANALYSIS. According to this strategy, a possibility modal is interpreted as necessity modal and vice versa to match the superlative modifier.[15] Since both *at least* and necessity modals express a lower bound, they seem to go together naturally. So do *at most* and possibility modals, which both express an upper bound. These are also the combinations for which we clearly found the authoritative reading to be available.

The modal reanalysis would explain why in the majority of the cases participants interpreted the numeral as specifying the lower bound when it was modified by *at least*, independently of whether it co-occurred with a necessity or a possibility modal, and why they interpreted the numeral as specifying the upper bound when it was modified by *at most*, whether co-occurring with a possibility or a necessity modal. This would also explain why we got mixed results for $\Diamond$ + *at least* and $\Box$ + *at most*: It seems that in the majority of cases, participants opted for the strong authoritative reading that necessitated modal reanalysis, but in the minority of cases, participants nevertheless opted for a compositional interpretation resulting in the weak reading conveying speaker ignorance.

Taking into account the bias towards authoritative readings that our experimental design might have led to and the resulting modal reanalysis participants might have resorted to, what can we

[15]This is reminiscent of Nouwen's 2010 claim that possibility modals are interpreted as expressing necessity when minimality is at stake. But we assume that this also occurs when *maximality* is at stake and that this reanalysis comes at a cost and is only used as a last resort strategy.

conclude from the results regarding which of the theoretical analysis makes the best predictions? Note that the grouping we found, whereby $\Box$ + *at least* and $\Diamond$ + *at most* have the authoritative reading and $\Diamond$ + *at least* and $\Box$ + *at most* don't, matches the results from Experiment 2. Thus the results of Experiment 3 also support the analyses of Geurts and Nouwen (2007) and Penka (2014), which are the only ones that correctly predict the grouping.

6 Experiment 4: Processing superlative modifier-modal combinations

6.1 Research question

In Experiment 3, we found clear response patterns suggesting that the authoritative reading is available for the $\Box$ + *at least* and $\Diamond$ + *at most* combinations. The results for the other two superlative modifier-modal combinations, $\Diamond$ + *at least* and $\Box$ + *at most*, were mixed and suggested that in addition to the semantic and pragmatic inferences predicted by the various analyses of these constructions, an additional interpretive strategy, modal reanalysis, emerged. The aim of Experiment 4 is to replicated the task in Experiment 3 in German and implement a self-paced reading task in order to determine whether the time-course of the detected interpretations could shed light on the semantic and pragmatic complexity of the inferences required to arrive at the attested interpretations.

6.2 Methods

We conducted an incremental self-paced reading experiment, in which 40 German speakers (27 Female, Mean Age: 24.5) read scenarios similar to the ones used in Experiment 3, as illustrated in (56-58).[16] The context was incrementally introduced on the screen, sentence by sentence, and the utterance and description sentences were introduced region by region, where each region was a syntactic constituent (see (57)-(58) for illustration.) Then, participants were asked whether the description was in accordance with the utterance.

(56) CONTEXT: (German)
John möchte einen Kuchen backen. | Deshalb fragt er seine Mutter nach dem Rezept für seinen Lieblingskuchen. | Nachdem er alle Schritte befolgt hat, schiebt er den Kuchen in den Ofen. | Da er nicht möchte, dass der Kuchen verbrennt oder roh ist, fragt er seine jüngere Schwester Lisa, wie lange er den Kuchen backen soll. | Sie sagt ihm: |

CONTEXT: (English translation)
John wants to bake a cake. | So he asks his mother for the recipe of his favourite cake. | After he follows all the instructions, he puts the cake in the oven. | As he doesn't want the cake

[16] The vertical lines represent breaks in the text, and participants were required to press the space bar to view the subsequence text chunk.

to be under- or over-baked, he asks his younger sister Lisa how long he should bake the cake for. | She tells him: |

(57) UTTERANCE:

Region 1	*Modal*	*Superlative Modifier*		*Region 4*	*Region 5*
„Der Kuchen	{darf / muss}	{mindestens / höchstens}	50 Minuten	im Ofen	backen."
"The cake	{can / has to}	for {at least / at most}	50 minutes	in the oven	bake."

"The cake {can / has to} bake in the oven for {at least / at most} 50 minutes."

(58) DESCRIPTION:

			Evaluation region
John	bäckt	den Kuchen	{47 / 53} Minuten lang.
John	baked	the cake	{47 / 53} minutes long.

"John baked the cake for {47 / 53} minutes."

Similarly to Experiment 3, the experimental design was 2 (modal conditions) × 2 (superlative modifier conditions) × 2 (under/over conditions). There were five observations for each one of the conditions, totalling in 40 target items. Eight lists were created using a latin square design so that each participant saw each context with only one modal-superlative modifier-over/under combination. In addition to the 40 target times, there were also 20 control items and 20 filler items. The control items consisted of 10 violation conditions, in which the description sentence violated the utterance, and 10 compliance conditions, in which the description sentence was in accordance with the utterance. The task was conducted at the Zukunftskolleg Psychology Lab at the University of Konstanz. Participants received €6 for their participation. The task took between 30 and 40 minutes.

The purpose of the online task was two-fold: First, we wanted to see whether some of the superlative modifier-modal combinations are more difficult to interpret and thus lead to processing difficulty. We expect such an effect to manifest itself in two main regions: the first one being the superlative modifier region in the utterance (57) and any spill-over effects in the following regions, and the second one being the EVALUATION REGION, in which a precise value is specified in the description sentence ({*47*/*53*} *minutes* in (58)) and where we hypothesize participants arrived at a decision. Second, as in Experiment 3 we aimed to determine what the preferred readings were for the various superlative modifier-modal combinations in terms of upper and lower bound of permissible values.

6.3 Results

We start with the readings in terms of lower and upper bound of permissible values. As shown in Table 3, in the □ + *at least* combination, the vast majority of participants accepted the description in the over condition and rejected it in the under condition. This means that the number was interpreted as denoting the lower bound of the deontic range. The number in the □ + *at most* combination was interpreted as specifying the upper bound, shown by the lower Yes rates in the over condition. In the ◇ + *at most* combination, the number was interpreted as the upper bound, shown by the fact that most participants rejected the description in the over condition and accepted it in the under condition. The pattern we found for ◇ + *at least* was less clear cut. Participants tended to choose the lower-bound reading, as shown by the fact that participants said Yes in 85.89% of the cases in the over condition. But in the under condition, in a reliable minority (23.36%) of the cases, values lower than the ones specified by the modified numeral (e.g. *47 minutes* in (58)) were accepted (W = 16297, $p<0.01$), which is incompatible with the lower-bound reading.

Discrepancy Condition	□ + *at least*	◇ + *at least*	□ + *at most*	◇ + *at most*
Over	94.56%	85.89%	5.13%	1.67%
Under	2.87%	23.36%	89.83%	93.57%

Table 3: Experiment 4: Means of “Yes” Responses

Similarly to the response rates in Experiment 3, we converted the Yes responses in Experiment 4 to 1 and a No responses to 0 and fitted the data to a linear mixed model with subjects and items as random effects and eight grouped conditions (e.g. □ + *at least* + under as one condition) as fixed effects. We have found that □ + *at least* and ◇ + *at most* are inversely parallel: the response pattern for the former combination in the over condition patterns the response pattern for the latter combination in the under condition, and the same for the former combination in the under condition and the latter in the over condition. ◇ + *at least* and □ + *at most*, on the other hand, are inversely parallel only in one respect: the response pattern for the former combination in the over condition parallels the response pattern in the latter combination in the under condition, but not vice versa. Instead, □ + *at most* in the over condition patterns together with □ + *at least* (with respect to the under condition) and ◇ + *at most* (in the over condition). The other combinations were different from each other (See detailed results in Appendices D.1.1-D.1.8.)

The reading time results of the utterance are given in Figure 6, where the reading times per region for each one of the modal-superlative modifier combinations are shown. We found that the reading times were significantly longer for *at least* following ◇ both for the superlative modifier region and region 4. An ANOVA of the utterance showed no main effect of superlative modifier or modal but there was an interaction between the two (Superlative modifier region: $F_{Within}(1,1548) = 5.36$, $p<0.05$; Region 4: $F_{Between}(1,36) = 7.14$, $p<0.05$; $F_{Within}(1,1503) = 25.21$, $p<0.01$). We also fitted the data to a linear mixed model with subjects and items as random effects and the the four different superlative modifier-modal combinations as fixed effects. We found that reading times for ◇ + *at least* were significantly longer than for all other combinations. Reading times for *at most* following a necessity modal were marginally longer than when following a possibility modal, with this difference bordering on statistical significance. Reading times for the other combinations were not different from each other. These findings were true both of the superlative modifier as well as spill over regions. (See detailed results in Appendices D.2-D.2.2.)

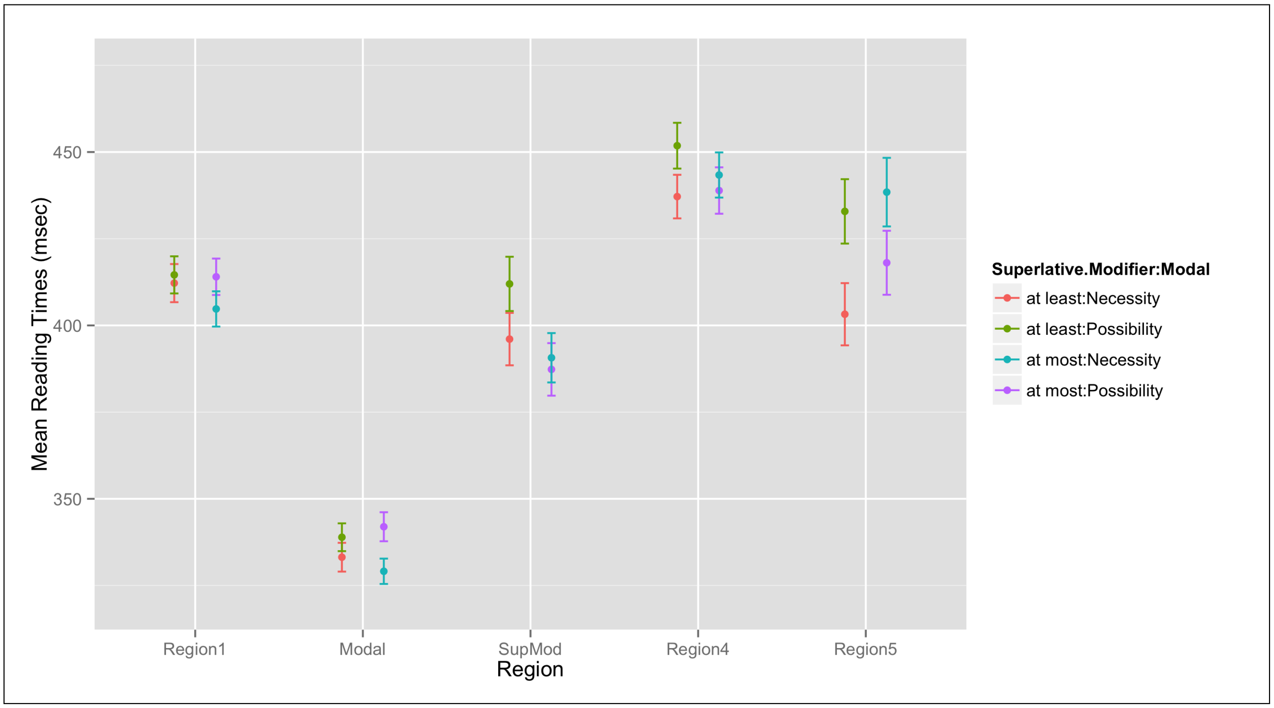


Figure 6: Experiment 4: Reading Times per Region in the Utterance

We also found a few significant differences in the reading times of the evaluation region of the description sentence. An ANOVA of the evaluation region shows no main effect of superlative modifier but a significant interaction between superlative modifier and modal ($F(1,1472)_{Within}$ = 6.68, p<0.01). A Wilcoxon signed-ranked test comparing specific conditions with a sufficient number of observations, revealed that arriving at a No answer in the under condition—that is, ruling out values lower than the numeral modified by *at least*, thereby interpreting it as specifying the lower bound— took significantly longer for ◇ + *at least* than for □ + *at least* (W = 14968, p<0.05). In addition, arriving at a Yes answer in the under condition took significantly longer for □ + *at most* than arriving at the same response for ◇ + *at most* (W = 15350.5, p<0.01). This difference suggests that deciding that lower values are permitted is more difficult for □ + *at most* than it is for ◇ + *at most*. No other comparisons reached significance. Although only marginally statistically significant (W = 14648, p=0.06), the mean reading time in the evaluation region leading to a "No" response seems shorter for □ + *at most* in the over condition (i.e. when interpreting the number as specifying the upper bound under an authoritative reading) than for ◇ + *at least* in the under condition (i.e. when interpreting the number as specifying the lower bound under an authoritative reading).

Condition	*Reply*	□ + *at least*	◇ + *at least*	□ + *at most*	◇ + *at most*
Over	Yes	930.64	1135.21	784.87	725.23
	No	584.20	987.10	950.99	1008.57
Under	Yes	685.54	1239.54	1207.14	1095.43
	No	938.37	1204.68	814.01	781.88

Table 4: Experiment 4: Mean Reading Times of the Evaluation Region in the Description Sentence

6.4 Discussion

Before we discuss the results in light of the predictions the various analyses make regarding the interpretation of modal-superlative modifier combinations, a few remarks about the implications of the bias towards epistemic authority we discussed in connection with Experiment 2 and 3 on the reading times are in order. Although in the present experiment, the context left open whether the speaker had the relevant knowledge, it is important to keep in mind that our task might have biased participants towards an authoritative reading. In addition to assuming that participants would prefer the authoritative reading over the speaker insecurity reading whenever it's compositionally available, we further assume that authoritative readings are faster and easier to compute than speaker uncertainty readings. Moreover, we interpret significantly prolonged reading times of a certain type of superlative modifier following a certain type of modal in comparison with the reading times of the same type of superlative modifiers following a different type of modal as an indication that the authoritative reading is compositionally unavailable for this combination.

In the discussion of the results, we start with the two combinations □ + *at least* and ◇ + *at most*. For these, we clearly found the authoritative readings to be available. We also didn't find any evidence in the reading times that would indicate that arriving at authoritative readings is associated with higher semantic or pragmatic complexity.

We turn next to the combination ◇ + *at least*, for which we found significantly longer reading times in the utterance compared to the other combinations. In line with the assumption that authoritative readings are faster and easier to compute, we conclude that the authoritative reading isn't compositionally available for this combination. This is also suggested by the mixed response pattern we found for ◇ + *at least* as well as the higher processing cost detected in the evaluation region. Recall that participants accepted lower numbers in 23.36% of the cases while in the remaining 76.64% they interpreted the modified number as the lower bound. This suggests that participants arrived at the speaker insecurity reading in about a quarter of the cases, while they got the authoritative reading in three quarter of the cases.[17] As in Experiment 3, we assume that the

[17]This is also compatible with the results in the over condition, where higher values were accepted in 85.89% of the cases. Assuming that participants chose to answer with Yes with a 50% chance when they had to decide about a value falling in the epistemic range (where the value might or might not be permissible according to what the speaker knows), we expect the overall rate of Yes responses in the over condition to be around 88% (23.36% x 0.5 + 76.64% x 1 = 88.32%), a calculated rate that is close to the attested one (85.89%). However, calculating the response rate for the over condition from the under condition for ◇ + *at least* or vice versa for □ + *at most* yields a prediction that isn't borne out in Experiment 3 for English. We leave for future research the question of how participants deal with values

availability of the authoritative reading would have trumped the speaker insecurity reading, and take the fact that we detected the insecurity reading in about a quarter of the cases to suggest that the authoritative reading isn't compositionally available. As discussed in section 5.4, we hypothesize that participants applied a last-resort interpretive strategy of modal reanalysis to arrive at an authoritative reading, which they may have felt better matched the task than the compositionally-available speaker insecurity reading. The longer reading times in the evaluation region for responses that favoured the authoritative reading for $\Diamond$ + *at least* suggest that this reading for this combination requires additional processing. Recall that arriving at a No answer in the under condition took significantly longer for $\Diamond$ + *at least* than for $\Box$ + *at least*. This means that arriving at the authoritative reading where the modified number specifies the lower bound of the deontic range was harder when *at least* followed $\Diamond$ than when it followed $\Box$. This supports our hypothesis that the authoritative reading of $\Diamond$ + *at least* was derived via a non-compositional interpretation strategy like modal reanalysis, which came at a processing cost.

The results for the combination $\Box$ + *at most* are interesting in several respects. First, the response pattern we found doesn't correlate to that of any of the other combinations. It was neither as clear-cut as for $\Box$ + *at least* and $\Diamond$ + *at most* nor as mixed as for $\Diamond$ + *at least*. Second, the reading times from the utterance suggest that processing *at most* following a necessity modal is somewhat harder than when following a possibility modal, but again the difference is not as striking as for *at least* following a possibility modal compared to it following a necessity modal. Taken together, the results suggest that the combination $\Box$ + *at most* is different from $\Box$ + *at least* and $\Diamond$ + *at most* as well as $\Diamond$ + *at least* regarding available readings. Although we didn't reliably detect the speaker insecurity reading, there is evidence that arriving at the authoritative reading for this combination comes with increased processing costs. Recall that in addition to the slightly longer reading times in the utterance, we also found that arriving at a Yes answer in the Under Condition took significantly longer for $\Box$ + *at most* than for $\Diamond$ + *at most*. This difference would be unexpected if the authoritative readings were equally available for both combinations of *at most* with modals. So the authoritative reading doesn't seem to be as readily available for $\Box$ + *at most* as it is for $\Diamond$ + *at most*. But at the same time, arriving at the authoritative reading doesn't seem to be as hard as for the combination $\Diamond$ + *at least*. And indeed the mean reading time in the evaluation region leading to a "No" response seems shorter for $\Box$ + *at most* in the over condition (i.e. when interpreting the number as specifying the upper bound under an authoritative reading) than for $\Diamond$ + *at least* in the under condition (i.e. when interpreting the number as specifying the lower bound under an authoritative reading). As mentioned in 6.3, however, this difference is only bordering on statistical significance.

Turning to the question which of the theoretical analyses accounts best for the three way grouping we found, we first note that several analyses are incompatible with the results of this experiment. Under the analysis of Nouwen (2010), under which superlative modifiers go well together with possibility modals, it is unexpected that $\Diamond$ + *at least* is harder to process. Coppock and Brochhagen (2013b) predict that for all of the combinations, both the authoritative and the speaker insecurity reading are available, which should make all combinations equally natural and comparable in processing, contrary to our results. Under neo-Gricean accounts of ignorance inferences of

---

falling in the epistemic range.

superlative modifiers (Büring, 2008; Schwarz, 2011, 2013; Kennedy, 2013), in contrast, we would expect that superlative modifiers yield authoritative readings in combination with necessity, but not possibility modals. This is at odds with our finding that $\Diamond$ + *at most* is not associated with higher processing costs and readily yields an authoritative reading. While the analysis of Geurts and Nouwen (2007) correctly predicts that $\Box$ + *at least* and $\Diamond$ + *at most* naturally go together to yield an authoritative reading and are fast to process, it doesn't explain the difference in reading times we found between $\Diamond$ + *at least* and $\Box$ + *at most*.

The analysis in Penka 2014 fairs best in being able to account for our results. Combining a neo-Gricean approach to ignorance inferences of superlative modifiers with a decompositional analysis of *at most*, this analysis predicts that *at most*, but not *at least*, is able to obviate ignorance inferences in combination with possibility modals. While this approach inherits from neo-Gricean approaches the prediction that ignorance is obviated and an authoritative reading arises if a superlative modifier is interpreted in the scope of a necessity modal, Penka argues that the antonymizing operator, which is one of the two ingredients of which *at most* is composed (*at least* being the second ingredient), exhibits certain scope preferences vis-à-vis modals. Since the German modal verb *müssen* 'have to', which we used as the necessity modal in our experiment, clearly prefers narrow scope with respect to negation and other negative operators (Penka and von Stechow, 2001), this makes the authoritative reading for $\Box$ + *at most*, which would require that the antonymizing operator (as well as *at least*) take scope below the necessity modal, strongly dispreferred. The dispreferred scope order required for the authoritative reading of $\Box$ + *at most* could explain that processing this combination under a bias towards authoritative readings is associated with somewhat higher costs.

As a consequence, there would in fact be two different strategies to arrive at an authoritative reading for the combinations for which this reading isn't readily available: Whereas in the case of $\Box$ + *at most* a dispreferred scope configuration is sufficient in order to derive the authoritative reading, there is no way to arrive at the authoritative reading for $\Diamond$ + *at least* in a compositional way. Thus, the modal reanalysis strategy might be what is behind the authoritative reading of $\Diamond$ + *at least*, whereas the authoritative reading of $\Box$ + *at most* is arrived at via a compositional but nevertheless cost-intensive mechanism. Assuming that compositionally-available readings, however dispreferred and convoluted, are nevertheless faster to arrive at than last-resort interpretive mechanisms that have no compositional basis, this would explain the difference in processing costs and available readings that we found for $\Diamond$ + *at least* and $\Box$ + *at most*.

7 General discussion

Taking together the results from all four experiments, what can we conclude regarding which of the analyses proposed in the literature fares best? Our first finding (from Experiment 1 and 2) points to ignorance inferences arising with superlative modifiers being pragmatic, rather than semantic, in nature. This is in line with neo-Gricean approaches and Coppock and Brochhagen's 2013b analysis couched in the framework of Inquisitive Semantics, but not with analyses hard-wiring ignorance inferences in the lexical semantics (Geurts and Nouwen, 2007) or deriving them at the syntax-semantics interface (Nouwen, 2010).

The second finding sheds light on the question which combinations of modals and superlative modifiers can suppress ignorance inferences and yield an authoritative reading. Here the results from Experiments 2 and 3 suggest that the combinations which have an authoritative reading are $\Box$ + *at least* and $\Diamond$ + *at most*, while $\Diamond$ + *at least* and $\Box$ + *at most* only allow for the speaker insecurity reading. This grouping corresponds to the one predicted by Geurts and Nouwen (2007) and Penka (2014). The other accounts each make some predictions which weren't borne out: Under Nouwen's 2010 analysis we would expect that both *at least* and *at most* would obviate ignorance inferences in combination with possibility modals, while under the basic neo-Gricean approach both superlative modifiers should obviate ignorance inferences in combination with necessity modals. Coppock and Brochhagen (2013b) predict that for all four combinations, both the authoritative and the speaker insecurity reading are available.

The more find-grained methodology used in Experiment 4, which also takes processing costs into account, suggests a further difference between the two combinations $\Diamond$ + *at least* and $\Box$ + *at most*. While for both the authoritative reading doesn't seem to be as easily available as for the other two combinations, arriving at the authoritative reading seems more demanding for $\Diamond$ + *at least* than for $\Box$ + *at most*. This is in line with Penka's (2014) analysis, where the authoritative reading of $\Box$ + *at most* necessitates a dispreferred scope order, while there is no way of deriving an authoritative reading for $\Diamond$ + *at least* in a compositional way.

All in all, the analysis that fares best vis-à-vis our experimental results is Penka's (2014), being the only one combining a pragmatic approach to ignorance inferences with the correct predictions which superlative modifier-modal combinations can suppress ignorance inferences. This account is also the only one able to explain the difference in processing costs that we found between the two combinations that do not easily obviate ignorance inferences. This analysis could potentially also account for another finding emerging from our studies that is worth pointing out: The readings we detected for the $\Box$ + *at most* combination in English in Experiment 3 were different than those we detected for German in Experiment 4. While in English the response pattern corresponded to the speaker insecurity reading in a reliable minority of the cases (16.67%), we weren't able to find clear evidence for this reading from the response pattern of the German participants. We thus detected a cross-linguistic difference in attested readings for the $\Box$ + *at most* combination, which we didn't find for any of the other combinations. Penka's (2014) analysis might provide an explanation for this cross-linguistic difference. Under this analysis the authoritative reading for $\Box$ + *at most* requires a scope configuration where the antonymizing operator is interpreted in the scope of the necessity modal. Further assuming, as Penka does, that deontic modals show certain scope preferences with respect to the antonymizing operator, the difference in the availability of the authoritative reading could be related to a difference in scope preferences between the necessity modals we used in our experiments in English and German. It might be that *müssen* in German somewhat easier allows dispreferred wide scope of the antonymizing operator than *have to* in English. To test whether this explanation for the cross-linguistic difference is on the right track, further experimental studies with different types of modals as well as different constructions which have been argued to involve the antonymizing operator would be necessary.

As discussed before, the primary objective of the experimental investigation in this article is to adjudicate between the different analyses of superlative modifiers. We had, however, an additional objective with a larger theoretical-empirical scope to put forth new methodologies (and modified methodologies with substantial improvement) that can be easily adapted and utilized in the growing field of experimental semantics and pragmatics. More specifically, we developed several types of experimental tasks to shed light on the extent to which superlative modifiers give rise to ignorance inferences when embedded under deontic modals and how these inferences interact with the epistemic state of the speaker (Experiment 2) and to determine the readings of the various superlative modifiers-deontic modal combinations (Experiment 3) and the time-course associated with each attested interpretation (Experiment 4). This is to our knowledge the first experimental paradigm that aims to examine ignorance inferences directly and explicitly and in addition to pit them against a contextually-inferred speaker's epistemic state. In Experiment 2, the results from the +knowledgeable condition clearly show lower coherence rates when the knowledgeable speaker produced an utterance that obligatorily conveys speaker ignorance. The results from the −knowledgeable condition, on the other hand, suggest that utterances produced by a contextually-inferred unknowledgeable speaker are considered degraded even if this contextually-inferred ignorance matches the ignorance inferences to which the utterance itself gives rise. This penalty for ignorance, so to speak, may be related to participants' preference for authoritative readings—and perhaps in general for authoritative and knowledgeable speakers—as attested in Experiments 3 and 4. We believe that the methodologies we developed in this study can be adapted to future research on various linguistic expressions that involve ignorance inferences and convey information about the speaker's epistemic state, e.g. modal expressions, vague predicates, and free choice items. We have also shown that the online methodology we developed reveals differences that don't necessarily manifest themselves in strategic verification and evaluation tasks.

Participants' strong preference for authoritative readings we detected in our tasks in Experiment 3 and 4 prevents us from discussing the availability and probability of ignorance inferences. It is only in cases where the authoritative readings were not compositionally available that we detected responses that correspond to the uncertainty readings, and even then they were detected in a (statistically significant) minority of the cases. This finding opens a new avenue for future research, in which participants are explicitly biased towards ignorance inferences. This bias may have to be done explicitly at the utterance level—e.g. be prefixing "I don't really know but ..." to the utterance—and not only inferred from the context, as was done in the experiments here. We expect a successful biasing toward ignorance inferences to increase the rate of the weak uncertainty readings in $\Diamond$ + *at least* and $\Box$ + *at most* and decrease the rate of the modal reanalysis strategy. An online task coupled with a bias toward ignorance inferences would reveal whether processing difficulties can be associated with accepting or rejecting values that lie in the epistemic range, i.e. values that for all the speaker knows may be permissible. We have found reliably longer processing times for rejecting lower values for $\Diamond$ + *at least* and for accepting lower values for $\Box$ + *at most*—that is, for arriving at the compositionally-unavailable authoritative reading. Facilitating ignorance inferences by framing the speakers as uncertain and their utterances as tentative may elucidate what the processing cost would be of assigning authoritative readings even when the context further biased participants toward speaker insecurity readings. Recall that in Experiment 2 in the −knowledgeable speaker condition, coherence rates of all utterances were low in comparison with the other speaker conditions. This was contrary to our predictions that a statement giving rise to ignorance inference uttered by an unknowledgeable speaker would receive higher coherence rates than in cases in which there was a mismatch between the contextual speaker's epistemic state and the one inferred from the utterance. The question is what processing effect would be associated with the degraded coherence rates and which response patterns we would observe with respect to the preferred readings, namely speaker insecurity or authoritative readings.

The question still remains whether participants' strong preference for authoritative readings is the sole factor that prevented participants from consistently assigning the speaker insecurity reading to $\diamond$ + *at least* and $\Box$ + *at most*. We hypothesize that they applied an interpretive strategy we dubbed as modal reanalysis. But this strategy may be just one of the situations in which participants in an experiment, and perhaps interlocutors in a conversation, may disregard semantically-entailed material in order to evaluate an utterance. Such behaviour has been found in various verification and interpretive tasks. For example, Tversky and Kahneman (1974) found that people overestimated the usefulness of some information that led them to prefer a less likely interpretation over a logically more accurate one. Additionally, Sanford and Moxey (2004) have found that participants ignored the monotonocity of some quantifiers and resulting entailments (e.g. that *At most n students came* is true in a situation in which 0 students came) and attributed it to the speaker's expectations and framing. Similarly to these cases, it may be that participants in Experiment 3 and 4 focussed on the (presumably authoritative) speaker's choice of superlative modifier, ignored the semantic material that the deontic modal contributed to the proposition, and arrived at the reading in which the superlative modifier and modal align in their range of permissible values.

Leaving these questions for future research, we showed that the experimental investigation of superlative modifiers and modals touches on important theoretical and experimental issues that pertain to ignorance inferences and modelling speakers' epistemic state in discourse while helping adjudicate between different theoretical analyses of these linguistic expressions.

A Appendix A: statistical tests of results from Experiment 1

A.1 Experiment 1 - scalar Items : Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: Entailments	Estimate	Std. Error	z value	Pr(> $|z|$)	Significance
Weaker	-1.6540	0.1285	-12.871	<2e-16	***
Similar	-0.2211	0.1320	-1.675	0.094	.
Stronger	-2.2273	0.1340	-16.615	<2e-16	***

Reference: Contradictions	Estimate	Std. Error	z value	Pr(> $|z|$)	Significance
Weaker	3.5699	0.1367	26.116	<2e-16	***
Similar	5.0028	0.1473	33.963	<2e-16	***
Stronger	2.9966	0.1380	21.717	<2e-16	***

Reference: Weaker	Estimate	Std. Error	z value	Pr(>—z—)	Significance
Similar	1.4329	0.1552	9.235	<2e-16	***
Stronger	-0.5733	0.1526	-3.757	0.000172	***

Reference: Similar	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
Weaker	-1.4329	0.1551	-9.236	<2e-16	***
Stronger	-2.0062	0.1597	-12.561	<2e-16	***

Reference: Stronger	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
Weaker	0.5733	0.1525	3.760	0.00017	***
Similar	2.0063	0.1596	12.568	<2e-16	***

A.2 Experiment 1 - Bare Numerals : Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: Entailments	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N-1	-3.4082	0.1352	-25.200	<2e-16	***
N	1.7744	0.1960	9.054	<2e-16	***
N+1	-4.0737	0.1355	-30.064	<2e-16	***

Reference: Contradictions	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N-1	1.8157	0.1294	14.035	<2e-16	***
N	6.9983	0.2091	33.472	<2e-16	***
N+1	1.1502	0.1267	9.081	<2e-16	***

Reference: N-1	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N	5.1825	0.2185	23.714	<2e-16	***
N+1	-0.6656	0.1469	-4.532	5.84e-06	***

Reference: N	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N-1	-5.1826	0.2186	-23.704	<2e-16	***
N+1	-5.8482	0.2189	-26.717	<2e-16	***

Reference: N+1	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N	5.8481	0.2187	26.738	<2e-16	***
N-1	0.6655	0.1467	4.536	5.74e-06	***

A.3 Experiment 1 - *at least*: Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: Entailments	Estimate	Std. Error	z value	Pr(> $\|z\|$)	Significance
N-1	-3.4082	0.1352	-25.200	<2e-16	***
N	0.95086	0.12923	7.358	1.87e-13	***
N+1	1.4408	0.1248	11.547	<2e-16	***

Reference: Contradictions	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	1.8157	0.1294	14.035	<2e-16	***
N	-4.27308	0.14198	-30.096	<2e-16	***
N+1	-3.7831	0.1350	-28.021	<2e-16	***

Reference: N-1	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N	0.4900	0.1484	3.302	0.000961	***
N+1	-2.4265	0.1509	-16.080	<2e-16	***

Reference: N	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	-0.48993	0.14846	-3.300	0.000966	***
N+1	-2.91651	0.15693	-18.585	<2e-16	***

Reference: N+1	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	2.4265	0.1510	16.074	<2e-16	***
N	2.9165	0.1569	18.587	<2e-16	***

A.4 Experiment 1 - *at most*: Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: Entailments	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	-3.4082	0.1352	-25.200	<2e-16	***
N	0.92813	0.13296	6.980	2.94e-12	***
N+1	3.2226	0.1317	24.469	<2e-16	***

Reference: Contradictions	Estimate	Std. Error	z value	Pr(>—z—)	Significance
N-1	1.8157	0.1294	14.035	<2e-16	***
N	-4.29578	0.14520	-29.585	<2e-16	***
N+1	-2.0013	0.1286	-15.562	<2e-16	***

Reference: N-1	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N	2.7357	0.1618	16.909	<2e-16	***
N+1	0.4411	0.1495	2.951	0.00317	**

Reference: N	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	-2.73564	0.16167	-16.922	<2e-16	***
N+1	-2.29446	0.15621	-14.688	<2e-16	***

Reference: N+1	Estimate	Std. Error	z value	Pr(> $\lvert z \rvert$)	Significance
N-1	-0.4412	0.1495	-2.951	0.00317	**
N	2.2945	0.1564	14.674	<2e-16	***

B Appendix B: statistical tests of results from Experiment 2

B.1 Experiment 2 - −Knowledgeable Condition: Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: □ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-0.59523	0.17976	-3.311	0.000929	***
□ + *at most*	-0.21792	0.17740	-1.228	0.219306	
◇ + *at most*	0.03431	0.17439	0.197	0.844044	

Reference: ◇ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at least*	0.5952	0.1798	3.311	0.000929	***
□ + *at most*	0.3773	0.1779	2.121	0.033916	*
◇ + *at most*	0.6295	0.1750	3.598	0.000321	***

Reference: □ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-0.3773	0.1779	-2.121	0.0339	*
□ + *at least*	0.2179	0.1774	1.228	0.2193	
◇ + *at most*	0.2522	0.1735	1.453	0.1461	

Reference: ◇ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at most*	-0.25224	0.17355	-1.453	0.146096	
◇ + *at least*	-0.62954	0.17497	-3.598	0.000321	***
□ + *at least*	-0.03432	0.17439	-0.197	0.843980	

B.2 Experiment 2 - ±Knowledgeable Condition: Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: □ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-0.506246	0.004796	-105.55	¡2e-16	***
□ + *at most*	-0.055287	0.005100	-10.84	<2e-16	***
◇ + *at most*	0.225409	0.005110	44.11	¡2e-16	***

Reference: ◇ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at least*	0.5103	0.2615	1.952	0.05098	.
□ + *at most*	0.4521	0.2632	1.718	0.08584	.
◇ + *at most*	0.7331	0.2667	2.749	0.00598	**

Reference: □ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-0.45081	0.26634	-1.693	0.0905	.
□ + *at least*	0.05881	0.26244	0.224	0.8227	
◇ + *at most*	0.28147	0.26849	1.048	0.2945	

Reference: ◇ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at most*	-0.281319	0.005143	-54.7	<2e-16	***
◇ + *at least*	-0.733405	0.004836	-151.6	<2e-16	***
□ + *at least*	-0.223079	0.005105	-43.7	<2e-16	***

B.3 Experiment 2 - +Knowledgeable Condition: Results of a Cumulative Link Mixed Model fitted with the Laplace approximation

Reference: □ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-1.3181	0.2979	-4.425	9.64e-06	***
□ + *at most*	-0.9804	0.3008	-3.259	0.00112	**
◇ + *at most*	-0.5426	0.2888	-1.878	0.06033	.

Reference: ◇ + *at least*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at least*	1.3180	0.2979	4.425	9.66e-06	***
□ + *at most*	0.3382	0.2823	1.198	0.2310	
◇ + *at most*	0.7754	0.2723	2.848	0.0044	**

Reference: □ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
◇ + *at least*	-0.3382	0.2823	-1.198	0.23103	
□ + *at least*	0.9799	0.3010	3.256	0.00113	**
◇ + *at most*	0.4373	0.2840	1.540	0.12367	

Reference: ◇ + *at most*	Estimate	Std. Error	z value	Pr(> |z|)	Significance
□ + *at most*	-0.4379	0.2838	-1.543	0.1229	
◇ + *at least*	-0.7756	0.2723	-2.848	0.0044	**
□ + *at least*	0.5426	0.2888	1.878	0.0603	.

C Appendix C: Statistical tests of results from Experiment 3

C.1 Experiment 3 - Results of a linear mixed model I. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NALO	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.99059	0.06590	34.01000	15.033	< 2e-16	***
NALU	-0.93247	0.08128	177.28000	-11.473	< 2e-16	***
NAMO	-0.84091	0.08313	101.78000	-10.116	< 2e-16	***
NAMU	-0.23635	0.09056	39.00000	-2.610	0.0128	*
PALO	-0.09659	0.08974	47.54000	-1.076	0.2872	
PALU	-0.77853	0.08615	60.60000	-9.037	7.85e-13	***
PAMO	-0.87532	0.08795	62.28000	-9.953	1.69e-14	***
PAMU	-0.03255	0.08218	123.13000	-0.396	0.6927	

C.2 Experiment 3 - Results of a linear mixed model II. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NALU	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.05812	0.06488	34.10000	0.896	0.3767	
NALO	0.93247	0.08128	177.28000	11.473	< 2e-16	***
NAMO	0.09156	0.08391	87.71000	1.091	0.2782	
NAMU	0.69611	0.08757	49.80000	7.949	2.05e-10	***
PALO	0.83588	0.08693	59.87000	9.615	9.44e-14	***
PALU	0.15394	0.08896	46.01000	1.730	0.0902	.
PAMO	0.05715	0.08751	59.28000	0.653	0.5163	
PAMU	0.89992	0.08324	96.48000	10.811	< 2e-16	***

C.3 Experiment 3 - Results of a linear mixed model III. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NAMO	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.14968	0.06247	32.91000	2.396	0.0224	*
NALU	-0.09156	0.08391	87.71000	-1.091	0.2782	
NALO	0.84091	0.08313	101.78000	10.116	2e-16	***
NAMU	0.60455	0.08704	42.88000	6.946	1.56e-08	***
PALO	0.74432	0.08640	50.91000	8.614	1.64e-11	***
PALU	0.06238	0.08698	43.87000	0.717	0.4770	
PAMO	-0.03441	0.08337	74.11000	-0.413	0.6810	
PAMU	0.80836	0.08540	57.10000	9.466	2.66e-13	***

C.4 Experiment 3 - Results of a linear mixed model IV. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NAMU	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.75423	0.06409	23.87000	11.768	2.01e-11	***
NAMO	-0.60455	0.08704	42.88000	-6.946	1.56e-08	***
NALU	-0.69611	0.08757	49.80000	-7.949	2.05e-10	***
NALO	0.23635	0.09056	39.00000	2.610	0.0128	*
PALO	0.13976	0.08526	53.95000	1.639	0.1070	
PALU	-0.54217	0.08839	34.02000	-6.134	5.76e-07	***
PAMO	-0.63897	0.08795	37.76000	-7.265	1.12e-08	***
PAMU	0.20381	0.08939	31.80000	2.280	0.0295	*

C.5 Experiment 3 - Results of a linear mixed model V. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PALO	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.89400	0.06359	27.33000	14.058	4.97e-14	***
NAMU	-0.13976	0.08526	53.95000	-1.639	0.107	
NAMO	-0.74432	0.08640	50.91000	-8.614	1.64e-11	***
NALU	-0.83588	0.08693	59.87000	-9.615	9.44e-14	***
NALO	0.09659	0.08974	47.54000	1.076	0.287	
PALU	-0.68194	0.08660	44.82000	-7.874	5.39e-10	***
PAMO	-0.77873	0.08617	50.63000	-9.037	3.86e-12	***
PAMU	0.06404	0.08759	41.59000	0.731	0.469	

C.6 Experiment 3 - Results of a linear mixed model VI. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PALU	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.21206	0.06379	25.51000	3.324	0.00269	**
PALO	0.68194	0.08660	44.82000	7.874	5.39e-10	***
NAMU	0.54217	0.08839	34.02000	6.134	5.76e-07	***
NAMO	-0.06238	0.08698	43.87000	-0.717	0.47705	
NALU	-0.15394	0.08896	46.01000	-1.730	0.09025	.
NALO	0.77853	0.08615	60.60000	9.037	7.85e-13	***
PAMO	-0.09679	0.08631	47.78000	-1.122	0.26767	
PAMU	0.74598	0.08661	44.71000	8.613	4.74e-11	***

C.7 Experiment 3 - Results of a linear mixed model VII. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PAMO	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.11527	0.06346	28.11000	1.816	0.080	.
PALU	0.09679	0.08631	47.78000	1.122	0.268	
PALO	0.77873	0.08617	50.63000	9.037	3.86e-12	***
NAMU	0.63897	0.08795	37.76000	7.265	1.12e-08	***
NAMO	0.03441	0.08337	74.11000	0.413	0.681	
NALU	-0.05715	0.08751	59.28000	-0.653	0.516	
NALO	0.87532	0.08795	62.28000	9.953	1.69e-14	***
PAMU	0.84277	0.08597	52.97000	9.803	1.69e-13	***

C.8 Experiment 3 - Results of a linear mixed model VIII. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PAMU	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.95804	0.06384	25.70000	15.008	3.15e-14	***
PAMO	-0.84277	0.08597	52.97000	-9.803	1.69e-13	***
PALU	-0.74598	0.08661	44.71000	-8.613	4.74e-11	***
PALO	-0.06404	0.08759	41.59000	-0.731	0.4688	
NAMU	-0.20381	0.08939	31.80000	-2.280	0.0295	*
NAMO	-0.80836	0.08540	57.10000	-9.466	2.66e-13	***
NALU	-0.89992	0.08324	96.48000	-10.811	< 2e-16	***
NALO	0.03255	0.08218	123.13000	0.396	0.6927	

D Appendix D: Statistical tests of results from Experiment 4

D.1 Linear mixed models on the "Yes" responses

D.1.1 Experiment 4 - Results of a linear mixed model I. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NALO	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	9.402e-01	2.090e-02	8.343e+02	44.991	<2e-16	***
NALU	-9.050e-01	2.893e-02	1.556e+03	-31.285	<2e-16	***
NAMO	-8.632e-01	2.896e-02	1.554e+03	-29.807	<2e-16	***
NAMU	-5.506e-02	2.847e-02	1.556e+03	-1.934	0.0533	.
PALO	-7.037e-02	2.879e-02	1.562e+03	-2.444	0.0146	*
PALU	-6.785e-01	2.882e-02	1.557e+03	-23.545	<2e-16	***
PAMO	-9.013e-01	2.862e-02	1.561e+03	-31.496	<2e-16	***
PAMU	-1.551e-03	2.897e-02	1.560e+03	-0.054	0.9573	

D.1.2 Experiment 4 - Results of a linear mixed model II. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NALU	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	3.520e-02	2.110e-02	8.502e+02	1.668	0.0957	.
NALO	9.050e-01	2.893e-02	1.556e+03	31.285	< 2e-16	***
NAMO	4.182e-02	2.911e-02	1.554e+03	1.437	0.1510	
NAMU	8.499e-01	2.862e-02	1.556e+03	29.701	< 2e-16	***
PALO	8.346e-01	2.893e-02	1.559e+03	28.847	< 2e-16	***
PALU	2.265e-01	2.898e-02	1.562e+03	7.815	9.99e-15	***
PAMO	3.740e-03	2.876e-02	1.561e+03	0.130	0.8966	
PAMU	9.034e-01	2.912e-02	1.562e+03	31.022	< 2e-16	***

D.1.3 Experiment 4 - Results of a linear mixed model III. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NAMO	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.07702	0.02115	854.20000	3.641	0.000288	***
NALU	-0.04182	0.02911	1554.40000	-1.437	0.150984	
NALO	0.86317	0.02896	1553.50000	29.807	< 2e-16	***
NAMU	0.80811	0.02866	1556.60000	28.200	< 2e-16	***
PALO	0.79280	0.02898	1561.10000	27.361	< 2e-16	***
PALU	0.18463	0.02901	1560.30000	6.365	2.57e-10	***
PAMO	-0.03808	0.02881	1563.70000	-1.322	0.186394	
PAMU	0.86162	0.02916	1560.70000	29.550	< 2e-16	***

D.1.4 Experiment 4 - Results of a linear mixed model IV. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: NAMU	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
NAMO	-0.80811	0.02866	1556.60000	-28.200	<2e-16	***
NALU	-0.84993	0.02862	1555.50000	-29.701	<2e-16	***
NALO	0.05506	0.02847	1556.50000	1.934	0.0533	.
PALO	-0.01531	0.02848	1558.90000	-0.538	0.5909	
PALU	-0.62348	0.02852	1561.60000	-21.863	<2e-16	***
PAMO	-0.84619	0.02830	1559.90000	-29.901	<2e-16	***
PAMU	0.05351	0.02867	1563.90000	1.866	0.0622	.

D.1.5 Experiment 4 - Results of a linear mixed model V. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PALO	Estimate	Std. Error	df	t value	Pr(> |*t*|)	Significance
(Intercept)	0.86981	0.02090	834.30000	41.624	<2e-16	***
NAMU	0.01531	0.02848	1558.90000	0.538	0.5909	
NAMO	-0.79280	0.02898	1561.10000	-27.361	<2e-16	***
NALU	-0.83462	0.02893	1558.60000	-28.847	<2e-16	***
NALO	0.07037	0.02879	1561.70000	2.444	0.0146	*
PALU	-0.60816	0.02882	1556.60000	-21.104	<2e-16	***
PAMO	-0.83088	0.02860	1553.40000	-29.053	<2e-16	***
PAMU	0.06882	0.02896	1554.80000	2.376	0.0176	*

D.1.6 Experiment 4 - Results of a linear mixed model VI. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PALU	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.26165	0.02095	838.20000	12.491	< 2e-16	***
PALO	0.60816	0.02882	1556.60000	21.104	< 2e-16	***
NAMU	0.62348	0.02852	1561.60000	21.863	< 2e-16	***
NAMO	-0.18463	0.02901	1560.30000	-6.365	2.57e-10	***
NALU	-0.22645	0.02898	1562.10000	-7.815	9.99e-15	***
NALO	0.67854	0.02882	1557.20000	23.545	< 2e-16	***
PAMO	-0.22271	0.02864	1554.50000	-7.777	1.33e-14	***
PAMU	0.67699	0.02900	1553.60000	23.348	< 2e-16	***

D.1.7 Experiment 4 - Results of a linear mixed model VII. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PAMO	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	0.03894	0.02065	811.40000	1.885	0.0598	.
PALU	0.22271	0.02864	1554.50000	7.777	1.33e-14	***
PALO	0.83088	0.02860	1553.40000	29.053	< 2e-16	***
NAMU	0.84619	0.02830	1559.90000	29.901	< 2e-16	***
NAMO	0.03808	0.02881	1563.70000	1.322	0.1864	
NALU	-0.00374	0.02876	1561.10000	-0.130	0.8966	
NALO	0.90125	0.02862	1560.90000	31.496	< 2e-16	***
PAMU	0.89970	0.02879	1555.20000	31.253	< 2e-16	***

D.1.8 Experiment 4 - Results of a linear mixed model VIII. Abbreviations: N = □; P = ◇; AL = *at least*; AM = *at most*; O = over; U = under

Reference: PAMU	Estimate	Std. Error	df	t value	Pr(> |t|)	Significance
(Intercept)	9.386e-01	2.115e-02	8.542e+02	44.377	<2e-16	***
PAMO	-8.997e-01	2.879e-02	1.555e+03	-31.253	<2e-16	***
PALU	-6.770e-01	2.899e-02	1.554e+03	-23.348	<2e-16	***
PALO	-6.882e-02	2.896e-02	1.555e+03	-2.376	0.0176	*
NAMU	-5.351e-02	2.867e-02	1.564e+03	-1.866	0.0622	.
NAMO	-8.616e-01	2.916e-02	1.561e+03	-29.550	<2e-16	***
NALU	-9.034e-01	2.912e-02	1.562e+03	-31.022	<2e-16	***
NALO	1.551e-03	2.897e-02	1.560e+03	0.054	0.9573	

D.2 Linear mixed models on reading times per region in the utterance

D.2.1 The superlative modifier (third) region

Reference: □ + *at least*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	416.580	25.999	16.023	<2e-16	***
□ + *at most*	-2.284	13.432	-0.170	0.6689	
◇ + *at least*	15.077	13.409	1.124	<2e-16	***
◇ + *at most*	-3.746	13.398	-0.280	0.2368	

Reference: ◇ + *at least*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	431.66	26.04	16.579	<2e-16	***
□ + *at most*	-17.36	13.49	-1.287	<2e-16	***
□ + *at least*	-15.08	13.41	-1.124	<2e-16	***
◇ + *at most*	-18.82	13.41	-1.404	<2e-16	***

Reference: □ + *at most*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	414.296	26.049	15.905	<2e-16	***
□ + *at least*	2.284	13.432	0.170	0.1055	
◇ + *at least*	17.361	13.486	1.287	<2e-16	***
◇ + *at most*	-1.462	13.526	-0.108	0.0546	.

Reference: ◇ + *at most*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	412.835	26.032	15.859	<2e-16	***
◇ + *at least*	18.823	13.408	1.404	<2e-16	***
□ + *at most*	1.462	13.526	0.108	0.0510	.
□ + *at least*	3.746	13.398	0.280	0.6595	

D.2.2 The fourth region

Reference: □ + *at least*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	442.751	16.789	26.371	<2e-16	***
□ + *at most*	6.541	12.945	0.505	0.5387	
◇ + *at least*	19.080	12.872	1.482	<2e-16	***
◇ + *at most*	6.876	12.683	0.542	0.1066	

Reference: ◇ + *at least*	Estimate	Std. Error	t value	Pr(> |*t*|)	Significance
(Intercept)	461.24	16.33	28.237	<2e-16	***
□ + *at most*	-11.98	13.36	-0.897	<2e-16	***
□ + *at least*	-18.73	13.07	-1.434	<2e-16	***
◇ + *at most*	-12.72	13.05	–0.975	<2e-16	***

Reference: □ + *at most*	Estimate	Std. Error	t value	Pr(> |t|)	Significance
(Intercept)	449.2610	16.3898	27.411	<2e-16	***
□ + *at least*	-6.7569	13.0894	-0.516	0.3458	
◇ + *at least*	11.9777	13.3591	0.897	<2e-16	***
◇ + *at most*	-0.7441	13.1838	-0.056	0.0544	.

Reference: ◇ + *at most*	Estimate	Std. Error	t value	Pr(> |t|)	Significance
(Intercept)	448.5169	16.1725	27.733	<2e-16	***
◇ + *at least*	12.7217	13.0466	0.975	<2e-16	***
□ + *at most*	0.7441	13.1838	0.056	0.0545	.
□ + *at least*	-6.0129	12.8758	-0.467	0.4425	

References

Alxatib, S. (2013). *'Only' and association with negative quantifiers*. Ph. D. thesis, MIT.

Beck, S. (2012). DegP scope revisited. *Natural Language Semantics 20*, 227–272.

Büring, D. (2007a). Cross-polar nomalies. In *Proceedings of Semantics and Linguistics Theory 17*, University of Connecticut.

Büring, D. (2007b). *More* or *less*. In *Proceedings the Chicago Linguistic Society Annual Meeting*, Chicago.

Büring, D. (2008). The least *at least* can do. In C. B. Chang and H. J. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, Somerville, MA, pp. 114–120. Cascadilla Proceedings Project.

Christensen, R. H. B. (2012). Ordinal Regression Models for Ordinal Data R, Package Version 2012.01-19.

Cohen, A. and M. Krifka (2011). Superlative quantifiers as modifiers of meta-speech acts. *The Baltic International Yearbook of Cognition, Logic and Communication 6*, 1–56.

Coppock, E. and T. Brochhagen (2013a). Diagnosing truth, interactive sincerity, and depictive sincerity. In *Proceedings of Semantics and Linguistic Theory 23*.

Coppock, E. and T. Brochhagen (2013b). Raising and resolving issues with scalar modifiers. *Semantics & Pragmatics 6*, 1–57.

Cummins, C. and N. Katsos (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics 27*(3), 271 –305.

Doran, R., G. Ward, M. Larson, Y. McNabb, and R. E. Baker (2012). A novel experimental paradigm for distinguishing between 'what is said' and 'what is implicated'. *Language 88*(1), 124–154.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics*, pp. 71–120. Palgrave Mc Millan.

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

Geurts, B. and R. Nouwen (2007). 'At least' et al.: The semantics of scalar modifiers. *Language 83*(3), 533–559.

Heim, I. (2001). Degree Operators and Scope. In C. Féry and W. Sternefeld (Eds.), *Audiatur vox sapientiae: A Festschrift for Arnim von Stechow*, Studia grammatica 52, pp. 214–239. Berlin: Akademie Verlag.

Katsos, N. and D. V. Bishop (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition 120*(1), 67–81.

Katsos, N. and N. Smith (2010). Pragmatic tolerance or a speaker–comprehender asymmetry in the acquisition of informativeness. In F. K., K. Iserman, and L. Keil (Eds.), *Proceedings of the 34th Annual Boston Conference in Language Development*, MA, USA. Cascadilla Press.

Katzir, R. (2007). Structurally defined alternatives. *Linguistics and Philosophy 30*(6), 669–690.

Kennedy, C. (2013). A "de-Fregean" semantics for modified and unmodified numerals. University of Chicago.

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases.* Plenum Press.

Meyer, M.-C. and U. Sauerland (2009). A pragmatic constraint on ambiguity detection: A rejoinder to Büring and Hartmann and to Reis. *Natural Language and Linguistic Theory 27*, 139–150.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics 3*, 1–41.

Penka, D. (2014). 'At most' at last. Handout for talk at Sinn und Bedeutung 19, Universität Göttingen, September 15, 2014.

Penka, D. and A. von Stechow (2001). Negative Indefinita unter Modalverben. In R. Müller and M. Reis (Eds.), *Modalität und Modalverben im Deutschen*, Linguistische Berichte, Sonderheft 9, pp. 263–286. Hamburg: Helmut Buske Verlag.

Sanford, A. and L. Moxey (2004). Exploring quantifiers: Pragmatics meets the psychology of comprehension. In I. A. Noveck and D. Sperber (Eds.), *Experimetal Pragmatics*, pp. 116–137. Palgrave Macmillan.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy 27*(3), 367–391.

Schwarz, B. (2011). Remarks on class B numeral modifiers. Handout of a talk at the workshop *Indefinites and Beyond*, Universität Göttingen, November 19, 2011.

Schwarz, B. (2013). 'At least' and quantity implicature: Choices and consequences. In M. Aloni, M. Franke, and F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium*, pp. 187–194.

Traugott, E. C. and R. B. Dasher (2002). *Regularity in Semantic Change*. Cambridge University Press.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.