



Ignorance implicatures of modified numerals

Alexandre Cremers¹ · Liz Coppock² · Jakub Dotlačil³ · Floris Roelofsen⁴

Accepted: 27 April 2021 / Published online: 16 August 2021

© The Author(s) 2021

Abstract

Modified numerals, such as *at least three* and *more than five*, are known to sometimes give rise to ignorance inferences. However, there is disagreement in the literature regarding the nature of these inferences, their context dependence, and differences between *at least* and *more than*. We present a series of experiments which sheds new light on these issues. Our results show that (a) the ignorance inferences of *at least* are more robust than those of *more than*, (b) the presence and strength of the ignorance inferences triggered by both *at least* and *more than* depends on the question under discussion (QUD), and (c) whether ignorance inferences are detected in a given

We would like to thank Ivano Ciardelli for his contribution to the design of the experiments and various aspects of the project. For all the precious discussions, we would like to thank Stavroula Alexandropoulou, Dominique Blok, Paul Boersma, Brian Buccola, Andreas Haida, Dora Mihoc, Rick Nouwen, Patrick Spaans, Benjamin Spector, Jennifer Spenader, Matthijs Westera and the audiences of SALT29, the Göttingen LinG Colloquium, the XPrag workshop “The meaning of numerals”, McGill linguistics group, the LINGUAE seminar (Paris), and the first InqBNB workshop. The experiments presented in this paper were made possible by Alex Drummond’s Ibox Farm platform. We would like to thank our editors, Katrin Schulz and Stefan Kaufmann, as well as all anonymous reviewers, whose comments and suggestions greatly improved the paper. We gratefully acknowledge funding from the Netherlands Organization for Scientific Research (NWO), the European Research Council (ERC, grant agreement number 680220), and the Agence Nationale de la Recherche (FrontCog Project ANR-17-EURE-0017 and ProbaSem project ANR-19-CE28-0004-01).

✉ Alexandre Cremers
alexandre.cremers@gmail.com

Liz Coppock
eecoppock@gmail.com

Jakub Dotlačil
j.dotlacil@gmail.com

Floris Roelofsen
floris.roelofsen@gmail.com

¹ ENS-Paris and Vilnius University, Vilnius, Lithuania

² Boston University, Boston, MA, USA

³ Utrecht University, Utrecht, Holland

⁴ University of Amsterdam, Amsterdam, Holland

experimental setting depends partly on the task that participants are asked to perform (e.g., an acceptability task versus an inference task). We offer an Optimality Theoretic account of these findings. In particular, the task effect is captured by assuming that in performing an acceptability task, participants take the *speaker's* perspective in order to determine whether an expression is optimal given a certain epistemic state, while in performing an inference task they take the *addressee's* perspective in order to determine what the most likely epistemic state of the speaker is given a certain expression. To execute the latter task in a fully rational manner, participants have to perform higher-order reasoning about alternative expressions the speaker could have used. Under the assumption that participants do not always perform such higher-order reasoning but also often resort to so-called *unidirectional optimization*, the task effect finds a natural explanation. This also allows us to relate our finding to asymmetries between comprehension and production that have been found in language acquisition.

1 Introduction

Modified numerals like *at least n* and *more than n* contrast in interesting ways with each other and with bare numerals in the implicatures that they give rise to. The basic empirical picture assumed in most work on the topic is as follows:

- (1) a. The house has three bedrooms. \rightsquigarrow exactly three $\not\rightsquigarrow$ ignorance
 b. The house has at least three bedrooms. $\not\rightsquigarrow$ exactly three \rightsquigarrow ignorance
 c. The house has more than two bedrooms. $\not\rightsquigarrow$ exactly three $\not\rightsquigarrow$ ignorance

At least three and *more than two* do not trigger an 'exactly three' implicature, in contrast with the bare numeral *three*. On the other hand, it is often assumed that *at least* triggers an ignorance implicature, unlike *more than* or bare numerals.

While the contrast in 'exactly *n*' implicatures is uncontroversial, the ignorance inferences triggered by modified numerals are subject to ongoing debate. Do superlative modifiers always convey ignorance? Do comparative modifiers never do so? And, moving from empirical issues to theoretical ones, how do these ignorance inferences come about? What is the semantics of numerals and of superlative and comparative modifiers? And what are the crucial pragmatic factors involved in deriving ignorance? In order to delineate the contribution of the present paper, let us look at these issues in some more detail.

1.1 The empirical debate

Consider the following scenario. The emergency department of a certain hospital is required to have three physicians present at all times. Following a complaint by a patient who had to wait for several hours on Tuesday last week, police officers are investigating whether the requirement was satisfied. An employee of the hospital tells the officers:

- (2) a. There were at least three physicians last Tuesday.
 b. There were more than three physicians last Tuesday.

Can the officers conclude that the employee doesn't know exactly how many physicians there were? The received view is that the superlative modifier *at least* in (2a) does imply such ignorance, while the comparative modifier *more than* in (2b) doesn't (see, e.g., Geurts and Nouwen 2007; Büring 2008; Cummins and Katsos 2010; Nouwen 2010; Kennedy 2015). This assumption, however, has recently been challenged by various authors.

Westera and Brasoveanu (2014) (henceforth W&B) hold that both *at least* and *more than* can in principle generate ignorance implicatures; whether they do is determined by the question under discussion (QUD). More specifically, W&B propose that in response to a *how many* question like (3a), both *at least* and *more than* imply ignorance, while in response to a *polar* question like (3b), neither does.

- (3) a. How many physicians were there last Tuesday? \rightsquigarrow ignorance
 b. Were there enough physicians last Tuesday? $\not\rightarrow$ ignorance

Mayr and Meyer (2014) (henceforth M&M) also propose that ignorance implicatures are sensitive to the QUD. However, the specific empirical generalization that they suggest is different from that of W&B. According to them, both types of modified numerals convey ignorance in response to *how many* questions such as (3a), while only *at least* conveys ignorance in response to polar questions such as (3b).

A third account where ignorance implicatures are sensitive to the QUD has been proposed by Coppock and Brochhagen (2013b) (henceforth C&B). However, the specific predictions of this account again differ from those of W&B and those of M&M. Namely, on the account of C&B neither *at least* nor *more than* conveys ignorance in response to polar questions, while only *at least* conveys ignorance in response to *how many* questions. Note that this is in a sense the reverse image of M&M's view: QUD sensitivity is not assumed for *more than* but rather for *at least*.

Cummins, Sauerland and Solt (2012) (henceforth CSS) suggest that, besides the QUD, there may be another factor determining whether ignorance implicatures arise or not. Namely, focusing on the case of *comparative* modified numerals, CSS propose that *more than n* only conveys ignorance if the numeral *n*, which we will refer to as the *base numeral*, is not made particularly salient in the immediately preceding discourse context.¹ To illustrate the effect of salience of the base numeral, consider the following minimal pair from CSS:

- (4) A: We can provide lunches for the guests. How many will stay for lunch?
 B: More than sixty will stay for lunch.
- (5) A: We can provide sixty lunches for the guests. How many will stay for lunch?
 B: More than sixty will stay for lunch.

¹ Ignorance implicatures of *superlative* modified numerals are not directly discussed by CSS. However, they build on prior work by Cummins (2011), who proposes that *at least n* and *at most n* "can be regarded as disjunctions, at some level of representation" and assumes that, as such, they always yield ignorance implicatures, in line with the received view (pp. 79–80).

Table 1 Empirical views on the ignorance implicatures of modified numerals

| Modifier | QUD | Received view | W&B | M&M | C&B | CSS |
|------------------|-----------------|---------------|-----|-----|-----|-----|
| <i>At least</i> | <i>how many</i> | Yes | Yes | Yes | Yes | Yes |
| | polar | | No | Yes | No | Yes |
| <i>More than</i> | <i>how many</i> | No | Yes | Yes | No | Yes |
| | polar | | No | No | No | No |

CSS propose that comparative modified numerals trigger ignorance inferences only to the extent that no alternative explanation can be found for the use of a relatively uninformative expression (e.g., *more than sixty* rather than a bare numeral). CSS assume that priming of the numeral *sixty*, as in (5), constitutes such an alternative explanation. So they predict that (4) carries a stronger ignorance implicature than (5). More generally, the strength of ignorance implicatures arising from comparative modified numerals is diminished in this view if the base numeral is particularly salient in the given discourse context.

Another take on the contrast between (4) and (5), as pointed out by Cummins (2013, p.105), is that A's mentioning of the fact that sixty lunches can be provided, in (5), affects the QUD addressed by B. In particular, it may implicitly raise the polar QUD of whether up to sixty guests or more will stay. Under this view, the contrast between (4) and (5) is in line with the empirical generalizations concerning *more than* of W&B and M&M: ignorance inferences arise in response to *how many* QUDs but not in response to polar QUDs. Note that, even if contrasts like that between (4) and (5) are fundamentally due to the different salience levels of the number *sixty* and not to the presence of an implicit polar QUD, in any context with an explicit polar QUD as to whether a certain quantity exceeds a given threshold *n* or not, the number *n* is particularly salient. This means that, albeit indirectly, CSS also predict QUD sensitivity of ignorance implicatures. In particular, they predict that comparative modified numerals do not give rise to ignorance implicatures in the presence of a polar QUD.²

The experiments presented here will be primarily concerned with QUD sensitivity. In our theoretical proposal, which builds on Cummins (2013), both QUD sensitivity and salience of the base numeral will be taken into account. The different empirical views on QUD sensitivity discussed above are summarized in Table 1. The fact that different authors have disagreed on the basic data shows the need for experimental investigation.

Ideally, such an investigation should not only establish which of the proposed empirical generalizations is correct (if any), but also explain why there has been so much disagreement. One of our main findings will be that whether ignorance inferences are detected depends (among other things) on the task that participants are asked to perform, in particular on whether they are given an acceptability task or an inference task. Our interpretation of this finding will be that, whether ignorance inferences

² Another observation in Cummins et al. (2012) is that non-round comparative modified numerals such as *more than 53* are somewhat degraded when the non-round numeral is not salient or particularly relevant in the context of utterance. Their explanation for this observation is that *more than 53* competes in such contexts with *more than 50*, which involves the round number 50. The fact that *more than 53* is slightly more informative may be outweighed by the fact that *more than 50* is easier to process.

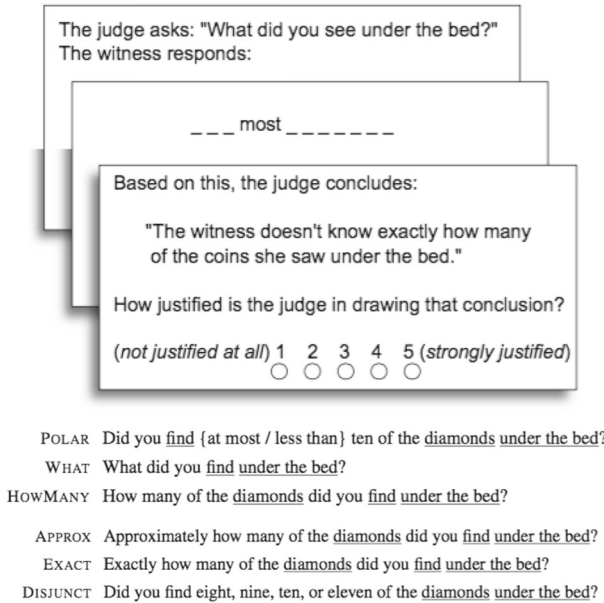


Fig. 1 Westera and Brasoveanu’s (2014) experimental design

are derived depends on whether participants take the speaker’s or the addressee’s perspective, and in the latter case, on whether they perform higher-order pragmatic reasoning or not. This, then, offers a possible explanation of the fact that the empirical generalizations proposed in previous work, which were mostly based on introspective judgments, do not align.

1.2 Existing experimental work

To our knowledge, the only existing experimental study of modified numerals that explicitly probes the QUD sensitivity of their ignorance implicatures is Westera and Brasoveanu (2014) (W&B).³

As depicted in Fig. 1, W&B presented experimental participants with courtroom dialogues such as the one in (6):

- (6) Judge: What did you see under the bed?
Witness: I saw at most 10 diamonds under the bed.

The type of question was experimentally manipulated as indicated in Fig. 1, and the witness’s response always contained either a superlative modified numeral, *at most 10*, or a comparative modified numeral, *less than 10*. Participants were then told that the judge concluded that the witness does not know exactly how many diamonds she saw under the bed, and were asked how justified the judge was in drawing that conclusion, on a 1–5 scale.

³ For a recent overview of experimental studies of the semantics and pragmatics of modified numerals in general, see Nouwen et al. (2019).

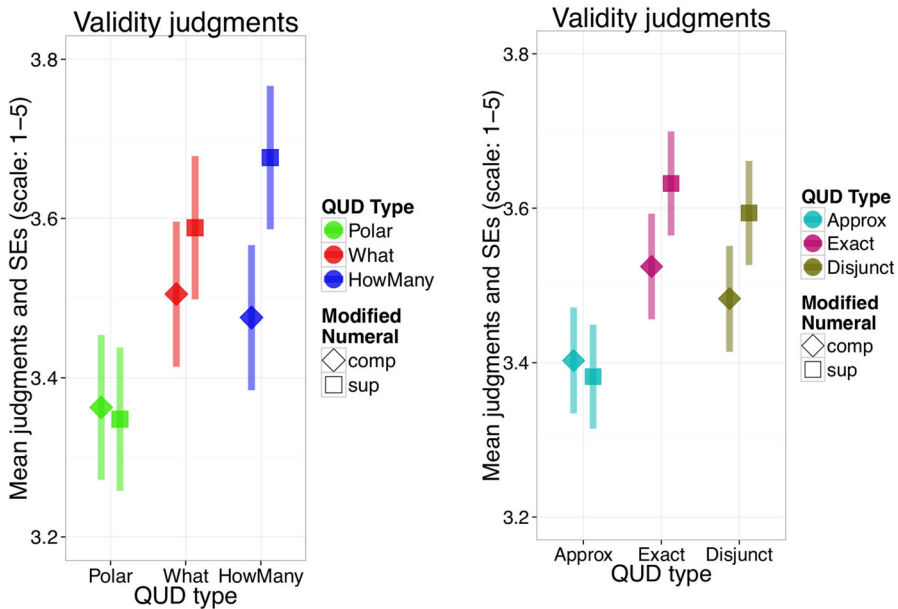


Fig. 2 Westera and Brasoveanu's (2014) results

The results are given in Fig. 2. Among the six types of context that W&B considered, a significant difference between superlative and comparative modifiers was only found in *how many* contexts. This result is surprising on the received view, on which superlative modifiers convey ignorance and comparative ones do not. From these results, W&B conclude that, contra the received view, comparative modifiers sometimes *do* signal ignorance (e.g. in response to *how many exactly* questions), and also that superlative modifiers sometimes *don't* convey ignorance (most clearly in polar contexts).

These findings leave open a number of empirical questions. First, the absence of a significant difference between the two types of modifier (in most contexts) is a null result, and may be due in part to the experimental setting. That is, it may well be that in other experimental settings, differences between superlative and comparative modifiers can more easily be detected. This will indeed be the case in some of the experiments presented below.

Second, it should be noted that the polar contexts that W&B considered are quite special, because they involve responses that literally 'echo' the question, as exemplified in (7).

- (7) A: Did you find at most 10 of the diamonds under the bed?
B: I found at most 10 of the diamonds under the bed.

This 'echoing' may have certain idiosyncratic effects. In order to draw general conclusions about the behaviour of superlative and comparative modifiers in polar contexts it would be preferable to avoid such effects.

The remainder of the paper is structured as follows. First, in Sect. 2 we present a series of experiments that build on the initial work of W&B, confirming the QUD sensitivity that they found but also identifying differences between superlative and comparative modifiers, as well as the task-dependence mentioned earlier. In Sect. 3 we spell out an Optimality Theoretic account of our experimental findings, building on Cummins (2013). Finally, Sect. 4 compares our proposal with a number of previous accounts in light of the experimental results, and Sect. 5 concludes.

2 New experimental data

2.1 Experiment 1

2.1.1 Goals

The goal of our first experiment was to establish some basic facts regarding ignorance implicatures triggered by *more than* and *at least* in the context of two types of QUD: *how many* questions and polar questions. In particular, we wanted to see whether the patterns found by W&B, reviewed above, could be replicated using a different experimental setup which does not explicitly focus on the speaker's ignorance, and is closer to the truth-value judgment task commonly used in the experimental literature on implicatures.

Participants had to judge the acceptability of statements made during a card game. This setup made it very easy to construct complete information and ignorance scenarios. In the former kind of scenario, a player makes a statement about her cards at a point in the game where she knows what all her cards are. In the latter, she makes a statement about her cards at a point where she has not seen all her cards yet. Both scenarios can be represented visually, by means of a simple picture (see Fig. 3 for an example).

In this setup, we can test the presence of ignorance inferences by studying how people judge interpretations that violate such inferences. This is a common strategy in other pragmatic studies, for example, on scalar implicatures (Bott and Noveck 2004). The specific design that we used has been employed successfully in a study of ignorance implicatures in Dieuleveut et al. (2019).

Given that the scenario is made very explicit, it is possible to ensure that polar questions and *how many* questions have exactly the properties that we want them to have. In particular, we can specify details of the scenario in such a way that it is extremely unlikely that *how many* questions would be treated underlyingly as polar questions.

We can also avoid echoic responses to polar questions that W&B used, capitalizing on the fact that *I have at least six clubs* entails a 'yes' answer to the question *Will you win this round?* (see below for details).

Finally, our setup allows us to test differences between answers which exactly match the 'yes' answer to a polar question and answers which provide more information than necessary.

Mary's cards:

2
♣

6
♣

7
♣

J
♣

Q
♣

K
♣

Sue: "How many face cards do you have?"

Mary: "At least three of my eight cards are face cards."

Is Mary's answer appropriate?

Not at all

1

2

3

4

5

Completely

Fig. 3 Example of a target item in Experiment 1 (*how many* QUD, *Superlative* quantifier, *Ignorance* situation)

2.1.2 Methodology

Participants: 50 participants were recruited on MTurk for the experiment. The participants were self-reported adult English speakers (age: 18–67, mean: 35). They received \$2.50 for their participation in the experiment.

Materials and procedure: The task of every participant was to judge the appropriateness of statements on a 5-point scale (1 signaled the lowest level of appropriateness, and 5 signaled the highest level). Statements were presented as answers to questions and were accompanied by a pictorial situation. An example of an item is given in Fig. 3.

The following background story preceded the actual experiment:

Mary is playing a card game online. Each round consists of a betting phase and a playing phase.

Betting phase: At the beginning of the betting phase, each player receives 8 cards: 6 visible (to them) and 2 face down. Players look at their first six cards. Based on what they see, all players place a bet on their chance to win the round. Then the players see their seventh card and place a second bet. A final bet is placed after players see their eighth and last card. When all players have placed their last bets, the playing phase begins.

Playing phase: The exact rules for this part of the game are not relevant here. Clubs are the trump suit (a club card beats any card from other suits), and if Mary receives six or more clubs, she is bound to win (she has an unbeatable strategy).

Scoring: If Mary wins the round, she earns the number of points she bet, plus one bonus point for each face card she had (Jack, Queen or King). Face cards only influence scoring and do not play any special role in the playing phase. Sue is Mary's best friend. She knows the rules of the game and Mary's strategy, but is not currently playing. Sue walks into the room at some point during the betting phase and asks Mary a question. Sue cannot see Mary's screen, so she doesn't know which cards Mary got or whether Mary has seen all of her cards yet when she asks her question. Mary has no reason to hide information from Sue.

Each stimulus consisted of (i) a picture showing Mary's cards, (ii) a question posed by Sue, and (iii) Mary's answer to the question (see Fig. 3).

The experiment consisted of 108 experimental stimuli and 51 fillers. The experimental stimuli tested 3 quantifiers (Condition: QUANTIFIER), 3 QUD-Answer relations (Condition: QUD) and 4 situations (Condition: SITUATION). Each combination of conditions was repeated three times.

The following three quantifiers were used for the QUANTIFIER condition: *at least n*, *more than n*, and the bare numeral *n*. For example, for the situation depicted in Fig. 3, these three versions were used:

- (8) a. At least three of my eight cards are clubs.
 b. More than two of my eight cards are clubs.
 c. Three of my eight cards are clubs.

The QUD manipulation consisted of 3 conditions. In the first (*how many*), Sue asked a *how many* question about the number of face cards (see Fig. 3). The other two conditions involved polar questions, and differed with respect to whether the answer was directly relevant to the question. In the *PolarRelevant* condition, Sue asked a polar question (*Will you win this round?*) and Mary responded using the relevant information to determine winning (e.g., *Yes, at least six/more than five of my eight cards are clubs*).⁴ In the *PolarOverInf* condition, Sue asked the same polar question and Mary responded providing more information than necessary (e.g., *Yes, at least seven/more than six of my eight cards are clubs*).

The SITUATION manipulation consisted of 4 conditions. In the first case (*False*), all the cards were revealed and Mary's answer was simply false: there were 2 clubs/face cards less than what she reported. In the second case (*Ignorance*), Mary's answer was given during the uncovering phase (one or two cards were still covered). The cards revealed so far supported the lower bound set by her response. In the third case (*Exceed*), all the cards were revealed and there was one more club/face card than the lower bound conveyed by Mary's answer. In the last condition (*Exact*), all the cards were revealed and the number of club/face cards was exactly the lower bound conveyed by the answer. The number of relevant cards in each condition is summarized in Table 2.

The actual experiment was preceded by 6 practice trials. The practice trials were similar to the experimental stimuli in their design, but unlike experimental items they

⁴ Recall that in order to be sure she will win, Mary had to have 6 or more clubs.

Table 2 Experiment 1: Number of cards matching Mary's answer in each situation. For *Ignorance*, only a range could be known because some cards were face down. For *how many* questions different base numbers were used, but they all followed the same pattern

| QUD | Expression | Situation | | | |
|-----------------------|---|--------------|--------------|---------------|------------------|
| | | <i>False</i> | <i>Exact</i> | <i>Exceed</i> | <i>Ignorance</i> |
| <i>how many</i> | <i>three, at least three, more than two</i> | 1 | 3 | 4 | [3,5] |
| <i>Polar Relevant</i> | <i>six, at least six, more than five</i> | 4 | 6 | 7 | [6,8] |
| <i>Polar OverInf</i> | <i>seven, at least seven, more than six</i> | 6 | 7 | 8 | [7,8] |

indicated whether Mary's answer was appropriate or not and why this was so. None of the practice trials used any modified numerals.

Pre-processing and data analysis: We first removed the fastest and slowest 1% of responses as outliers. We then calculated individual error rates on control items for which we expected a clear answer. In such cases, '3' never counted as a correct answer (hence the theoretical chance level is at 40%). Four participants were removed because their error rate on control items was at least one standard deviation above the mean error rate (threshold: 22.7%). The mean error rate on remaining participants was 7.6%.

For statistical analysis, responses were treated as a continuous variable and normalized by participant. We fitted mixed-effects linear models with the `lme4` package in R (R Core Team 2014; Bates et al. 2015b), following the recommendations of Bates et al. (2015a) regarding the specification of the random effects structure. For the calculation of *p*-values, we approximated the *t*-distribution with a Gaussian curve. (This approximation should not be problematic given the number of participants.)

2.1.3 Results

The detailed results on target conditions are presented in Fig. 4. In the rest of this section and for the statistical analyses, we leave aside the *False* situations, which gave rise to very low ratings in all conditions, as expected.

We fitted a linear mixed-effects model on responses to the *Exact*, *Exceed* and *Ignorance* conditions for sentences with *at least* and *more than*, in response to *how many* and polar questions. All three factors (QUANTIFIER, QUD, and SITUATION) were treatment-coded, with *Ignorance* as the SITUATION reference level, *more than* as the QUANTIFIER reference level, and *how many* as the QUD reference level. The results are given in Table 3.

The middle column in Fig. 4 shows that, across all QUD conditions (corresponding to the rows), ratings for *more than* sentences in the *Exact*, *Exceed*, and *Ignorance* situations are all roughly on a par with each other. Thus *more than* does not appear to exhibit a preference for ignorance situations. This conclusion is supported by our statistical analysis shown in Table 3. With *more than* sentences, across all QUD conditions, the differences between the *Ignorance* condition (reference level) and the non-*Ignorance* conditions are non-significant (with one possible exception: for sentences giving a relevant answer to a polar QUD, the ratings were 0.41 scale points higher on average

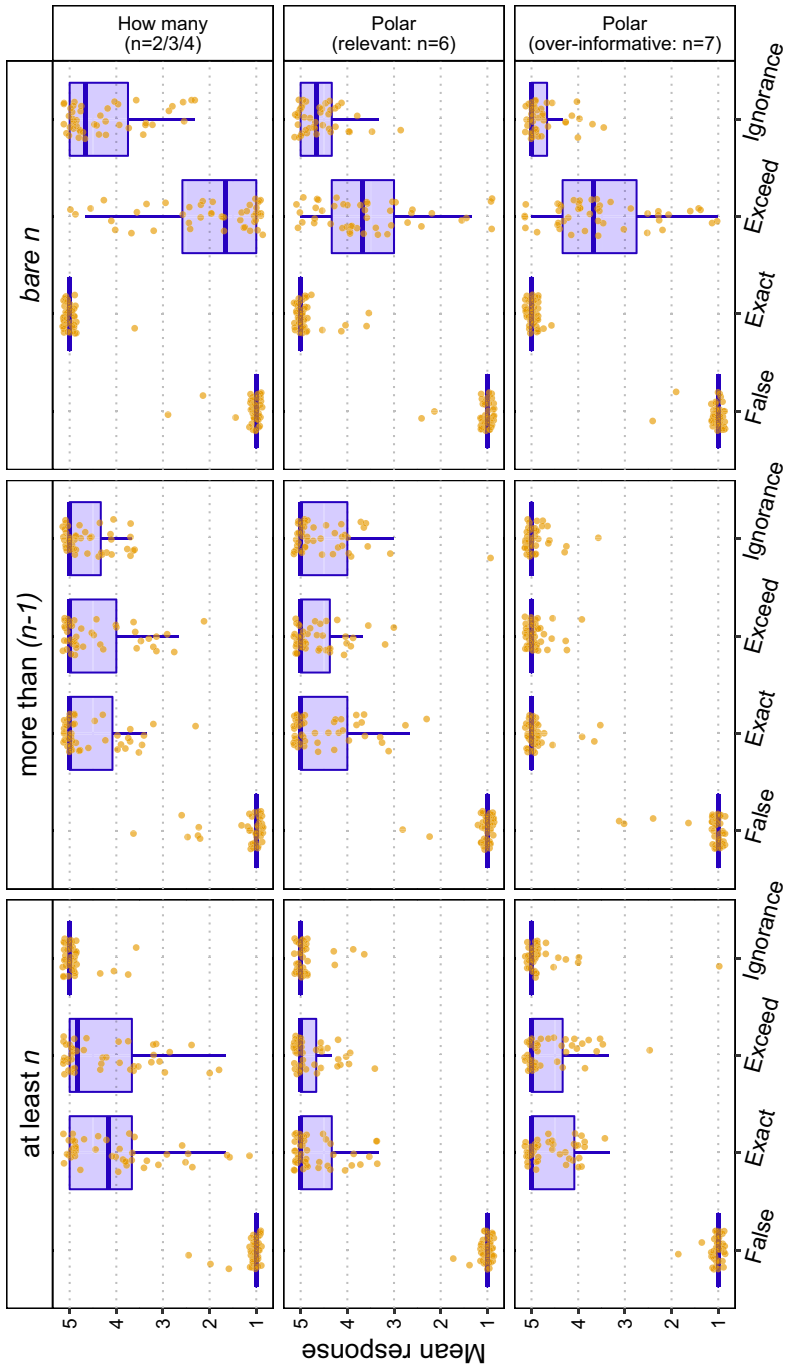


Fig. 4 Mean acceptability of Mary's answer in each target condition by individual participants in Experiment 1

Table 3 Estimates for the fixed effects of the model fitted on target conditions for Experiment 1. All factors treatment-coded with reference levels Situation[*Ignorance*], QUANTIFIER[*more than*], QUD[*how many*]

| | β | <i>t</i> -value | <i>p</i> -value |
|--|---------|-----------------|-----------------|
| (Intercept) | -0.05 | -0.6 | .58 |
| SITUATION[<i>Exact</i>] | -0.07 | -0.6 | .57 |
| SITUATION[<i>Exceed</i>] | -0.24 | -1.8 | .073. |
| QUANTIFIER[<i>at least</i>] | 0.42 | 3.2 | .001** |
| QUD[<i>PolarRelevant</i>] | -0.11 | -0.8 | .44 |
| QUD[<i>PolarOverInf</i>] | 0.41 | 3.5 | .001** |
| SITUATION[<i>Exact</i>]:QUANTIFIER[<i>at least</i>] | -1.13 | -5.5 | < .001*** |
| SITUATION[<i>Exceed</i>]:QUANTIFIER[<i>at least</i>] | -0.66 | -3.8 | < .001*** |
| SITUATION[<i>Exact</i>]:QUD[<i>PolarRelevant</i>] | -0.02 | -0.1 | .93 |
| SITUATION[<i>Exceed</i>]:QUD[<i>PolarRelevant</i>] | 0.41 | 2.3 | .019* |
| SITUATION[<i>Exact</i>]:QUD[<i>PolarOverInf</i>] | 0.06 | 0.4 | .69 |
| SITUATION[<i>Exceed</i>]:QUD[<i>PolarOverInf</i>] | 0.24 | 1.6 | .11 |
| QUANTIFIER[<i>at least</i>]:QUD[<i>PolarRelevant</i>] | 0.10 | 0.6 | .54 |
| QUANTIFIER[<i>at least</i>]:QUD[<i>PolarOverInf</i>] | -0.48 | -3.1 | .002** |
| SITUATION[<i>Exact</i>]:QUANTIFIER[<i>at least</i>]:QUD[<i>PolarRelevant</i>] | 0.90 | 3.8 | < .001*** |
| SITUATION[<i>Exceed</i>]:QUANTIFIER[<i>at least</i>]:QUD[<i>PolarRelevant</i>] | 0.29 | 1.3 | .20 |
| SITUATION[<i>Exact</i>]:QUANTIFIER[<i>at least</i>]:QUD[<i>PolarOverInf</i>] | 0.82 | 3.5 | < .001*** |
| SITUATION[<i>Exceed</i>]:QUANTIFIER[<i>at least</i>]:QUD[<i>PolarOverInf</i>] | 0.32 | 1.6 | .12 |

in the *Exceed* situation, compared to the *Ignorance* situation, and this difference is significant, though only at the 0.05 level, which is arguably too lax given the number of comparisons we are testing).

Things were different in the case of *at least*, as seen in the leftmost column in Fig. 4, showing that ratings for *at least* sentences in non-*Ignorance* situations were generally lower than in *Ignorance* situations. This is supported by the statistical analysis. Across all QUD types, the ratings for sentences in the *Exact* condition were significantly lower compared to the *Ignorance* reference level (between 0.82 and 1.13 scale points lower on average). The finding supports previous claims that precise knowledge negatively affects the acceptability of *at least* more than the acceptability of *more than*.

The relative acceptability of *at least* in the *Exceed* situation, compared to the *Ignorance* situation, depended on the nature of the QUD. Ratings for *at least* sentences were significantly lower in the *Exceed* situation in the context of a *how many* question (0.66 scale points lower on average), but not in the context of a polar question (0.29 or 0.32 scale points lower on average, depending on whether the answer was relevant). Hence, broadly speaking, the deviance due to ignorance inferences of *at least* is not as strong in the context of polar questions.

As a post hoc analysis, we fitted a model on the results of *at least* which pooled the *Exact* and *Exceed* situations together as a new condition *Precise*. The *PolarRelevant* and *PolarOverInf* conditions were also pooled together as a new condition, *Polar*. The model did not explain significantly less variance than a full model ($\chi^2(5) = 8.5$, $p = .13$), and it showed a strong negative effect of the *Precise* situation ($t = -5.2$, $p < .001$) as well as a strong positive interaction of the *Precise* situation with the *Polar* QUD ($t = 4.9$, $p < .001$).

We analyzed the results for bare numerals separately. Unsurprisingly, we observed that the *Exact* condition was rated higher than the *Ignorance* condition ($t = 5.2$, $p <$

.001), which in turn was higher than the *Exceed* condition ($t = 13, p < .001$). The *False* condition was still lower than the *Exceed* condition ($t = 5.7, p < .001$). Interestingly, the intermediate *Exceed* and *Ignorance* conditions were both rated higher with *Polar* QUDs ($t = 3.7$ and $t = 7.8$, respectively, both $p < .001$), while the *False* and *Exact* conditions were unaffected by the QUD type (both $t < .4, p > .72$).

2.1.4 Discussion

We saw no trace of ignorance inferences with *more than*, except for a small trend in responses to *HowMany* questions (in the *Exceed* situation). By contrast, *at least* gave rise to clear ignorance inferences, as evidenced by its lower acceptability in *Exceed* and *Exact* situations compared to *Ignorance* situations, although the ignorance inferences were found to be weaker in the context of a polar question. The present results are therefore at odds with the results reported by W&B, who report similar ignorance inferences with *at least* and *more than*.

Note that the absence of ignorance inferences with *more than* cannot be fully explained by a lack of statistical power, since we were able to clearly detect QUD effects on *at least*. Our experimental design differed from W&B's on a few key points, which could explain the observed differences. First of all, our participants' task was to evaluate the appropriateness of an utterance given the knowledge state of a speaker, while W&B required participants to judge an inference about a speaker's knowledge state, given their utterance. In short, we could say that our task was more "speaker oriented" (evaluating whether an utterance is acceptable is similar to considering whether one could utter such an utterance) while W&B's task was more "hearer oriented". Second, we situated our conversations in a more casual setup (a discussion between friends), while W&B used a very formal context (a witness testifying in front of a judge). Third, we only tested positive modified numerals (*at least* and *more than*), whereas W&B only tested their negative counterparts (*at most* and *less than*).⁵

To find out which of these factors is responsible for the discrepancies between our results and those reported by W&B, in the next two experiments we adopt a design much closer to W&B's original design (while of course still sidestepping the methodological issue concerning 'echo' responses discussed above). To foreshadow the results, the first factor ("speaker" vs. "hearer" orientation) seems to best explain the differences, but a contrast between positive and negative modified numerals may have played a role as well.

Before turning to the next experiment, let us briefly discuss the results obtained with bare numerals, even though this is not the main focus of the paper. In general, our findings seem to support the view that numerals start with a one-sided denotation and receive their exact reading through implicatures (Horn 1972, see Spector 2013 for a review) over the view that the basic meaning of numerals is exact and that the 'at least' reading is derived by pragmatics (Breheny 2008), although we will see in Sect. 3 that the results can also be captured under semantic ambiguity accounts such as Geurts

⁵ However, Coppock and Brochhagen (2013a) tested both positive and negative modified numerals and found a stronger contrast between *at most* and *fewer than* than between *at least* and *more than*, suggesting that polarity alone cannot explain the difference between our results and W&B (if anything, they should have observed a stronger contrast than we did).

(2006) or Kennedy (2015). First, we see that the *Exceed* condition (which involved a situation with more than n items) is rated much higher than the *False* condition (which involved a situation with less than n items), suggesting that bare numerals are more sensitive to violations in one direction than in the other. Second, the implicature account offers an explanation for the fact that *Exceed* is less accepted than *Ignorance* (where the exact reading is neither supported nor excluded). Indeed, both *Exceed* and *Ignorance* violate the upper-bounding implicature (the speaker believes the stronger alternative to be false; secondary implicature in the nomenclature of Sauerland 2004), but *Exceed* further violates the primary ignorance implicature (the speaker does not believe the stronger alternative to be true). The fact that *Ignorance* is less accepted than *Exact* could be indicative of either secondary implicatures or a semantic ambiguity between one-sided and two-sided readings (Geurts 2006; Kennedy 2015). Third, we found that both the contrasts between *Exceed* and *Ignorance* and between *Ignorance* and *Exact* were sensitive to the QUD, as implicatures would be. Overall, these results are surprising because in most experimental settings, numerals diverge from archetypal scalar implicatures (see, e.g., Papafragou and Musolino 2003 for acquisition, Huang and Snedeker 2009 for processing; but see Panizza et al. 2009 for a view in line with our finding). In particular, Dieuleveut et al. (2019), using a task very similar to ours, detected primary implicatures with the quantifier *some* but not with numerals.⁶

2.2 Experiment 2

2.2.1 Goals

The main goals of our second experiment were (a) to understand why the results of our first experiment differed so sharply from W&B's results, and (b) to attempt once more, in a different experimental setting, to detect ignorance inferences with comparative modifiers, in order to test their possible QUD-dependency. For this purpose, we adopted a design very close to W&B's. In particular, we switched to an inferential task, in which participants had to evaluate how much a speaker knew given her answer to a question.

The characters involved in this experiment were police officers or investigators (who asked questions) and witnesses (who responded to these questions). This brings us closer to W&B's judge/witness situation, while allowing more variety in the situations we considered. The conversation was also likely to be more casual than in a courtroom.

As discussed above, one other possible source of the discrepancies between the results of our first experiment and those of W&B, is the fact that we considered positive modifiers (*at least* and *more than*) while they considered negative ones (*at most* and *less than*). In our second experiment, we tested both positive and negative modifiers.

Finally, as in our first experiment, we wanted to get a better understanding of the potential contrast between propositions which match a complete answer to a given polar question and propositions which are over-informative. In the previous experiment

⁶ Something that might explain this difference is that Dieuleveut et al. (2019) used binary response options while we collected graded judgments.

there was a possible confound: we saw that the acceptability of *more than* was overall degraded when it combined with a numeral which was not salient, even though it provided a complete answer to the question, whereas the over-informative condition was overall well accepted, possibly because it involved the salient numeral *six*. In our second experiment, we varied the polar question in such a way that *more than n* was sometimes but not always over-informative, while keeping *n* salient. How this was achieved is explained in more detail below.

2.2.2 Manipulating relevance through context

We manipulated the relevance of the response provided by a modified numeral to a polar question by varying contexts instead of the numeral (as we did in Experiment 1). Schematically, if *n* is the numeral that was modified, “upward” contexts involved a polar question equivalent in terms of resolution conditions to “is it at least *n*?”, while “downward” contexts involved a polar question equivalent to “is it at most *n*?”.⁷ However, we made sure that the questions did not explicitly contain a modified numeral. Upward contexts typically established a requirement for *n* items, whereas downward contexts specified a maximum number of items. This way, we could simply ask whether the rules had been respected, without having to mention any modified numeral in the question.

Let us illustrate this with some concrete examples. In an upward context, an investigator may ask whether there were enough seat belts for the five passengers in a car. If the witness answers that there were *more than five* seat belts, she gave an over-informative positive answer. *Fewer than five* would be a negative relevant answer. Finally, *at least five* would be a positive relevant answer, but *at most five* would not resolve the question (there may have been five, and there may have been fewer than five).

The roles are reversed in a downward context, in which an investigator may for instance ask whether the maximum load of ten people was exceeded during an elevator incident. In this case, saying that *more than ten* people were present in the elevator is a positive relevant answer, while *fewer than ten* is a negative over-informative answer. Finally, *at most ten* is a negative relevant answer, while *at least ten* does not resolve the question.

To sum up, our manipulation allowed us to test comparatives both as relevant and as over-informative answers, and the numeral being modified was always salient and round, whether the resulting construction matched a complete answer or not. Superlatives always corresponded to relevant answers (cases in which superlatives did not answer the question were excluded from the experiment, since they would introduce an orthogonal issue).

⁷ Concretely, the questions we used were phrased positively, so that positive quantifiers would always be associated with the response particle ‘Yes’, while negative ones were associated with the particle ‘No’. In this sense, they are closer to “is it more than *n*?” than “is it at most *n*?”.

The emergency department of a hospital is required to have ten physicians present at all times. Following a complaint by a patient who had to wait for several hours, police officers want to know if this requirement was satisfied.

Investigator: "Were there enough physicians on Tuesday last week?"
Witness: "No, there were fewer than ten physicians."

Would you conclude that the witness knows exactly how many physicians were present?

Definitely not 1 2 3 4 5 Definitely yes

Fig. 5 Example of a target item in Experiment 2 with an *upward* context, *polar* QUD, and a *comparative-relevant* construction (i.e. *fewer than*)

2.2.3 Methodology

Participants: 95 participants were recruited on MTurk for the experiment. The participants were self-reported adult English speakers (age: 20–63, mean: 35). They received \$1.80 for their participation in the experiment.

Materials and procedure: Every participant read the following instructions on the first screen:

In this survey, you will see short dialogues between police officers and witnesses in some legal cases. Each example will come with a few sentences giving the context of this discussion. The witnesses are neither suspects nor plaintiffs, so they have a neutral position in the cases. They have no reason to hide information from the investigators, and are therefore being as cooperative as they can.

Each experimental item consisted of a context story, a question-answer pair, and a prompt. An example is presented in Fig. 5. We manipulated the following factors: QUD (*Polar* or *how many*), QUANTIFIERTYPE (*superlative*, *comparative-relevant*, *comparative-over-informative*, and *bare numeral*), and CONTEXTTYPE (*upward* or *downward*). All factors were within subject. All factors but CONTEXTTYPE were within item.

The QUD factor determined whether the investigator's question was a polar or a *how many* question. Each context story came in two versions: one that mentioned an explicit threshold, for the polar QUD, and one which did not specify any threshold, for the *how many* QUD. Responses to polar questions always involved a response particle ('yes' or 'no' depending on the case). The concrete quantifier used in the witness's response depended on both QUANTIFIERTYPE and CONTEXTTYPE, as explained above: in upward contexts, the superlative modifier was *at least*, the relevant comparative was *fewer than*, and the over-informative comparative was *more than*. In downward contexts, the superlative modifier was *at most* and the roles of *fewer than* and *more than* were reversed. Note that this contrast between over-informative and relevant answers only makes sense for polar questions, but to keep the comparison minimal, we kept the same quantifiers in *how many* QUDs. The context fixed the numeral used in all conditions. It was always a round number, ranging from 5 to 1000, and it was written in words.

In all target items, the prompt was a question of the form “Would you conclude that the witness knows exactly how many ...?”. Participants responded using a 5-point Likert scale from “Definitely not” to “Definitely yes”.

We designed 24 contexts (12 upward, 12 downward). The 8 possible combinations of QUD and QUANTIFIERTYPE were presented three times to each participant following a latin-square design. This means that some participants would see some combinations of quantifiers and QUD as upward twice and downward once, while other participants would see the same combinations as downward twice and upward once. The order was fully randomized across participants.

In addition to the 24 targets, each participant saw the same 10 fillers (5 true, 5 false).

Pre-processing and data analysis: Pre-processing was done in the same way as for Experiment 1. We decided to ignore three fillers when computing error rates because they tested participants’ attention in a convoluted way and were answered incorrectly in 90% of cases on average. Eleven participants were removed because their error rate on the rest of filler items was at least one standard deviation above the mean error rate (threshold: 31.8%). The mean error rate on remaining participants was 6.1%. The analyses followed the methods of the first experiment, with the only difference that the used mixed-effects models also had items random effects (in addition to subjects random effects).

2.2.4 Results

The results are presented in Fig. 6. The higher the response, the more exact knowledge (i.e. the less ignorance) participants attributed to the witness.

For the statistical analysis, we defined a POLARITY factor: *at least* and *more than* are positive quantifiers, whereas *at most* and *fewer than* are negative quantifiers. We first ran a model on responses to all target items with the following predictors: QUANTIFIERTYPE (comparative vs. superlative), POLARITY (centered, with negative at -0.5 , and positive at $+0.5$) and QUD (centered, with *HowMany* at -0.5 , *Polar* at $+0.5$), all interactions between the previous three predictors, as well as a NUMERAL predictor (scaled log of the numeral used in the witness’s response). The results, given in Table 4, showed significant main effects of POLARITY (positive quantifiers give rise to less ignorance) and QUD (*Polar* QUDs give rise to less ignorance). These two effects interacted positively, suggesting that the effect of QUD was stronger for positive quantifiers. NUMERAL was also significant (higher numerals give rise to more ignorance). No other effect was significant, in particular, none of the interactions associated with QUANTIFIERTYPE.

We fitted a second model focused on data from comparative modifiers and polar QUD to test for effects of relevance. This model confirmed the results of the first model (clear main effects of POLARITY), but showed no effect of RELEVANCE ($t = 0.23$, $p = .81$) and no interaction between POLARITY and RELEVANCE ($t = -1.3$, $p = .21$: if anything *more than* gives rise to more ignorance when providing a relevant answer, which is the opposite of what one would expect in light of the work of Cummins et al. 2012 discussed above).

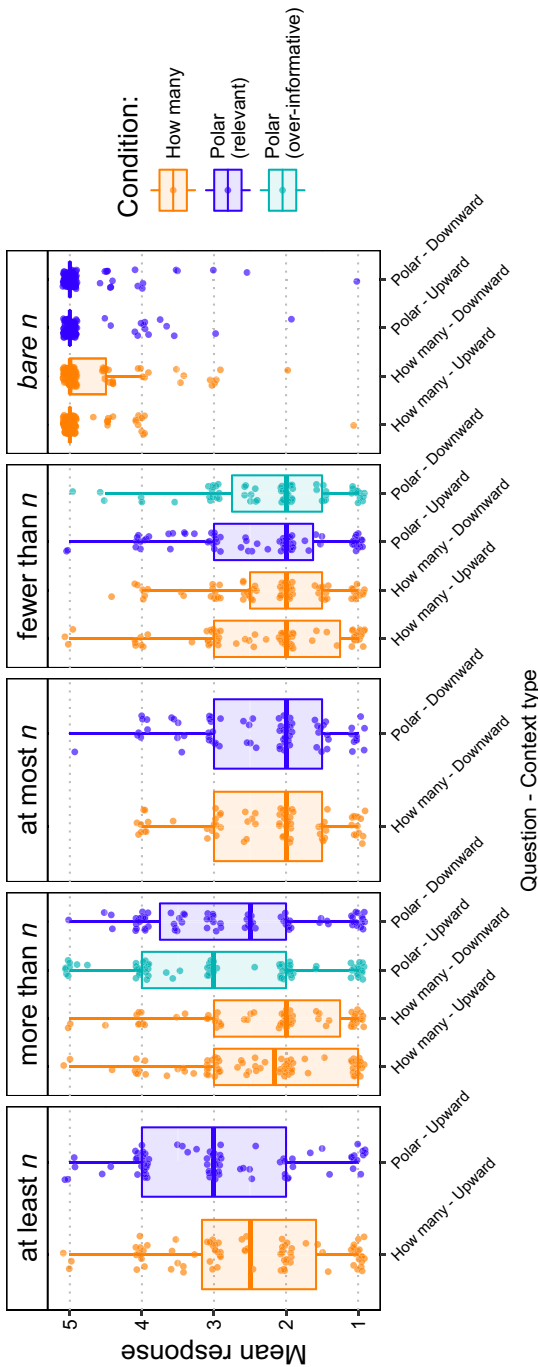


Fig. 6 Mean strength of the “exact knowledge” inference in each target condition for individual participants in Experiment 2. The relevance of the answer provided by a comparative modified numeral to the polar question, marked by the blue-green contrast here, depended on an interaction between context type (upward or downward) and polarity (negative *fewer than* or positive *more than*)

Table 4 Estimates for the fixed effects of the model fitted on target conditions for Experiment 2. For QUANTIFIERTYPE, *comparative-relevant* and *comparative-over-informative* were analyzed together and used as the baseline

| | β | t -value | p -value |
|---|---------|------------|------------|
| (Intercept) | -0.59 | -17.16 | < .001*** |
| QUANTIFIERTYPE[<i>Superlative</i>] | 0.06 | 1.20 | .23 |
| POLARITY | 0.24 | 3.85 | < .001*** |
| QUD | 0.18 | 4.03 | < .001*** |
| NUMERAL | -0.06 | -2.87 | .004** |
| QUANTIFIERTYPE[<i>Superlative</i>]:POLARITY | 0.09 | 1.09 | .28 |
| QUANTIFIERTYPE[<i>Superlative</i>]:QUD | 0.05 | 0.66 | .51 |
| POLARITY:QUD | 0.24 | 2.75 | .006** |
| QUANTIFIERTYPE[<i>Superlative</i>]:POLARITY:QUD | -0.04 | -0.28 | .78 |

We also fitted a model on the results involving bare numerals. This showed a strong effect of NUMERAL ($t = -2.9$, $p = .003$: higher numerals give rise to more ignorance), but no effect of QUD, CONTEXTTYPE or interaction (all $|t| < 1$, $p > .33$).

2.2.5 Discussion

In line with Coppock and Brochhagen's (2013a) finding that negative quantifiers are judged false more often than positive ones when the speaker has precise knowledge, we found that participants took the speaker to be less knowledgeable when she used negative quantifiers than when she used positive quantifiers.

We did not observe any effect of relevance on responses to polar QUDs. Recall that in Experiment 1, we observed that for *more than*, *PolarOverInf* was rated higher than *PolarRelevant*. This, we conjectured, was likely not due to the difference in relevance, but rather to the fact that *PolarOverInf* required mentioning a number that was salient in that context (*more than six*), unlike *PolarRelevant* which used the numeral *five*. The null effect of relevance in this experiment suggests that the contrast observed in Experiment 1 was indeed caused by the salience of *six* compared to *five* (recall that having six clubs was crucial to win the game).

We also confirmed that ignorance inferences are QUD-sensitive, as we had observed in Experiment 1. However, unlike in Experiment 1, we observed no difference between superlative and comparative modified numerals regarding the strength of ignorance inferences. This is in line with the results of W&B, and it raises an important question: what could explain the difference between our first experiment on the one hand, and W&B and our second experiment on the other hand? The third experiment addressed this issue.

2.3 Experiment 3

2.3.1 Goals

The main goal of this experiment was to gain a better understanding of the apparent conflict between the results of the first two experiments. We hypothesized that, among the various factors we identified in Sect. 2.1.4, the difference in the task was the most

A circuit breaker is designed to support up to ten outlets. Police officers are investigating how many outlets were connected to a circuit breaker that malfunctioned and caused a fire. The witness knows that between 11 and 13 outlets were connected to the circuit breaker.

Investigator: “Were there too many outlets connected to this circuit breaker?”
Witness: “Yes, there were more than ten outlets connected to the circuit breaker.”

Is the witness's answer appropriate?

Definitely not 1 2 3 4 5 Definitely yes

Fig. 7 Example of a target *Approximate* KNOWLEDGE item in Experiment 3 with a *downward* context, *polar* QUD, and *more than* (i.e., the *comparative-relevant* QUANTIFIERTYPE, since *more than* exactly matches the positive answer to the QUD)

likely explanation. In Experiment 1, participants had to judge the acceptability of an utterance, while they were informed about the knowledge state of the speaker. In Experiment 2, participants had to decide whether or not to draw an ignorance inference, based on the context and the given utterance. In order to directly assess the hypothesis that the difference in task was responsible for the observed contrasts between the results of the two experiments, we adapted the task of Experiment 2, making it very similar to the one of Experiment 1. At the same time, the materials of Experiment 2 were preserved as much as possible. How this was done is described in more detail below.

2.3.2 Methods

Participants: 96 participants were recruited on MTurk for the experiment. The participants were self-reported adult English speakers (age: 20–67, mean: 35). They received \$1.50 for their participation in the experiment.

Materials and procedure: The experiment mainly differed from Experiment 2 in the following two respects. First, we added information to each item on what the witness actually knew. Second, we changed the prompt from “Would you conclude that the witness knows...” to “Is the witness’s answer appropriate?”. An example is given in Fig. 7.

We added a two-level factor KNOWLEDGE determining how much the witness knew about each situation. In the *Precise* condition, the witness knew exactly what the number of relevant items was. This would be in conflict with potential ignorance inferences triggered by her answer. In the *Approximate* condition, the witness only knew a range of possible values that was chosen to be maximally compatible with a potential ignorance inference.⁸ The range of possible values was determined by the rules in Table 5 and the upper and lower bound of the range were written in numbers, as exemplified in Fig. 7.

⁸ In addition to being compatible with potential ignorance inferences, the range of possible values in the *Approximate* condition respected the exhaustivity inference discussed in Cummins et al. (2012).

Table 5 Description of the witness’s knowledge in each condition for each quantifier, where n is the numeral used by the witness and δ is a lower level of granularity (e.g., $n = 10$ and $\delta = 1$, or $n = 500$ and $\delta = 25$)

| | ‘at least n ’ | ‘more than n ’ | ‘at most n ’ | ‘fewer than n ’ |
|-----------------------|-------------------------------|--|-------------------------------|--|
| Precise knowledge | $n + \delta$ | $n + \delta$ | $n - \delta$ | $n - \delta$ |
| Approximate knowledge | between n and $n + 3\delta$ | between $n + \delta$ and $n + 3\delta$ | between $n - 3\delta$ and n | between $n - 3\delta$ and $n - \delta$ |

The addition of the KNOWLEDGE manipulation would have doubled the number of items in comparison with Experiment 2. We dropped certain conditions to compensate for this: since we did not observe any difference between relevant and over-informative answers in Experiment 2, we kept only relevant answers this time. This means that in the *Polar QUD* cases, *more than* was not tested in upward contexts, and *fewer than* was not tested in downward contexts.

The modifiers and the QUDs were balanced across contexts, but their combinations were not: *at least* and *fewer than* appeared with the *how many* QUD of downward contexts, and *at most* and *more than* appeared with the *how many* QUD of upward contexts. This was because, as explained in the method section of the previous experiment (Sect. 2.2.2), in case of *Polar QUDs* not every context could appear with every modifier. That is, only upward contexts could be combined with *at least* and *fewer than* and downward contexts with *at most* and *more than*. Since *how many* questions do not pose such a restriction, they were used to balance the design. This move was justified by a post hoc analysis of data from the previous experiment, showing that CONTEXT-TYPE did not exhibit any interaction with QUD or POLARITY ($\chi^2(3) = .90, p = .83$). Bare numerals appeared with both QUDs in each context.

We used a latin-square design with 24 context stories (the same stories as in Experiment 2). The experiment also included 2 training items (immediately after the instructions) and 12 fillers (8 were false controls, where the witness’s knowledge directly contradicted her statement; in the other 4 the witness had approximate knowledge and used the expression *about n* in her response).

Pre-processing and data analysis: Pre-processing and analysis were done in the same way as for Experiment 2. Ten participants were removed because their error rate on the filler items was at least one standard deviation above the mean error rate (threshold: 30.8%). The mean error rate for remaining participants was 2.7%.

2.3.3 Results

The results are presented in Fig. 8. We ran a model on responses to all target items with POLARITY, QUD, and QUANTIFIERTYPE encoded as in Experiment 2: QUANTIFIERTYPE was treatment-coded (comparative vs. superlative), POLARITY was centered (with negative at -0.5 , and positive at $+0.5$) and QUD was centered (with *how many* at -0.5 , and *Polar* at $+0.5$). KNOWLEDGE was treatment-coded with Approximate as the reference level. All interactions between the previous factors were present as fixed effects, as well as (scaled log) NUMERAL and the interaction between NUMERAL and

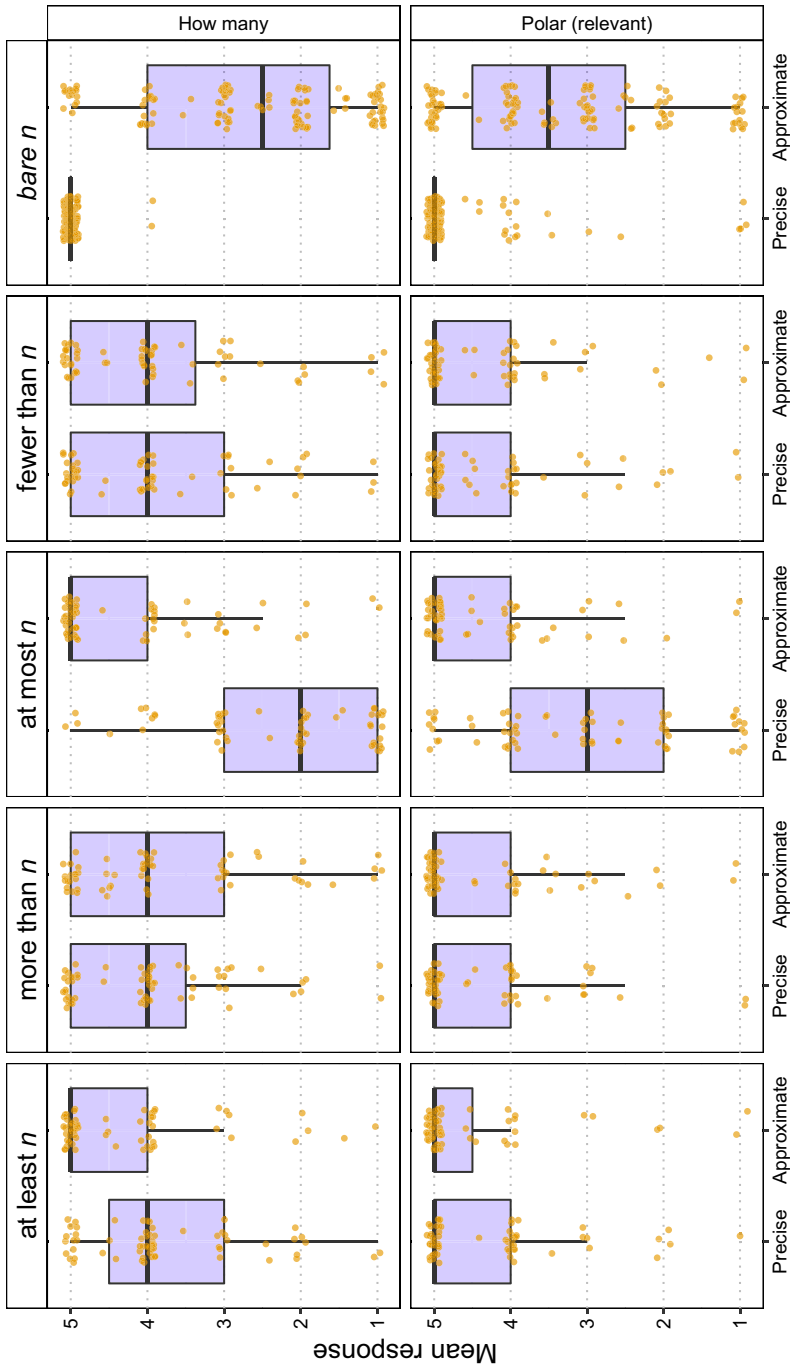


Fig. 8 Mean acceptability of the witness's answer in each target condition by individual participants in Experiment 3. Similar to Experiment 1, ignorance inferences manifest themselves as differences between the Precise and Approximate conditions

Table 6 Estimates for the fixed effects of the model fitted on target conditions for Experiment 3. QUANTIFIERTYPE and KNOWLEDGE were treatment coded with reference levels *Comparative* and *Approximate* respectively

| | β | t-value | p-value |
|--|---------|---------|-----------|
| (Intercept) | 0.56 | 13.51 | < .001*** |
| QUANTIFIERTYPE[Sup.] | 0.15 | 2.82 | .005** |
| POLARITY | -0.04 | -0.44 | .66 |
| KNOWLEDGE[Precise] | 0.02 | 0.47 | .64 |
| QUD | 0.27 | 2.46 | .014* |
| NUMERAL | -0.02 | -0.77 | .44 |
| QUANTIFIERTYPE[Sup.]:POLARITY | 0.14 | 1.37 | .17 |
| QUANTIFIERTYPE[Sup.]:KNOWLEDGE[Precise] | -0.59 | -7.88 | < .001*** |
| POLARITY:KNOWLEDGE[Precise] | 0.03 | 0.27 | .79 |
| QUANTIFIERTYPE[Sup.]:QUD | -0.14 | -0.95 | .34 |
| POLARITY:QUD | 0.21 | 1.45 | .15 |
| KNOWLEDGE[Precise]:QUD | -0.02 | -0.15 | .88 |
| KNOWLEDGE[Precise]:NUMERAL | 0.08 | 2.38 | .017* |
| QUANTIFIERTYPE[Sup.]:POLARITY:KNOWLEDGE[Precise] | 0.65 | 4.64 | < .001*** |
| QUANTIFIERTYPE[Sup.]:POLARITY:QUD | -0.10 | -0.41 | .68 |
| QUANTIFIERTYPE[Sup.]:KNOWLEDGE[Precise]:QUD | 0.27 | 1.55 | .12 |
| POLARITY:KNOWLEDGE[Precise]:QUD | -0.27 | -1.43 | .15 |
| QUANTIFIERTYPE[Sup.]:POLARITY:KNOWLEDGE[Precise]:QUD | 0.15 | 0.52 | .60 |

KNOWLEDGE. The results, given in Table 6, showed a significant effect of QUANTIFIERTYPE (superlative quantifiers are more acceptable in Approximate KNOWLEDGE situations than comparatives are), an effect of QUD (the acceptability of comparatives is higher with *Polar* QUDs), a strong interaction between KNOWLEDGE and QUANTIFIERTYPE (superlative quantifiers are clearly degraded in Precise KNOWLEDGE situations), and a triple interaction between KNOWLEDGE, QUANTIFIERTYPE and POLARITY (*at most* is more sensitive than *at least* to ignorance violations). Most notably, KNOWLEDGE had no effect whatsoever on comparative quantifiers, showing that these quantifiers do not convey ignorance in this experimental setup. We found no main effect of NUMERAL, but an interaction with KNOWLEDGE: when the speaker has precise knowledge, modified numerals involving high numerals are more acceptable than those involving low numerals, suggesting that higher numerals give rise to weaker ignorance inferences. In Experiment 2, we observed an effect in the opposite direction: higher numerals gave rise to stronger ignorance inferences.

Overall, these results are very close to what we found in Experiment 1: strong ignorance effects with superlative modifiers, and none with comparative modifiers. Interestingly, the three-way interaction between KNOWLEDGE, QUANTIFIERTYPE and QUD, which would have indicated QUD sensitivity for the ignorance inferences of superlative modifiers, did not reach significance ($p = .11$). While this might seem surprising at first sight, this finding also matches the results of Experiment 1. There, we observed a three-way interaction, indicating that the ignorance inferences of superlative modifiers are QUD-sensitive. However, this three-way interaction was only present in the *Exact* SITUATION, while in the *Exceed* SITUATION, it was of the same magnitude as in the current experiment and failed to reach significance. This is not surprising since

the *Precise* KNOWLEDGE condition in this experiment is more similar to the *Exceed* than to the *Exact* SITUATION in Experiment 1.

In line with Experiment 2, ignorance inferences affected the acceptability of the negative modifier *at most* more than that of the positive modifier *at least*.

Finally, we analyzed the results on bare numerals and replicated the findings of Experiment 1: we found a clear effect of KNOWLEDGE ($t = -11.5$, $p < .001$: bare numerals are degraded when the speaker doesn't have Precise knowledge), and an interaction between KNOWLEDGE and QUD ($t = -5.3$, $p < .001$: the effect of ignorance is weaker with a *Polar* QUD). No other effect was significant (all $|t| < 1.3$, $p > .22$), and in particular, we found no effect of context type, and no effect of numeral size.

2.4 Experiment 4 (replication of Experiment 2)

2.4.1 Goal

One possible shortcoming of Experiment 2 is the prompt we used. We felt it was more natural to ask “Would you conclude that the witness knows exactly how many items there were?”, instead of the negative question “Would you conclude that the witness doesn't know exactly how many items there were?”. Intuitively, a “No” response to the first question can convey that the witness is in fact ignorant, but it may also convey the weaker conclusion that the participant isn't in a position to conclude that the witness has exact knowledge, but does not exclude that this could be the case. In short, our method may not allow us to distinguish between utterances that convey ignorance and utterances which are compatible with ignorance but do not convey it. This is important because this distinction is sometimes assumed to be the distinguishing factor between superlative and comparative modified numerals, and only in this experiment did we find the same pattern for the two types of modifiers.

We therefore decided to replicate Experiment 2 using a prompt that would unambiguously target ignorance, and not just compatibility with ignorance.

2.4.2 Methodology

Participants: 95 participants were recruited on MTurk for the experiment. The participants were self-reported adult English speakers (age: 19–65, mean: 33). They received \$1.80 for their participation in the experiment.

Materials and Procedure: The design was strictly identical to that of Experiment 2, except that the prompt in (9a) was replaced with (9b), which was unambiguous and more natural than the plain negation of (9a).

- (9) a. Would you conclude that the witness knows exactly how many...?
 b. Would you conclude that the witness is unaware precisely how many...?

The rating for utterances that are compatible with both ignorance and knowledgeability should be more or less unaffected, as they shouldn't convey ignorance much more than they conveyed knowledgeability. By contrast, the ratings for sentences which do

convey either ignorance or knowledgeability should be reversed (they should go from low to high, and from high to low, respectively).

Pre-processing and data analysis: Pre-processing and analysis were done in the same way as for Experiment 2 and 3. Eleven participants were removed because their error rate on the filler items was at least one standard deviation above the mean error rate (threshold: 32%). The mean error rate on remaining participants was 4.4%.

2.4.3 Results

The results are presented in Fig. 9. The higher the response, the *more* ignorance participants attributed to the witness (opposite of Experiment 2).

We ran the same statistical analysis as for Experiment 2. We first ran a model on responses to all target items with the following predictors: QUANTIFIERTYPE (comparative vs. superlative), POLARITY (centered, with negative at -0.5 , and positive at $+0.5$) and QUD (centered, with *HowMany* at -0.5 , *Polar* at $+0.5$), all interactions between the previous three predictors, as well as a NUMERAL predictor (scaled log numeral). The results, given in Table 7, showed significant main effects of NUMERAL (higher numerals tend to give rise to more ignorance), POLARITY (positive quantifiers give rise to less ignorance), and QUD (*Polar* QUDs give rise to less ignorance), as in Experiment 2. Surprisingly however, the interaction between POLARITY and QUD was completely absent this time. No other interaction was significant either.

We fitted a second model focused on data from comparative modifiers and polar QUD to test for effects of relevance. This model didn't show any significant effect. Unlike in Experiment 2, the main effects of POLARITY didn't exactly reach significance ($t = -1.96$, $p = .0501$). Most importantly, we again found no significant effect of RELEVANCE ($t = -1.44$, $p = .15$) and no interaction ($t = 0.72$, $p = .47$).

We also fitted a model on the results involving bare numerals, with results similar to Experiment 2: a strong effect of NUMERAL ($t = 3.0$, $p = .002$; more ignorance with higher numerals), but no effect of QUD or CONTEXTTYPE (all $|t| < 1.4$, $p > .17$).

2.5 General discussion of experimental findings

Figure 10 presents a condensed summary of the most important results obtained in the four experiments. The graphs highlight the main findings: in both Experiments 1 and 3, the acceptability of superlative modifiers was affected by the speaker's knowledge level, unlike that of comparative modifiers. Superlative modifiers therefore give rise to ignorance implicatures more robustly than comparative modifiers. In all experiments, moreover, we observed that whenever an effect of ignorance was present, it was sensitive to the QUD.

This empirical picture is complicated by several factors, however. First, ignorance effects were not always stronger for superlative modifiers. In Experiments 2 and 4, in particular, we observed no difference between the two types of modifiers. The results of Experiment 3 provide an indication as to how to reconcile the conflicting results obtained in the previous two experiments. Altering the setup of Experiment 2 only in the task that participants performed (judging acceptability rather than inferences), we

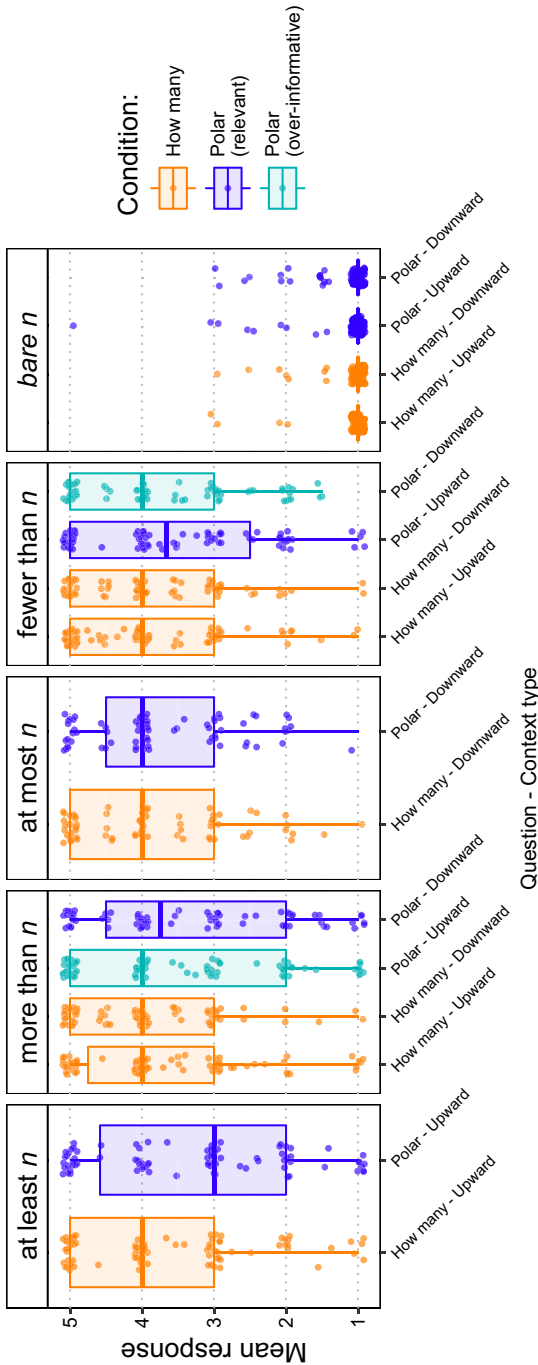


Fig. 9 Mean strength of the ignorance inference in each target condition for individual participants in Experiment 4. As in Experiment 2, the relevance of the answer provided by a comparative modified numeral to the polar question, marked by the blue-green contrast here, depended on an interaction between context type (upward or downward) and polarity (negative *fewer than* or positive *more than*)

Table 7 Estimates for the fixed effects of the model fitted on target conditions for Experiment 4

| | β | t -value | p -value |
|---|---------|------------|------------|
| (Intercept) | 0.59 | -21.18 | < .001*** |
| QUANTIFIERTYPE[<i>Superlative</i>] | -0.04 | -0.99 | .32 |
| Polarity | -0.13 | -2.70 | .007** |
| QUD | -0.26 | -5.36 | < .001*** |
| Numeral | 0.03 | 2.29 | .022* |
| QUANTIFIERTYPE[<i>Superlative</i>]:Polarity | -0.09 | -1.29 | .20 |
| QUANTIFIERTYPE[<i>Superlative</i>]:QUD | 0.09 | 1.23 | .22 |
| Polarity:QUD | .001 | 0.01 | .99 |
| QUANTIFIERTYPE[<i>Superlative</i>]:Polarity:QUD | -0.05 | -0.34 | .74 |

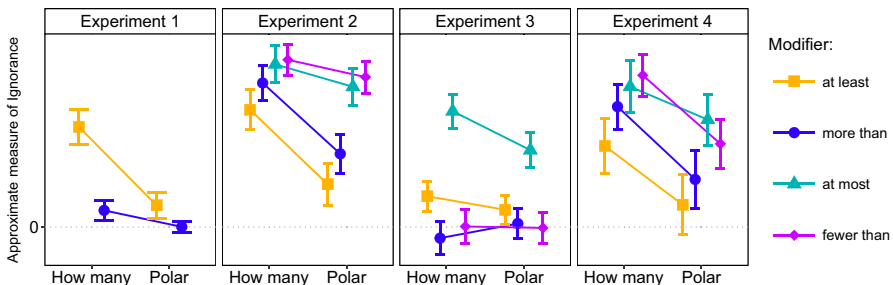


Fig. 10 Mean and SE for a derived measure of Ignorance in each of the four experiments. 0 corresponds to a no-ignorance baseline, and higher values mean stronger ignorance. For Experiments 1 and 3, we computed the mean difference between conditions that respected and violated ignorance. For Experiment 2 and 4, we computed the difference between the mean answer and an estimated middle-of-the-scale point corresponding to no ignorance (this point was estimated from a regression of the data on Experiment 2 and 4 to be around 3.2). Relevant and over-informative answers to polar questions were pooled together since we observed no difference between them

obtained results very similar to those of Experiment 1. We thus conclude that the source of the contrast observed between the results of Experiments 1 and 2, respectively, indeed lies in the difference between the two tasks: (i) when participants judge the appropriateness of an expression while having access to the speaker’s knowledge, they only take superlative modifiers to convey ignorance; (ii) when they have to draw inferences about the speaker’s knowledge given what the speaker said in a certain context, they take both superlative and comparative modifiers to convey ignorance, which is modulated by QUDs.

Second, we observed a systematic contrast between positive and negative quantifiers in Experiments 2 to 4 (negative quantifiers were not tested in Experiment 1). Similar contrasts are reported in Geurts and van der Slik (2005), Geurts et al. (2010), Coppock and Brochhagen (2013a). Having established that the contrast between positive and negative quantifiers is orthogonal to the task effect discussed above, we leave an explanation of the former for future research and focus on the latter.

In a series of experiments on scalar implicatures, Degen and Goodman (2014) observed a similar difference between inference and truth-value judgment tasks: the truth-value judgment task was sensitive to contextual factors that had no effect on the inference task. Our results are in line with their findings, and we follow their

explanation of the difference in assuming that the two tasks make participants take different perspectives in communication. In truth-value or acceptability judgment tasks (such as Experiment 1 and 3), participants are taking the speaker's perspective, as they have access to her internal mental state and are asked to judge the appropriateness of different utterances. In this sense, these tasks are relatively close to a production task. Degen and Goodman (2014) indeed show that a standard production experiment (word probability rating) gives results similar to the truth-value judgment task. In inference tasks (Experiments 2 and 4, as well as W&B's experiment), participants are more likely to take the hearer's perspective, as they have to draw inferences about the speaker's mental state based on what was said. This is a comprehension task proper. Our results therefore suggest that when taking the perspective of speakers, participants are sensitive to the comparative/superlative distinction with respect to ignorance inferences, but this distinction does not play an important role when they take the hearer's perspective. In Sect. 3 we develop an account that derives this connection.

3 An optimality theoretic account

3.1 Motivating an optimality theoretic account

We have seen that the extent to which experimental participants take modified numerals to convey ignorance depends on the task that they are asked to perform. Following a suggestion by Degen and Goodman (2014), who found a similar task-dependency in a different empirical domain, we take our findings to reflect an asymmetry between production and comprehension tasks. Degen and Goodman explain this difference by assuming that participants behave as rational Bayesian listeners in comprehension tasks, and that this results in a more noisy dependent variable (because uncertainties on the speaker model and on the prior add up), which gives rise to null effects. However, the results of our Experiments 2 and 4 appear to be qualitatively different from those of Experiments 1 and 3, not just noisier. For this reason and others detailed below, we will depart from the assumption that participants in a comprehension task are fully rational, and pursue a different model of comprehension tasks inspired by work in language acquisition.

While Degen and Goodman (2014) are the only ones, to our knowledge, to have empirically demonstrated production/comprehension asymmetries with adults, the literature on language acquisition has uncovered many such asymmetries in children (see, e.g., Hendriks and Koster 2010; Hendriks 2014). Of particular interest for us is a well-studied example in pronoun resolution. Unlike adults, who only allow a disjoint reference reading for sentences like (10), children up to the age of 6 often allow the coreference reading as well (Chien and Wexler 1990, among others). However, in production these children start behaving like adults much earlier on: when coreference is intended they use a reflexive pronoun (*herself*) and when disjoint reference is intended they use a non-reflexive pronoun (*her*). This has been found both in corpus studies and in elicited experimental data (Bloom et al. 1994; De Villiers et al. 2006, among others).

(10) This is Mama Bear; this is Goldilocks. Mama Bear_{*i*} is washing her_{*j*≠*i*}.

Hendriks and Spenader (2006) propose an account of this asymmetry couched in Optimality Theory (OT). In OT, a grammar is seen as an ordered set of constraints on possible form/meaning pairs. In comprehension, such a grammar selects the optimal interpretation(s) for a given expression, while in production it selects the optimal expression(s) to convey a given intended interpretation. Which interpretations/expressions are considered optimal depends on the ‘mode’ of optimization. We will first consider so-called *unidirectional optimization*, since it can account for production/comprehension asymmetries.

In production, a speaker who wants to express a certain meaning considers the candidate forms which could express it and selects the one which is optimal w.r.t. the constraints in the grammar (taking the ordering of these constraints into account). In comprehension, the listener considers candidate meanings for the form they heard and, again, selects the one which is optimal w.r.t. the constraints in the grammar. Unidirectional optimization in comprehension does not look at competing forms but only at competing meanings. Because of this, it occasionally results in production/comprehension asymmetries. Namely, it may be the case that the meaning *m* that is optimal for a given form *f* would be optimally expressed by a different form *f'*. In this case, a listener would interpret *f* as *m*, but a speaker would not use *f* to express *m*.

Hendriks and Spenader (2006) assume that children’s non-standard comprehension of pronouns results from unidirectional optimization, while adults perform so-called *bidirectional optimization*. Bidirectional optimization (Blutner 2000) requires looking at both competing forms and meanings in production and in comprehension, eliminating potential asymmetries. It has been argued that this is more taxing than unidirectional optimization on executive functions, in particular working memory (Hendriks 2014). Bidirectional optimization also requires extra time compared to unidirectional optimization during actual communication (van Rij et al. 2010). The hypothesis is that the limited working memory capacity of children does not facilitate bidirectional optimization. Children therefore resort to unidirectional optimization, which results in the observed production/comprehension asymmetries.

We will pursue an OT analysis of modified numerals in order to explain the task effect found in our experiments, which, as discussed above, plausibly reflects an asymmetry between production and comprehension comparable to the asymmetries found in child language. Of course, such an approach is only worth pursuing if it can be argued that it is plausible to assume that not only children, but adults too, may resort to unidirectional optimization in some cases, in particular when interpreting modified numerals in a psycholinguistic task.

We believe that there are multiple reasons why this is indeed a plausible assumption to make. First, van Rij et al. (2010, 2013) argue that bidirectional optimization involves ‘proceduralization’: when it is practiced enough, the best expression-interpretation pair can be selected in one step, without the need to go through two selection procedures. This, however, requires practice. Pronouns are extremely frequent, so it is to be expected that adults manage to perform bidirectional optimization in this case, unlike children who have not yet been exposed to them enough. Given the much lower

frequency of modified numerals compared to pronouns,⁹ it is likely that in the case of modified numerals bidirectional optimization has not been proceduralized to the extent that it has in the case of pronouns. Indeed, it has been found that deriving ignorance inferences of modified numerals comes with a significant processing cost in psycholinguistic tasks (Alexandropoulou et al. 2016, 2017; Alexandropoulou 2018), which would be unexpected if the process of identifying their optimal interpretation had been fully proceduralized.

A second reason why adults may resort to unidirectional optimization when interpreting modified numerals concerns the range of alternative candidates that have to be considered in the selection procedure. Even though OT in principle does not restrict the number of candidates, it seems plausible that language users only consider a limited range of alternatives. In the case of pronoun resolution, there are not many alternative expressions to be compared: the pronoun itself, a reflexive form and perhaps a definite description. On the other hand, in the case of modified numerals, there are many more alternatives to be considered. For instance, *at least eight* is plausibly compared to expressions which involve a different base numeral (e.g., *at least seven*, *at least nine*,...), expressions that involve a different modifier (e.g., *more than eight*, *more than nine*,...), and numeral expressions that do not involve a modifier at all (e.g., *eight*, *nine*,...). Reinhart (2006, §2.7) already hypothesized that adults may fail to correctly apply procedures such as the one necessary for pronoun resolution if there are many candidates to be compared. Bidirectional optimization in particular would be much more taxing in the case of modified numerals than in the case of pronoun resolution. Note that the same considerations apply more generally to any comprehension mechanism which relies on the comparison of alternative expressions (van Rooij et al. 2011), including the kind of Bayesian reasoning at play in Rational Speech-Act models (Frank and Goodman 2012).

Finally, we know that in psycholinguistic tasks adults often fail to derive implicatures, precisely because they ignore lexical alternatives which are too costly to retrieve (Van Tiel and Schaeken 2017, and references therein). This effect becomes even clearer when participants' working memory resources are restricted (De Neys and Schaeken 2007; Marty and Chemla 2013; Marty et al. 2013). Unidirectional OT is one of several ways to formalize pragmatic reasoning when ignoring lexical alternatives, but it comes with some advantages in comparison with models which simply take 'literal participants' to not derive any implicatures at all (e.g., the literal listener in RSA models). In particular, unidirectional OT will predict that some ignorance inferences survive even when alternative expressions are not considered, which will turn out to play an important role in modelling the participants' behavior in Experiments 2 and 4, and it offers an alternative explanation of the task effect reported in Degen and Goodman (2014).¹⁰ In this sense it may represent a plausible heuristic that listeners may rely on in situations where full-fledged pragmatic reasoning is too costly.

⁹ In the Corpus of Contemporary American English (CoCA; Davies 2008), *at least* (in any construction, not just modifying numerals) occurs around 150.000 times. Pronouns *him/her/she/he* occur around 10 million times.

¹⁰ In short, we would predict an effect of QUD for production as *some* violates QUANT in the \forall -world with an *all?* QUD but only SQUAL with an *any?* QUD. By contrast, our model predicts no effect of QUD in comprehension since *some* always violates higher constraints in the \forall -world than in the \exists -world.

While these considerations justify, in our view, the pursuit of a unidirectional OT account of our experimental findings, it is beyond the scope of this paper to provide a definitive, general answer to the question whether production/comprehension asymmetries are encountered in adult language ‘in the wild’. Outside psycholinguistic tasks, we suspect that listeners typically do consider alternative expressions and derive the full range of implicatures available, minimizing asymmetries with how speakers use modified numerals. Whether they achieve this with bidirectional OT or Bayesian reasoning is an open question. In any case, the unidirectional comprehension OT account explains the surprising contrast between, on the one hand, the robust introspective judgments reported in the literature that superlative modified numerals always give rise to ignorance inferences, and on the other hand, our experimental results (as well as Westera and Brasoveanu’s) which show that participants sometimes don’t derive these inferences in demanding experimental settings.

Readers who are not particularly interested in the psycholinguistic debate regarding inference tasks may want to skip discussion of our “listener” model (Sect. 3.3.4 and Sect. 3.4.3 specifically), as it really is a model of participants’ behavior in true comprehension tasks only. We take the “speaker” model, on the other hand, to be applicable to natural communication more generally.

3.2 Semantic assumptions

3.2.1 Bare numerals

We assume that a bare numeral n is semantically ambiguous between a one-sided reading (n or more) and a two-sided reading (exactly n). Relevant proposals include that of Geurts (2006) and Kennedy (2015), who take the two-sided reading to be basic and derive the one-sided reading through type-shifting operations, as well as the grammatical implicature account (Chierchia et al. 2012, a.o.), which takes the one-sided reading to be basic and derives the two-sided reading by application of a grammatical exhaustivity operator. For our purposes, the differences between these two ambiguity approaches (and others) do not matter, since we are only concerned here with unembedded uses of bare numerals describing cardinalities (see Spector 2013 for further discussion).¹¹

Our proposal is not compatible with approaches that take numerals to be semantically unambiguous and derive the missing interpretation through pragmatics. These include the neo-Gricean account (Horn 1972; Schulz and van Rooij 2006), which assigns bare numerals a one-sided semantic denotation only, and Breheny’s (2008)

¹¹ As pointed out by an anonymous reviewer, however, if we were to apply our proposed OT system beyond numerical expressions, we would need to postulate an exhaustivity operator. Indeed, in (i) the speaker can simply say “Mary” to mean “Mary and no one else” but nothing in the pragmatics we propose in Sect. 3.3 derives exhaustivity:

- (i) a. Who called?
- b. Mary.

In this sense we are following Fox’s (2007) approach to implicatures, where exhaustivity and upper-bounding effects are derived in the grammar, while ignorance implicatures are derived by pragmatics.

account, which assigns them a two-sided semantic denotation only. Besides previously identified shortcomings (Geurts 2006; Breheny 2008; Spector 2013; Kennedy 2015), the former would not allow us to explain the difference between bare numerals and superlative modified numerals (they would be semantically equivalent). The latter would allow us to account for our experimental results concerning modified numerals, but would be incompatible with the asymmetry observed in Experiment 1 for bare numerals. Namely, on this account *Six of my cards are clubs* is false not only when the speaker has 4 or 5 clubs, but also when she has 7 or 8 clubs. We found, however, that the sentence is much more degraded in the former situation (i.e., when it is false both on a one-sided and on a two-sided reading) than in the latter (i.e., when it is false only on a two-sided reading). Furthermore, we found that judgments in the latter, but not in the former case, were sensitive to the QUD, which would be unexpected under a purely semantic account.

3.2.2 Modified numerals

We assume that modified numerals are unambiguous and receive their traditional naive interpretation:

(11) “At least three students left” is true if and only if $|\llbracket \text{student} \rrbracket \cap \llbracket \text{left} \rrbracket| \geq 3$.

(12) “More than three students left” is true if and only if $|\llbracket \text{student} \rrbracket \cap \llbracket \text{left} \rrbracket| > 3$.

Note that under an exhaustification account of the ambiguity of bare numerals, something must be said as to why modified numerals are *not* ambiguous between the proposed denotation and a two-sided reading. Such accounts usually assume that exhaustification is vacuous for modified numerals (Fox and Hackl 2007, among others) though see Enguehard (2018) for an alternative view.

3.3 Pragmatic assumptions

As discussed above, we will model pragmatic behavior using Optimality Theory (OT). More specifically, we will build on Cummins’ (2011, 2013) OT account of modified numerals.

Cummins’ main idea is that the distribution of modified numerals results from a trade-off between three factors: complexity (modified numerals are more complex expressions than bare numerals), salience of the base numeral (round numerals and numerals that have been primed by previous context are easier to process), and informativity (bare numerals can convey exact quantities, so they are usually more informative). The net result is that modified numerals are used in two kinds of situations: when the speaker doesn’t have enough information to use a bare numeral, or when she has precise knowledge of a specific non-round number, but decides that using a round or salient number is more important than conveying the exact quantity.

We will now present each of the constraints we consider, and the restrictions they impose on a tuple $\langle \varphi, s, Q \rangle$ where φ is an expression (utterance), s is the speaker’s information state (a set of worlds), and Q the Question under Discussion (QUD), modeled here for simplicity as a partition on a set of worlds representing the common

ground. An expression φ may have several interpretations $\varphi_1, \varphi_2 \dots$ (due to different possible syntactic parses and/or optional type-shifting or exhaustivity operators). Given our semantic assumptions in Sect. 3.2, the only expressions which we take to be ambiguous are the bare numerals. We write $\llbracket \varphi_i \rrbracket$ for the denotation of φ_i , which we take to be a set of worlds.

3.3.1 Quality

We assume the usual Gricean maxim of Quality: if a speaker utters φ , her information state s should support φ under at least one interpretation:

- (13) Maxim of Quality (QUAL):
 $\langle \varphi, s, Q \rangle$ violates QUAL if for all interpretations φ_i of φ : $s \not\subseteq \llbracket \varphi_i \rrbracket$.

Note that Quality phrased this way is rather lenient since it only rules out the use of an ambiguous expression when *none* of its interpretations are supported by the speaker's information state. We propose a second constraint to further restrict the use of ambiguous expressions. If a speaker uses an ambiguous expression, she should make sure that it is true under all its different interpretations (even the ones she did not intend to convey).¹² This can be seen as a more stringent version of Quality (it prevents the speaker from unintentionally conveying false information), or as part of the Maxim of Manner (be clear, avoid confusion).¹³

- (14) Strong Quality / Avoid misleading ambiguities (SQUAL):
 $\langle \varphi, s, Q \rangle$ violates SQUAL if for some interpretation φ_i of φ : $s \not\subseteq \llbracket \varphi_i \rrbracket$.

3.3.2 Quantity

The maxim of quantity requires that the speaker make her contribution “as informative as is required (for the current purposes of the exchange)” (Grice 1975). We will adopt an implementation of Grice's maxim which requires that the speaker conveys all the information relevant to the QUD that she has access to:

- (15) Maxim of Quantity (QUANT):
 QUANT is violated if s resolves the QUD better than φ .
 More formally, $\langle \varphi, s, Q \rangle$ violates QUANT if s overlaps with a proper subset of the cells of Q that any interpretation of φ supported by s overlaps with:

¹² An anonymous reviewer points out that the range of interpretations would need to be contextually restricted, otherwise pronouns would systematically violate this constraint.

¹³ Similar principles have been proposed in Spector (2017) and Križ and Spector (2020) to account for homogeneity effects with definite descriptions. The idea has roots in supervaluationism (an idea first proposed to account for vague expressions) and Krifka's (1996) Strongest Meaning Hypothesis. Indeed, requiring that one uses an ambiguous expression only when one has enough evidence to support all of its interpretations amounts to requiring enough evidence for the strongest interpretation if these interpretations are ordered by entailment. In a situation where there are alternative expressions equivalent to the weaker interpretations, the listener can then infer that the speaker's use of an ambiguous expression conveys the strongest interpretation.

$$\forall i : s \subseteq \llbracket \varphi_i \rrbracket \implies \{q \in Q \mid s \cap q \neq \emptyset\} \subset \{q \in Q \mid \llbracket \varphi_i \rrbracket \cap q \neq \emptyset\}$$

In short, (15) states that at least one interpretation of the utterance should convey all the information available to the speaker that is relevant to the QUD. If the expression is ambiguous between an interpretation that is not supported by the speaker's information state, and one that *is* supported by the speaker's information state but under-informative, we take it to be a violation of Quantity, as the speaker could not possibly intend to resolve the QUD with the unsupported interpretation.

Our QUANT constraint differs from Cummins' *Informativeness* constraint in several respects (see Cummins 2011, §2.4.1). First, we directly encode relevance in Quantity by measuring informativeness with respect to the QUD (as first proposed in Matsumoto 1995). Second, we do not need to distinguish between varying degrees of violation. This would however become necessary if we were to look at a broader range of expressions or if we adopted Fox and Hackl's (2007) Universal Density of Measurement hypothesis, as discussed in Sect. 3.4.4.

3.3.3 Manner

Grice's maxim of manner addresses a variety of issues with the way speakers articulate their thoughts. We already mentioned how our second Quality constraint against misleading ambiguities could be related to Manner.

The maxim of manner also requires speakers to aim for the simplest expression that satisfies their communicative purpose. We will model this by penalizing more complex expressions (both in term of production and processing cost for the listener). Following Cummins (2011), we take superlative modifiers to be more complex than comparative ones (see Geurts et al. 2010; Cummins and Katsos 2010 for experimental validation of this assumption), and bare numerals to be the simplest form (this corresponds to *quantifier simplicity* in Cummins 2011).

- (16) Simplicity (SIMP):
 $\langle \varphi, s, Q \rangle$ incurs one violation of SIMP if φ contains a comparative modifier, and two violations if it contains a superlative modifier.

While SIMP affects the choice of a modifier, the choice of a numeral involves similar considerations. More specifically, "round" numbers are typically easier to process (Dehaene 2011, see also Jansen and Pollmann 2001 for a formal definition of roundness). Similarly, numbers mentioned in the previous context (primed) are salient and therefore also easier to process. While Cummins (2011) distinguishes between the two sources of salience for numerals (roundness and priming), we will model both with a single constraint NSAL, since this will be sufficient to explain our experimental results. Unlike Cummins (2011), we do not need to distinguish between different levels of roundness.

- (17) Numeral salience (NSAL):
 $\langle \varphi, s, Q \rangle$ violates NSAL if φ contains a numeral that is neither round nor primed by the context.

Finally, we assume that speakers also obey a faithfulness constraint that favors the use of numerals that are internally salient (i.e., salient to the speaker, though not necessarily to the addressee). We assume in particular that if a speaker knows that the value of a given quantity lies within a certain range $[n \dots m]$, then the boundaries of this range, n and m , are internally salient. For instance, if a speaker makes a statement about how many students smoke, and if she knows that between 7 and 12 students smoke, then 7 and 12 are internally salient for her.^{14, 15}

(18) Internal salience (ISAL):

$\langle \varphi, s, Q \rangle$ violates ISAL if φ contains a numeral that does not correspond to a boundary of the speaker's knowledge about the quantity under discussion.

Thus, for instance, if as above a speaker knows that between 7 and 12 students smoke, then *at least 7 students smoke* and *at most 12 students smoke* satisfy ISAL but *more than 6 students smoke* and *less than 13 students smoke* do not.

We assume that ISAL plays a role in speakers' choices of numerical expressions, because, other things being equal, internally salient numerals are presumably easier to produce for speakers than non-salient ones. Additionally, ISAL may be seen as a particular instance of a more general pressure to align what is externally salient with what is internally salient.

Cummins (2011, §2.5.2) also considers the possibility of including a faithfulness constraint in his OT system which would favor the use of numerals that are internally salient to the speaker. "One approach," he writes, "would be to change the definition of 'primed' numerals and quantifiers to include those which are activated in the mind of the speaker as well as those which are present in the preceding discourse." Something like this would be needed, he points out, in order for his system to license the use of numerals that are not salient in any external sense (i.e., round or contextually primed). However, he does not explicitly introduce such a constraint, and does not specify in more detail what it would mean for a numeral to be 'activated in the mind of the speaker'. Since this is exactly what our ISAL constraint does, we view it as being very much in the spirit of Cummins' proposal, and one that is indeed necessary in any OT system of this kind in order to license the use of numerals that are not externally salient.

3.3.4 Linking theory and experimental results

With our constraints in place, we need to specify how they are ranked, and how the predictions of the OT system translate into predictions about the behavior of participants in each of the tasks in our experiments.

¹⁴ Although we cannot point to concrete experimental evidence for this assumption at this point, it seems plausible that range knowledge is often encoded in terms of the boundaries of the range. This would indeed make the boundaries more salient than other values in the range. In Sect. 3.5.1, we will extend the notion of internal salience to scales other than cardinalities and show that it immediately captures previous observations in this domain, providing further support for the proposed constraint.

¹⁵ Besides the boundaries of the range of values that a speaker considers possible, other numerals may be internally salient for the speaker as well. For our purposes here, however, boundaries are of particular relevance. The question of which other factors give rise to internal salience is left open.

We assume the order in (19).¹⁶

(19) QUAL \gg QUANT \gg SQUAL \gg ISAL \approx NSAL \gg SIMP

Note that the two salience constraints are not ranked. Following Boersma (1997), among others, we interpret ‘ \approx ’ in a probabilistic manner: if NSAL and ISAL are in conflict, they do not cancel each other but either of them can take precedence at evaluation time. The result is that ties between these two constraints are not broken by lower constraints (e.g., SIMP here), but both candidates are possibly optimal. This contrasts with classical OT, where two forms can never be optimal at the same time.

In Experiments 1 and 3, participants had access to the information state of the speaker s , and had to evaluate the acceptability of candidate responses to a question (which we identify with the QUD Q). We assume that an utterance which is optimal in the OT system receives maximum acceptability. For utterances which are not optimal, we assume that the acceptability depends on the rank of the fatal violation (the highest violation that distinguishes the utterance from the optimal candidate). The higher this violation ranks, the more degraded the utterance.¹⁷

In Experiments 2 and 4, participants were placed in the position of a listener who had access to Q and φ , and needed to infer whether the speaker had precise knowledge or not (i.e., whether s specified a single value or not). Such a task can be modeled in several ways. Bayesian models and bidirectional OT are meant to capture optimal rational behavior in such circumstances. However, it is impossible that the behaviour of the participants in our experiments was completely rational in this sense. Otherwise, they should have always inferred ignorance from superlative modifiers, since *at least* was always degraded in Experiments 1 and 3 when the speaker had precise knowledge. We therefore propose that participants performed the task in Experiments 2 and 4 in a simpler, sub-rational way.

Specifically, we will model participants’ behavior in Experiments 2 and 4 using unidirectional OT: participants compare the different possible knowledge states a speaker could be in, and select the state(s) s in which the highest violation incurred by φ is lower than the highest violation incurred in any other state s' (i.e., they optimize s given φ and Q). This is sub-rational because some expression φ' may be more optimal than φ for a speaker with information state s , so it may very well be that φ would not actually be used to convey s . Nevertheless, it is a good heuristic because it allows participants not to take alternative expressions into consideration—a move that is known to be

¹⁶ Note that the order of constraints in OT systems for pragmatics is usually assumed to be relatively stable across languages, unlike the order of constraints in OT systems for phonology, where cross-linguistic variation is captured in terms of different constraint orders. While cross-linguistic variation is much less explored in pragmatics, we expect that such variation is rather limited. Of course, the competing expressions will vary cross-linguistically, but the constraints are probably ordered in a similar way.

¹⁷ Alternatively, we could formulate our linking hypothesis based on Linear Optimality Theory (Keller 2000, 2006), in which each constraint is assigned a weight and the acceptability of an utterance is assumed to correlate with the sum of the weights of all constraints that it violates. This model has been applied successfully to syntactic phenomena and can in principle be extended to pragmatics (see, e.g., Farkas and Roelofsen 2019), but here we will stick to a more traditional OT framework and our simple linking hypothesis, since they are sufficient to capture our data (in particular, we did not observe the *ganging up* effect described by Keller, whereby several low ranked violations can lead to the same degradation as a single high ranked violation).

cognitively costly—without entirely ignoring pragmatic constraints such as those corresponding to the Gricean maxims. The main effect of this heuristics is to completely disregard constraints which, for a given φ , do not depend on s . This includes so-called *markedness* constraints such as NSAL and SIMP, which only refer to properties of φ , as well as ISAL in the case of *more than*. Indeed, *more than* violates ISAL no matter what s is, so ISAL will have no effect on the interpretation of comparative modifiers. In fact, ISAL will not affect the decision between knowledge and ignorance for other expressions either, because for any given n , ISAL does not distinguish between the knowledge states $\{n\}$ and $[n, \dots)$. The end result is that only QUAL, QUANT, and SQUAL will affect the predictions for Experiments 2 and 4, effectively neutralizing the contrast between *at least* and *more than*.

We further assume that participants behave as follows. In case the optimal state s is an ignorance state, participants pick the maximal degree of ignorance on the scale that they are presented with. In case the optimal state s is an exact knowledge state, they go for the other extreme of the scale. Finally, if ignorance and exact knowledge states are in a tie, we assume that participants fall back on their prior expectations about the speaker, which doesn't depend on φ .

3.4 Deriving predictions and capturing the experimental data

We will first present the main predictions of the model and give some insight into how the different parts of the model interact with each other. We then describe the predictions of the production model in greater detail, and how it captures the results of Experiments 1 and 3, followed by the predictions of the comprehension model for Experiments 2 and 4. As discussed in Sect. 3.1, we don't take the comprehension model to faithfully represent listeners in natural conversations, but rather to explain the behavior of participants in a psycholinguistic comprehension task. The production model on the other hand, is assumed to generalize beyond the experimental setting of Experiments 1 and 3, and will be explored in greater detail and compared to the predictions of competing proposals in the next sections.

We will discuss the predictions of our account for *at least*, *more than* and bare numerals, but will not discuss *at most* and *fewer than*. The pattern observed for negative modified numerals is similar, but the effects are stronger.

3.4.1 Getting a sense of the model's inner workings

Before showing in detail how the model captures our experimental findings, we will first state the two most basic predictions and explain how they come about:

1. Comparative modified numerals are only optimal in combination with round numerals. They do not require speaker ignorance but can only be used when conveying a precise number is not necessary (e.g., with polar or coarse-grained QUDs).
2. Superlative modified numerals require speaker ignorance and the numeral they combine with must match the exact minimum the speaker considers possible.

The distribution of comparative modified numerals is determined by an interaction between their semantics and the ISAL/NSAL constraints. Specifically, their use implies that the numeral they combine with cannot be part of the speaker's knowledge: if someone says "More than twelve students smoke", they immediately exclude the possibility that exactly twelve students smoke. This means that *more than* cannot ever satisfy ISAL in plain affirmative sentences. To stand a chance in the competition with bare numerals and *at least*, *more than* must at least satisfy NSAL, i.e., it must combine with a contextually salient/round numeral. If the QUD calls for a precise answer, however, it is often impossible to fall back on the closest round numeral without incurring a violation of QUANT. We therefore predict that *more than* is only used in combination with round or salient numerals, usually in the context of coarse-grained QUDs.

Second, since superlative modifiers are more costly than comparative modifiers and bare numerals, *at least* is only used when (i) SQUAL rules out bare numerals (i.e. the speaker does not have exact knowledge), and (ii) *at least* beats *more than* by satisfying ISAL (which *more than* always violates). This means that *at least* requires partial ignorance and must combine with the lowest numeral the speaker considers possible. We thus capture the usually accepted felicity conditions for *at least k* (see, e.g., Schwarz 2016a): the speaker must consider both *k* and some value above *k* possible, but does not need to consider each value above *k* possible.

To sum up, while most accounts attempt to derive the ignorance effects associated with superlative modifiers from their semantics or the alternatives they give rise to, the present account relies primarily on the fact that comparative modifiers always violate the ISAL constraint. The distribution of *at least* simply reflects the gaps left by bare numerals and *more than*. At this point, the ISAL constraint may seem like a one-trick pony whose sole function is to rule out *more than* in situations where *at least* is typically used. We will see in Sect. 3.5 that ISAL in fact plays a crucial role in capturing a surprising range of empirical facts that go beyond ignorance implicatures. But first, let us now dive into the details of how the model captures our experimental findings.

3.4.2 Production: Capturing Experiments 1 and 3

The tableaux for *how many* and polar QUDs are presented in Tables 8 and 9, respectively. For simplicity, we identify *s* with the set of values the speaker considers possible. Note that QUANT only has an effect on *how many* questions, since all expressions considered in Table 9 completely resolve the polar QUD (though see Sect. 3.4.4). Let us now explain how the tableaux translate into predictions for Experiments 1 and 3. Whenever relevant, we indicate the experimental conditions testing a given prediction.

Bare numerals: When the speaker has exact knowledge, the matching bare numeral is predicted to be acceptable with any QUD (Exp. 1: *Exact*; Exp. 3: *Precise*). Higher numerals violate QUAL so they are maximally degraded (Exp. 1: *False*), while lower numerals violate QUANT in a *how many* QUD context, and SQUAL in a polar QUD context, leading to mild degradedness and a QUD effect (Exp. 1: *Exceed*). When the speaker is ignorant, the numeral matching the lower-bound incurs a violation of

SQUAL, which is always fatal because of the *at least* alternative (Exp. 1: *Ignorance*; Exp. 3: *Approximate*).¹⁸

More than: Table 8 shows that *more than* is almost always predicted to be degraded in response to a (fine-grained) *how many* question. Our results confirm this: *more than* is indeed degraded with *how many* QUD, whether the speaker has exact knowledge or not.

Turning to polar questions, we see that *more than n* is now in a tie for optimality in every situation where the speaker's knowledge excludes the salient *n* (whether the speaker knows a precise number above *n*, or only a range). This was the case in all our target conditions in Experiment 3, and in the *PolarOverInf* case in Experiment 1. We correctly predict that *more than* is fully acceptable in these conditions. The *Polar-Relevant* case in Experiment 1—where *more than* was observed to be degraded—is predicted to be out because it violates both NSAL and ISAL.

At least: With precise knowledge, *at least* as a response to a (fine-grained) *how many* question always incurs a violation of QUANT, and is therefore predicted to be clearly degraded (Exp. 1: *Exact, Exceed*; Exp 3: *Precise*). In ignorance cases, *at least* becomes optimal (or in a tie to be optimal), and is therefore fully acceptable (Exp. 1: *Ignorance*; Exp. 3: *Approximate*).

For *at least*, the ignorance effect is predicted to persist in polar questions, but due to violations of lower-ranked constraints with precise knowledge (ISAL, NSAL, or SIMP).¹⁹ The model thereby captures the observation that *at least* always shows an ignorance effect, but the amplitude of this effect is modulated by QUD.

3.4.3 Comprehension: Capturing Experiments 2 and 4

The tableaux for *how many* and polar QUDs are presented in Tables 10 and 11, respectively. Only round numerals are considered since we did not test non-round numerals in these experiments. As these tableaux model the hearer's perspective, φ is fixed and s needs to be inferred. Moreover, we take the QUD to be fixed by the overt questions preceding each utterance. In real-life conversations, the QUD isn't always clearly set and speakers may decide to refine it before answering, so Q would sometimes have to be inferred as well (see Klecha 2018 for a mechanism of QUD-revision in an OT framework).

¹⁸ Note that the model misses a QUD effect here. Capturing it would require breaking down SQUAL into two constraints, as Grice (1975) suggests for quality: one that penalizes the speaker for using an utterance they know to be false (under at least one interpretation), and a stronger constraint requiring them to have evidence supporting all interpretations (as we assumed in our system).

¹⁹ It may be worth taking a moment to explain how exactly the model rules out *at least* with precise knowledge in the case of polar questions, as this is far from obvious. Take the cell where s amounts to $\{k\}$ in Table 9. While *at least k* and *at least n* only violate one of ISAL/NSAL each, they are still harmonically bounded by one of the winning candidates. Indeed, if ISAL dominates NSAL, the bare numeral k is optimal, *at least k* loses by violating SIMP, and *at least n* by violating ISAL. Conversely, if NSAL dominates ISAL, *more than n* is optimal, *at least k* loses by violating NSAL, and *at least n* by violating SIMP. This means that despite uncertainty about the relative ranking of ISAL and NSAL, there is always an optimal candidate better than *at least* if the speaker has precise knowledge. Assuming that participants are charitable, they retain the lowest possible fatal violation, SIMP in each case.

Table 8 OT Tableaux for Experiments 1 and 3, *how many* question (understood as a fine-grained *how many* question). n is a round number, k is a non-round number between $n + 2$ and the next round number. This is a tableau for production, so s and Q are fixed and only expressions φ are compared and evaluated against each other. Blocks separated by a horizontal line should be considered separately

| s | φ | QUAL | QUANT | SQUAL | ISAL | NSAL | SIMP |
|------------------|-------------------|------|-------|-------|------|------|------|
| $\{n\}$ | n | *! | | | | | |
| | more than n | | *! | | * | * | * |
| $\{n + 1\}$ | more than $n - 1$ | | *! | | | | * |
| | at least n | | *! | * | | | * |
| | at least $n + 1$ | | *! | | | * | ** |
| $\{k\}$ | n | *! | | * | * | | |
| | more than n | | *! | | * | * | * |
| | more than k | | * | | | | * |
| | more than $k - 1$ | | *! | | * | * | * |
| | at least n | | *! | | * | | * |
| $[n, \dots)$ | at least k | | | | | * | ** |
| | n | *! | | *! | *! | *! | * |
| $[n + 1, \dots)$ | more than n | | *! | * | * | | * |
| | at least n | | *! | *! | * | * | ** |
| | at least $n + 1$ | | | | | * | ** |
| $[k, \dots)$ | n | *! | | * | * | | |
| | more than n | | *! | | * | * | * |
| | more than k | | * | | | | * |
| | more than $k - 1$ | | *! | | *! | * | * |
| | at least n | | *! | | * | * | |
| | at least k | | | | | * | ** |

The predictions are straightforward. In *how many* questions, the only optimal knowledge states for modified numerals are ones involving ignorance. Both superlative and comparative modifiers therefore lead to equally strong ignorance inferences, as observed in the experimental results.

In polar questions, both types of modifiers lead to a tie between precise and imprecise knowledge states. As a consequence, we predict that participants fall back on their prior expectations with regard to the speaker’s knowledge. Given our experimental results, this must be leaning towards ignorance, but the conclusion is weaker than when participants can infer ignorance directly from the utterance. Crucially, the effect

Table 9 OT Tableaux for Experiments 1 and 3, polar question. n is the threshold for the polar question and is round, k is a non-round number above $n + 2$. This is a tableau for production, so s and Q are fixed and only expressions φ are compared and evaluated against each other. Blocks separated by a horizontal line should be considered separately

| s | φ | QUAL | QUANT | SQUAL | ISAL | NSAL | SIMP |
|------------------|---|------|-------|----------|-------------------|-------------------|---------------------------|
| $\{n\}$ | n more than n more than $n - 1$ at least n | *! | | | *! | *! | * * **! |
| $\{n + 1\}$ | n $n + 1$ more than n at least n at least $n + 1$ | | | *! | * * * | * * | * **! **! |
| $\{k\}$ | n k more than n more than k more than $k - 1$ at least n at least k | *! | | *! | * * *! * | * * *! * | * * * **! **! |
| $[n, \dots)$ | n more than n more than $n - 1$ at least n | *! | | *! | *! | *! | * * ** |
| $[n + 1, \dots)$ | n $n + 1$ more than n at least n at least $n + 1$ | | | *! *! | * * * | * * | * **! ** |
| $[k, \dots)$ | n k more than n more than k more than $k - 1$ at least n at least k | *! | | *! *! | * * *! * | * * *! * | * * * **! ** |

is again the same for *at least* and *more than*, so the model captures the absence of a difference between superlatives and comparatives in this experimental setting.

For bare numerals, the model correctly predicts no ignorance inference whatsoever, since their optimal knowledge state in both polar and *how many* contexts is always precise knowledge. The model also captures the absence of a QUD effect on bare numerals in these experiments, in contrast with Experiments 1 and 3.

Table 10 OT Tableaux for Experiments 2 and 4, *HowMany* question (understood as a fine-grained *how many* question). n is a round number. Cases beyond $n + 2$ behave similarly, up to the next round numeral. This is a tableau for comprehension, so φ and Q are fixed, and only possible information states s are compared and evaluated against each other. Blocks are separated by a horizontal line. Each block should be considered separately

| φ | s | QUAL | QUANT | SQUAL | ISAL | NSAL | SIMP |
|---------------|--|------|-------|-------|------|------|------|
| n | $\{n\}$ | | *! | * | * | | |
| | $\{n + 1\}$ | | *! | * | * | | |
| | $\{n + 2\}$ | | | *! | | | |
| | $[n, \dots]$ | | *! | * | * | | |
| | $[n + 1, \dots]$ $[n + 2, \dots]$ | | *! | * | * | | |
| more than n | $\{n\}$ | *! | | | | | * |
| | $\{n + 1\}$ | | *! | | * | | * |
| | $\{n + 2\}$ | | *! | | * | | * |
| | $[n, \dots]$ | *! | | | | | * |
| | $\{n + 1, \dots\}$ $[n + 2, \dots]$ | | | *! | * | | * |
| at least n | $\{n\}$ | | *! | | | | ** |
| | $\{n + 1\}$ | | *! | | * | | ** |
| | $\{n + 2\}$ | | *! | | * | | ** |
| | $[n, \dots]$ | | | | | | ** |
| | $[n + 1, \dots]$ $[n + 2, \dots]$ | | | *! | * | | ** |

3.4.4 Universal density of measurement

So far we have implicitly assumed a discrete scale for cardinalities. In particular, we have assumed that QUANT never distinguishes between *more than n* and *at least $n + 1$* . However, Fox and Hackl (2007) have argued that all scales are dense (the so-called Universal Density of Measurement hypothesis, abbreviated UDM). Here we discuss how adopting the UDM would affect the predictions of our account.

The first effect of the UDM would be to make *more than* fully unacceptable with precise *how many* questions. Recall that the only situation in which *more than n* was predicted to be compatible with a precise *how many* QUD was one in which n is round and the speaker’s knowledge is a range with lower bound $n + 1$. This situation resulted in a tie between *more than n* and *at least $n + 1$* , but with the UDM *more than* would violate QUANT. In Experiment 3, this situation was not tested, and in Experiment 1, we didn’t properly control for roundness and salience.²⁰ Nevertheless, we think that the prediction of the UDM here is correct. In response to “How many students registered for the course exactly?”, *more than 30* intuitively does not convey as much information

²⁰ Specifically, only *more than 1/2/3* were tested with *how many* questions. Such small numbers are extremely easy to process (they fall in the so-called *subitizing range*) and all count as round by Jansen and Pollmann’s (2001) measure. When both n and $n + 1$ satisfy NSAL, we predict *more than n* to be harmonically bounded by either the bare numeral $n + 1$ or *at least $n + 1$* .

Table 11 OT Tableaux for Experiments 2 and 4, polar question. n is the threshold for the polar question and is round. Cases beyond $n + 2$ behave similarly. This is a tableau for comprehension, so φ and Q are fixed, and only possible information states s are compared and evaluated against each other. Blocks are separated by a horizontal line. Each block should be considered separately

| φ | s | QUAL | QUANT | SQUAL | ISAL | NSAL | SIMP |
|---------------|---|----------|-------|----------------------------|----------------------|------|----------------------------|
| n | \mathbb{E}^{Q} $\{n\}$ \mathbb{E}^{Q} $\{n + 1\}$ \mathbb{E}^{Q} $\{n + 2\}$ \mathbb{E}^{Q} $[n, \dots)$ \mathbb{E}^{Q} $[n + 1, \dots)$ \mathbb{E}^{Q} $[n + 2, \dots)$ | | | *! *! *! *! *! | * * * * | | |
| more than n | \mathbb{E}^{Q} $\{n\}$ \mathbb{E}^{Q} $\{n + 1\}$ \mathbb{E}^{Q} $\{n + 2\}$ \mathbb{E}^{Q} $[n, \dots)$ \mathbb{E}^{Q} $[n + 1, \dots)$ \mathbb{E}^{Q} $[n + 2, \dots)$ | *! *! | | | * * * * | | * * * * * |
| at least n | \mathbb{E}^{Q} $\{n\}$ \mathbb{E}^{Q} $\{n + 1\}$ \mathbb{E}^{Q} $\{n + 2\}$ \mathbb{E}^{Q} $[n, \dots)$ \mathbb{E}^{Q} $[n + 1, \dots)$ \mathbb{E}^{Q} $[n + 2, \dots)$ | | | | *! *! *! *! | | ** ** ** ** ** |

as *at least 31*. This prediction could be captured by other mechanisms however (for instance, Dehaene 2011 and Krifka (2009) show that round numbers tend to have an approximative meaning).

The second effect of the UDM would be to make *more than n – 1* even more degraded as an answer to a polar QUD. Without the UDM, it was already non-optimal (because it violated both ISAL and NSAL), but it would now violate QUANT as well. Experiment 1 includes cases corresponding to this situation (*PolarRelevant* QUD with *Exact* and *Ignorance* SITUATION), and they do appear to be as degraded as other violations of QUANT, suggesting that the UDM is again on the right track.

Finally, the UDM would affect the predictions for Experiments 2 and 4. In particular, since any information state would violate QUANT with *more than* after precise *how many* questions, all the information states that pass QUAL in Table 10 would be equally possible. This would predict no effect of QUD on the ignorance inferred from *more than*, in contradiction with our results. If we were to adopt the UDM, this last issue could be worked around by assuming different levels of violations for QUANT, as proposed in Cummins (2011, §2.4.1) or Klecha (2018) (so that *more than* incurs more violations if the speaker has exact knowledge and the QUD is precise).

To sum up, adopting the UDM would require a graded version of QUANT to keep the good predictions of the comprehension model intact, but would not affect the predictions of the production model much.

3.5 Further predictions

We will now discuss several predictions that the proposed model makes beyond the empirical data it was designed to account for. Discussing these predictions is particularly important in order to provide further support for the ISAL constraint. While the other constraints in our model are familiar from other work and have received much independent support in other empirical domains, the ISAL constraint is new, and has not so far been motivated on independent grounds. The fact that, as we will see, it makes correct predictions for a range of other puzzles related to modified numerals provides such independent motivation.

The first prediction relates to the observation that speakers, when using *at least* with a *non-numerical* prejacent, do not always convey that they consider the prejacent itself possible (Mendia 2016). The second prediction regards the effect of polarity on modified numerals. It has been noticed that *at least* is degraded under negation. Our account captures this effect and makes a number of subtle predictions for other embeddings, assuming ISAL is generalized to quantified sentences. We then show that the model may also shed new light on the behavior of modified numerals in the scope of quantifiers and modals, although a detailed account of this behavior is left for future work.

3.5.1 Mendia's observation

Consider sentence (20) below. We have seen that such sentences can only be optimal if the speaker considers it possible that exactly two students completed the quiz, and therefore implicate that she is uncertain as to whether exactly two students completed the quiz or more.

(20) At least two students completed the quiz.

However, Mendia (2016) argues that this inference depends crucially on the fact that, for numerals, the relevant comparison class is totally ordered. In cases which involve a partially ordered comparison class, such as (21), the speaker might in fact know that Ann and Bill are not the only students who completed the quiz.

(21) At least Ann and Bill completed the quiz.

To support this claim, Mendia provides experimental data showing that the answer in (22b) is regarded considerably more acceptable than the one in (23b).

- (22) a. Who completed the quiz?
 b. I don't remember, at least Ann and Bill. But not only them.
- (23) a. How many students completed the quiz?
 b. I don't remember, at least two. #But not only two.

This leads Mendia to propose the following generalization:

- (24) a. When the relevant comparison class forms a partial order, *at least* need not convey speaker ignorance as to whether the exhausted prejacent holds;

- b. When the relevant comparison class is a total order, *at least* does always convey speaker ignorance with respect to the exhausted prejacent.

To capture this contrast, we first need to specify how ISAL would extend to scales other than numerals, since this is the constraint mainly responsible for ignorance inferences concerning the prejacent (see Sect. 3.4.2). Given a partially ordered set, we can generalize ISAL using the least upper bound and greatest lower bound of the speaker’s information state, to the extent that they exist. Consider the case of (22b): the domain of individuals is partially ordered by the i-part relation (Link 1983). If the speaker knows that Ann, Bill, and exactly one of Clara, David and Elliot completed the quiz, then her information state amounts to the set $\{a \oplus b \oplus c, a \oplus b \oplus d, a \oplus b \oplus e\}$. While this set has neither a minimum nor a maximum in D_e , it has a greatest lower bound $a \oplus b$ and a least upper bound $a \oplus b \oplus c \oplus d \oplus e$. Assuming as we did above that these are the boundaries ISAL is concerned with, this means that the speaker can use “at least Ann and Bill” without violating ISAL, even when she knows that not only Ann and Bill completed the quiz.

We therefore predict that ISAL can be satisfied without considering the exhausted prejacent possible, to the extent that the set of values compatible with the speaker’s information state has a greatest lower bound that is not a minimum. This is a necessary condition for the use of *at least*, but it is not sufficient. Indeed open intervals in a totally ordered dense scale also make it possible to have a greatest lower bound without a minimum, but the oddness of (25b) suggests that *at least* still conveys ignorance towards the exhausted prejacent in this case.

- (25) a. How far is the university from here?
- b. I don’t remember, at least 3km. #But not (exactly) 3km.

Our account captures this fact as well. In order to satisfy ISAL without considering exactly 3 km possible, the speaker would have to consider any distance greater but arbitrarily close to 3 km possible. But if that were the case, *more than 3 km* would be optimal since it would satisfy every constraint *at least 3 km* satisfies, while incurring one less violation of SIMP, as the following tableau shows:

| <i>s</i> | φ | QUAL | QUANT | SQUAL | ISAL | NSAL | SIMP |
|----------|---------------|------|-------|-------|------|------|------|
| (3, ...) | 3km | | | *! | | | |
| | more than 3km | | | | | | * |
| | at least 3km | | | | | | **! |

The important difference between (25b) and (22b) is that—at least in English—partially ordered comparison classes do not license comparative modifiers (*more than Ann and Bill completed the quiz* is ungrammatical for some reason). Without a comparative competitor, *at least* only needs to beat its non-modified alternative to become optimal. This requires some ignorance, but not necessarily ignorance with respect to the exhausted prejacent.

To sum up, we derive Mendiá’s generalization (24) from two facts: (i) comparison sets can have a greatest lower bound which is not a minimum, and (ii) comparative

modifiers require a total ordered comparison set. The use of *at least* without speaker ignorance regarding the exhausted prejacent is licensed just in case the comparison set has a greatest lower bound which is not a minimum (needed to satisfy ISAL), and the set is not totally ordered (needed to block the competition with *more than*, which now also satisfies ISAL).

3.5.2 Polarity effects

Geurts and Nouwen (2007) noted that superlative but not comparative modified numerals are degraded under negation, as shown in (26).

(26) Betty didn't have ??at least / more than three martinis.

Nilsen (2007) further shows that *at least* is degraded in a number of other downward-entailing environments, as seen in (27), but not in the restrictor of a universal quantifier or in the antecedent of a conditional, as seen in (28).²¹ These contrasts have been confirmed experimentally in Mihoc and Davidson (2019), Mihoc (2019, chap. 5).

- (27) a. ??John hardly ate at least three apples
 b. ??Policemen rarely carry at least two guns.
 c. ??This won't take at least an hour.
- (28) a. Everybody who cancelled at least three days in advance got their money back.
 b. Nobody who cancelled at least three days in advance had to pay the fees.
 c. If Bob cancelled at least three days in advance, then he got his money back.

Our account directly captures the unacceptability of *at least* under negation. Remember that, given its additional violation of SIMP, *at least* can only win against *more than* by satisfying ISAL. However, negation flips the strict and non-strict comparison: *not . . . at least three* mentions but excludes 3, while *not . . . more than three* is compatible with 3. This means that under negation, *at least* cannot ever satisfy ISAL, thereby losing its natural advantage over *more than*.²² We further predict that *more than* loses its roundness/salience requirement under negation, as it can now satisfy ISAL. Our

²¹ Nilsen (2007) also shows an effect of valence: the acceptability of *at least* in a restrictor/antecedent seem to depend on whether the scope/consequent is seen as positive or negative. We will not discuss this effect here, and to the best of our knowledge, no account of this effect has been proposed.

²² An anonymous reviewer wonders whether our account wouldn't still allow *at least* under negation when it satisfies NSAL. This situation would arise when the speaker's knowledge falls just short of reaching a salient number (for instance, the speaker knows that Ugné ate up to 9 cookies, when 10 was a salient threshold). In this case, *not at least 10* would violate ISAL but satisfy NSAL, while *not more than 9* would violate NSAL but not ISAL. The reviewer points out that *less than* might enter the competition and be simpler than *not at least*, but that this would raise some questions about rescuing of *at least* in double negative environments. Yet, our account still predicts the former to be blocked by the negated bare numeral *not 10* which, in this case is ambiguous between a negated exact meaning and an 'at most' reading. Nevertheless, rescuing under double negation remains a puzzle, and competition between different negative forms is also known to give rise to evaluativity inferences (Moracchini 2018), which further complicates the matter.

intuitions, together with a quick corpus analysis, suggest that this prediction is correct.²³

Turning to restrictors, our current definition of ISAL falls short, since we're not talking about a single quantity but a set of quantities, for instance, the set containing every n such that some individual cancelled n days in advance in Nilsen's examples. Generalizing ISAL to sets of quantities would be necessary to make fully explicit predictions about such cases, but we can already point out one essential difference between negation and the restrictor of *every*: while *not . . . at least n* excludes *exactly n* , (28a) does not exclude that someone who cancelled *exactly three days* in advance got their money back, and in fact it entails that everyone who did cancel exactly three days in advance got their money back. This suggests that any reasonable generalization of ISAL would not in fact block *at least* in the restrictor of *every* or in antecedents of conditionals. By contrast, the sentence with *no* does exclude that anyone who cancelled exactly three days in advance had to pay the fees. Here the exact way we generalize ISAL may matter. We also want to point out that Nilsen's examples may not be ideal, because they involve very small numbers, so it could be that (28b) is only acceptable because 3 is in the subitizing range. Looking at higher numerals suggests that there is in fact a contrast between *everybody* and *nobody*:

- (29) A hotel's official policy is to charge a fee for cancellations within the last month, but they tend to be lenient for cancellations that are only slightly late.
- a. Everybody who cancelled at least 27 days in advance got the fees waived.
 - b. ?Everybody who cancelled more than 26 days in advance got the fees waived.
 - c. ?Nobody who cancelled at least 27 days in advance had to pay the fee.
 - d. Nobody who cancelled more than 26 days in advance had to pay the fee.

According to the few native speakers we consulted, (29b) and (29c) are slightly degraded compared to (29a) and (29d) respectively, as would be expected if, in the restrictor of negative quantifiers, it was *at least* and not *more than* that violated ISAL. However, the theoretical and empirical work that would be necessary to properly extend our account to these examples is beyond the scope of the present manuscript.

3.5.3 Quantified and modal sentences

Modified numerals are known to give rise to specific effects not only in the restrictor of a quantifier, but also in the scope of a quantifier or modal operator. In particular, superlative modified numerals in the scope of a universal quantifier do not necessarily give rise to ignorance effects, but can give rise to so-called *variation effects* instead (e.g., Alexandropoulou et al. 2015):

²³ We searched the CoCA corpus (Davies 2008) for *more than*, *no more than*, and *not more than*, combined with all numerals between 1 and 100. Counts for *not/no more than* were subtracted from counts for *more than*. Base roundness was set to 0 and increased by 1 for each divisor in {2, 5, 10, 25} (Jansen and Pollmann 2001), except for numerals 1–5, the roundness of which was set to 2. A linear regression was done on log-counts normalized by quantifier with roundness, presence of negation and their interaction as predictors. All factors were significant, and in particular we observed a significant negative interaction, indicating that roundness had less effect when negation was present ($t = -2.8, p < .01$).

- (30) Every street was guarded by at least three policemen.
 \rightsquigarrow Some streets were guarded by three policemen; some by more than three

This is entirely expected if bare numerals can receive an exact reading in the scope of the quantifier. In this case, it is possible for the sentence with the bare numeral to fail SQUAL even when the speaker is fully knowledgeable. Our account then predicts the following felicity conditions for (30):

- (i) Every street is guarded by three or more policemen (the literal reading), and
- (ii) The speaker doesn't know for sure that every street is guarded by exactly three policemen (bare numeral filtered by SQUAL), hence some streets may be guarded by more than three, and
- (iii) Some streets must/may be guarded by exactly three policemen (depending on how ISAL is generalized to quantified sentences).

We therefore predict that ignorance is possible but not necessary, and if we assume that the speaker is in fact knowledgeable about exactly how many policemen guarded each street, we capture the variation effect as described in Alexandropoulou et al. (2015): some streets are guarded by exactly three policemen, some by more than three.

Similarly, our account sheds new light on the interaction between modified numerals and modals, which motivated Geurts and Nouwen's assumption that superlative modifiers are undercover modals. In particular, our account immediately captures the observation in (31) that *at least* under possibility modals only has an ignorance reading.

- (31) The paper can be at least 12 pages long \rightsquigarrow only ignorance reading

To see what happens in (31), let us show why a knowledgeable speaker cannot use this sentence to answer the QUD "Which lengths (in pages) are allowed for this assignment?". Let us first set aside the inverse scope reading of (31): if the modified numeral has wide scope, it is unembedded and therefore necessarily conveys ignorance. For any number k , the speaker, by assumption, knows whether the paper is allowed to be k pages long or not, hence we can identify the smallest such k , k_{\min} . By definition, for any $k < k_{\min}$, the assignment is not allowed to be k -pages long, hence it follows that the paper is *required* to be at least k_{\min} pages long (here the assumption that the speaker is knowledgeable is crucial). If no other page length is allowed (i.e. if the assignment must be exactly k_{\min} -pages long), then "the paper can/must be at least k_{\min} pages" is out in favor of a bare numeral. Assuming that the paper can also be k -pages for some $k \neq k_{\min}$, and that $k_{\min} - 1$ isn't particularly salient (so that *more than* is out), "the paper can/must be at least k_{\min} pages long" become possible options, but the necessity modal is more informative than the possibility modal, hence *at least* can only be used under the necessity modal. In short, assuming that the speaker is fully knowledgeable, *at least* under a possibility modal only satisfies ISAL when a necessity modal could be used instead. As a result, we capture the observation that *at least* can only be used under a possibility modal when the speaker is ignorant of the exact numbers allowed.

This is far from saying that our account captures the interaction between modified numerals and modals. In particular, the account says nothing about the second puzzle

that motivated Geurts and Nouwen (2007), namely that *at most* under possibility modals always imposes a strict upper bound, as illustrated in (32) (see Blok 2015, 2019 for additional puzzles). Nevertheless, we hope to have shown that pragmatic solutions should be explored in greater detail before departing from a plain semantics for modified numerals.

(32) The paper can be at most 12 pages long \rightsquigarrow the paper cannot exceed 12 pages

4 Comparison with other recent proposals

Before concluding we briefly compare our proposal with a number of other recent accounts, in light of our experimental data and the additional predictions discussed in Sect. 3.5.

4.1 Other OT proposals

As mentioned already, our account strongly builds on Cummins (2011, 2013). However, it improves on it in a number of ways. First, we explicitly account for QUD effects through our QUANT constraint. Second, we account for the stronger ignorance effect of superlative modifiers without having to encode it in the meaning of non-strict comparison, as Cummins (2011) does (Cummins 2013 does not discuss superlative modifiers). Third, our proposal correctly predicts that superlative modifiers are *less* sensitive to roundness than comparative modifiers, while Cummins predicts the opposite.²⁴

There are however a few questions discussed in Cummins (2011) that we haven't delved into, including other modifiers such as *exactly* and *about*, as well as a more precise notion of granularity. Given the proximity between our models, the solutions he adopts could be transposed straightforwardly to our account.

Another relevant OT proposal—though not directly addressing comparative and superlative modified numerals—is Klecha (2018).²⁵ Klecha proposes an OT account of imprecision, in particular with bare numerals and approximatives (e.g., *about 50*).

While they were developed to account for different problems, our models have a lot in common. There are some crucial differences however: (i) Klecha's Quality constraint (Faithfulness) is QUD-dependent, (ii) his Quantity constraint (Informativity) is graded, (iii) his model has no constraint like ISAL. Without any constraint like ISAL, non-round numerals are always predicted to be degraded in low-precision contexts (and therefore require implicit adjustment of the context to a high-precision one when used). Our model occasionally allows non-round numerals in low-precision contexts, as long as they are particularly salient to the speaker. As a result, their use does not

²⁴ Indeed, Cummins derives the roundness preference of comparative modified numerals from the trade-off between simplicity (which favors bare numerals) and numeral salience (which favors round numerals), in situations where quantity is neutralized (i.e., polar questions or coarse-grained *how many* questions). Superlative modified numerals are assumed to be even more complex than comparative modified numerals, and as a result they can only ever be used in situations where the speaker wants to convey a range with a lower bound that is round or contextually salient (see Cummins 2011, p. 106).

²⁵ We'd like to thank an anonymous reviewer for bringing this paper to our attention.

require the listener to implicitly adjust the QUD in our model. We could in principle follow Klecha and prevent the acceptability of non-round numerals in imprecise context entirely, assuming that superlative modified numerals indicate a switch to a precise QUD (as Westera and Brasoveanu 2014 do). However, the crucial role of ISAL is to rule out comparative modifiers with non-round numerals, and it is unclear how this can be done without a constraint like ISAL. In other words, Klecha would have to explain why *at least 32* can force adjustment to a high-precision context, while *more than 32* conveys that there is something special about 32. Conversely, some of Klecha's ideas could be imported in our model, notably his approach to capturing imprecise uses of (bare) round numerals and the graded Quantity constraint (two desiderata we quickly mentioned in Sect. 3.4.4 when discussing the UDM hypothesis).

4.2 Neo-Gricean and grammatical implicature approaches

There has been a lot of work on modified numerals recently both within the neo-Gricean approach to implicatures (Nouwen 2015; Kennedy 2015; Alrenga 2018) and within the grammatical approach (Mayr 2013; Schwarz 2016b; Enguehard 2018; Buccola and Haida 2020; Mihoc 2019).

These proposals generally assume a naive denotation for modified numerals, but propose a specific set of formal alternatives for superlative and comparative modified numerals, and fine-tune the mechanism for implicature derivation/exhaustification in order to explain certain puzzles concerning the interpretation of modified numerals.

They offer in-depth analyses of certain complex issues (for instance, the interaction with modals and quantifiers). Our own model was originally designed to capture the distribution and interpretation of modified numerals in very simple affirmative sentences only, though we showed in Sect. 3.5 that it also makes promising predictions about the issues targeted by the neo-Gricean and grammatical approaches. Going in the other direction, it is less obvious that any of the proposals mentioned above could be extended to properly capture all properties of modified numerals. The main degrees of freedom in a neo-Gricean or grammatical approach are the alternatives that are assumed for each expression, and the exact mechanism for implicature derivation / exhaustification. The various existing proposals fix these parameters in a way that is incompatible with other proposals, which makes it impossible—for now at least—to synthesize the various pieces of work, each addressing different issues, into a single unified account.²⁶

4.3 Westera and Brasoveanu (2014)

Westera and Brasoveanu (2014) account for the contrast between ignorance inferences of superlative and comparative modified numerals under the assumption that superlatives are typically used to answer precise QUDs (e.g., “How many exactly?”) while

²⁶ Kennedy (2015), Schwarz (2016b) and Alrenga (2018) leave comparative modified numerals aside to focus on superlatives. Only Enguehard (2018) captures the roundness sensitivity of comparatives, but leaves superlatives out of consideration. Polarity effects are particularly difficult to capture (only addressed in Mihoc 2019, by postulating different exhaustification operators for scalar and domain alternatives).

comparatives are typically used to answer coarse-grained or polar QUDs. When the QUD is not fully specified and a speaker uses a superlative modified numeral, the listener is taken to infer a more precise QUD than if the speaker had used a comparative modified numeral. Given a QUD-sensitive mechanism for the derivation of quantity implicatures, this would lead to stronger ignorance inferences with superlative than with comparative modifiers. Their experimental results seem to support this conclusion, because when controlling for the QUD, they find no difference between the two types of modifiers. In addition, a corpus study shows that comparative modifiers associate more frequently with round numbers than superlative modified numerals, supporting the idea that they are used to answer coarser QUDs.

W&B's experimental results were replicated in our experiments 2 and 4, but our other experiments show that this picture is incomplete and more likely reflects a heuristic that participants use in the inference task. In our production-like acceptability tasks, we did find a clear contrast between the two modifiers even when controlling for the QUD. Examples like (33), where *at least* is infelicitous due to its ignorance implicature (even in the context of a polar QUD), also remain unexplained on W&B's account.²⁷

- (33) *Context: you must have three children or more to benefit from a tax ruling.*
 A: Do you qualify for the tax rebate?
 B: Yes, I have more than three children / ??at least three children.

In addition, the assumption that different modifiers evoke different QUDs, while indirectly supported by their corpus analysis, is not explained. Deriving it from the fact that superlative modifiers convey ignorance would be circular.

Our account improves on W&B's in three respects. First, it explains the corpus results without falling into circularity, since the explanation does not make any reference to the QUD or ignorance inferences. The interplay between NSAL and SIMP is sufficient to predict that comparative modifiers are only optimal when combined with round or contextually primed numerals.

Second, it derives the assumed difference in evoked QUD rather than postulating it. More precisely, we predict that comparative modifiers are usually unacceptable with fine-grained *how-many* QUDs, because of their roundness sensitivity and QUANT. As a consequence, a listener trying to reconstruct the QUD the speaker had in mind is very unlikely to assume a fine-grained QUD upon hearing a comparative modified numeral.

Finally, we predict that *at least* is unacceptable when the speaker is knowledgeable, even with coarse-grained or polar QUDs, as seen in (33) and in our Experiments 1 and 3 (while the ignorance effect is smaller with polar questions, it does not disappear).

²⁷ Barring the assumption that B is actually answering a fine-grained *how-many* QUD when using the superlative modifier. This assumption is problematic because one would have to assume that *at least* not only evokes, but *forces* accommodation of a fine-grained QUD, even when it is in contradiction with contextual information that the speaker is ignorant, and hence wouldn't use a superlative modifier to answer a fine-grained QUD.

4.4 Inquisitive semantics approaches

Ciardelli et al. (2018a) propose that the ignorance inferences of modified numerals involve both quantity implicatures (following Schwarz 2016b and others) and implicatures derived from a pragmatic maxim called *inquisitive sincerity* (following Coppock and Brochhagen 2013b). The latter is taken to trigger ignorance inferences with superlative modifiers, which are assumed to be semantically inquisitive, but not with comparative modifiers, which are taken to be non-inquisitive. Ignorance inferences can arise through the maxim of quantity with both types of modifiers, but only in the context of a *how many* QUD. This proposal correctly predicts the contrasts we found in Experiments 1 and 3, but it does not address roundness or salience effects, and it's unclear what its predictions beyond simple affirmative sentences would be.

There is also a conceptual issue with this approach, in that it relies on the assumption that declarative sentences with superlative modifiers such as (34) are semantically inquisitive. This means that (34) is semantically equivalent with the question in (35).

(34) Bill ate at least two apples.

(35) Did Bill eat exactly two apples or did he eat more than two?

As Ciardelli et al. (2018a, footnote 2) discuss, this requires a certain perspective on the connection between inquisitiveness, a semantic notion, and the communicative effects of sentences when uttered in discourse. The perspective that Ciardelli et al. (2018a) adopt, following Groenendijk (2009) and Coppock and Brochhagen (2013b), among others, is that even if a sentence is semantically inquisitive, i.e., even if its denotation contains multiple alternative propositions, a speaker who utters this sentence in discourse does not necessarily request a response from the addressee which confirms one of these alternatives. That is, under this perspective, uttering an inquisitive sentence does not necessarily amount to asking a question. This assumption is incompatible with much other work on inquisitive semantics (see, e.g., Farkas and Roelofsen 2017; Ciardelli et al. 2018b), in which uttering an inquisitive sentence does always amount to asking a question.

Even under the assumption that uttering an inquisitive sentence does *not* necessarily amount to asking a question, it is difficult to construe independent evidence for the assumption that sentences like (34) are semantically inquisitive. The parallel between *at least* and disjunction noted by Büring (2008) relies mostly on their similar ignorance effects, so using it to justify the inquisitiveness assumption would be circular.

Finally, Blok (2019) develops an extensive account of the interaction between modified numerals and modal operators which crucially relies on the assumption that superlative modified numerals are inquisitive, as assumed in Coppock and Brochhagen (2013b) and Ciardelli et al. (2018a). While we've shown that our pragmatic model can account for some basic facts about the interaction between modified numerals and modals, we leave open whether the more complex issues she addresses require a more sophisticated semantics or not.

5 Conclusion

The ignorance inferences triggered by modified numerals have received considerable attention over the past decades. However, the empirical generalizations that have been proposed partly contradict each other, and similarly, the various theoretical accounts that have been developed disagree on the differences between comparative and superlative modifiers, the role of the question under discussion, and the nature of ignorance inferences (although there seems to be a consensus that they constitute some sort of implicatures).

We have presented a set of experiments examining how ignorance inferences depend on the modifier and the QUD, as well as the task that is used to probe them. Overall, our results show that superlative modified numerals give rise to robust ignorance inferences, while comparative modified numerals *can* also convey ignorance, but do not necessarily do so. Furthermore, all ignorance inferences are affected by the QUD. Finally, we found a striking difference between experiments in which ignorance inferences were probed by means of an acceptability task, and ones where an inference task was used. This result is reminiscent of previous findings by Degen and Goodman (2014) regarding scalar implicatures.

Building on the work of Cummins (2011, 2013), we proposed a theoretical account of our results as well as previous experimental work (in particular Westera and Brasoveanu 2014). Couched in Optimality Theory, our proposal offers an explanation of the observed task effect and accounts for differences between superlative and comparative modifiers, effects of the QUD, and effects of roundness and contextual salience of the base numeral.

Crucial for our account of the differences between superlative and comparative modifiers is the ISAL constraint (already anticipated though not implemented in Cummins 2011), which favors the use of numerals that are internally salient to the speaker. This constraint does not look at the modifiers themselves but at the base numeral. The modifiers are assumed to receive a naive denotation corresponding to strict comparison for comparatives, and non-strict comparison for superlatives. The complex empirical picture is entirely captured by the interaction between a few constraints, all of which find independent support in the literature, except for ISAL.

Our proposal is compatible with the experimental literature on modified numerals more broadly, which strongly supports a pragmatic approach to ignorance inferences (Cummins and Katsos 2010; Alexandropoulou et al. 2016, 2017; Alexandropoulou 2018; see Nouwen et al. 2019 for a review). Unidirectional OT offers a good model of the kind of heuristics participants may be using to deal with psycholinguistic tasks which are too demanding for participants to fully reconstruct speakers' intentions by comparison with alternative expressions. Compared to simpler models of 'literal' behavior, unidirectional OT still predicts some implicatures to arise. In particular, modified numerals are predicted to give rise to ignorance inferences in response to precise QUDs because of Quantity.

Modified numerals present a number of puzzles beyond ignorance effects. We showed that our model accounts for a few of these (unacceptability of *at least* under negation, weaker ignorance with partial orders) and could possibly account for more if we generalized the ISAL constraint to more complex sentences (involving interaction

with quantifiers and modals). We take all these additional predictions as post hoc justification for the ISAL constraint we postulated, and for our model more generally. There are of course a number of issues that we must leave open for future work. Cummins et al. (2012) observe that comparative modified numerals (and possibly superlative too) can give rise to upper-bounding implicatures in the context of coarse-grained QUDs: for instance *more than 90* may imply not more than 100. Our account does not immediately capture this effect, but could easily do so if we adopted the granularity constraint from Cummins et al.'s system. Alternatively, we could assume that exhaustification is sensitive to granularity, as Enguehard (2018) does. There is also much more to say about the well-studied interaction between modified numerals and modals and quantifiers, we haven't touched upon the differences between positive and negative modified numerals either (in particular between *at least* and *at most*; Penka 2014 among others), and we restricted our attention to very simple knowledge states for the speaker (exact knowledge or knowledge of a lower bound).

When more complex knowledge states are considered, for instance "between 12 and 20, but not 16", the model would automatically give precedence to very complex expressions describing these possibilities exhaustively (because SIMP is the lowest-ranked constraint). This illustrates a possible limitation of OT compared to more flexible models such as the RSA model, which allow for a probabilistic trade-off between Manner and Quantity. Conversely, an RSA account would have difficulty capturing the complex interplay between the different constraints, as it would operate with a single cost variable to represent all Manner effects.

Overall, we think that our model has a better shot than competing accounts at ultimately capturing the behavior of modified numerals in a unified way. In particular, despite very active research on modified numerals within the Neo-Gricean and grammatical approaches to implicatures, no current proposal fully captures our experimental findings. Unlike recent accounts (Mihoc 2019; Buccola and Haida 2019, 2020) which adopt the purely grammatical view of Meyer (2013), where even ignorance implicatures are derived in the grammar, our approach can be seen as a revival of Fox (2007). Exhaustification is done in the grammar, but ignorance implicatures are derived by a pragmatic apparatus, albeit a more complex one than the usual "derive ignorance for every alternative that is neither entailed nor negated". We hope to have shown that the proposed division of labor between semantics and pragmatics offers a path towards a unified account of modified numerals, and possibly numerical expressions more generally.

Summing up, our experimental results have clarified the empirical properties of ignorance inferences triggered by modified numerals, and we have argued that these inferences, as well as roundness and priming effects, can be captured by a small set of pragmatic constraints, combined with the simplest possible semantic treatment of modified numerals. We further showed that this model could lead to new explanations for several of the puzzling phenomena associated with modified numerals that have been thought to require a more sophisticated semantics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexandropoulou, S. (2018). *On the pragmatics of numeral modifiers: The availability and time course of variation, ignorance and indifference inferences*. Utrecht: LOT.
- Alexandropoulou, S., Dotlačil, J., McNabb, Y., & Nouwen, R. (2015). Pragmatic inferences with numeral modifiers: Novel experimental data. *Semantics and Linguistic Theory (SALT)*, 25, 533–549.
- Alexandropoulou, S., Dotlačil, J., & Nouwen, R. (2016). *At least* ignorance inferences come at a processing cost: Support from eye movements. *Semantics and Linguistic Theory (SALT)*, 26, 795–813.
- Alexandropoulou, S., Dotlačil, J., & Nouwen, R. (2017). Pragmatic effects of *more than* and *at least* in incremental interpretation. *Semantics and Linguistic Theory (SALT)*, 27, 680–699.
- Alrenga, P. (2018). *At least* and *at most*: Scalar focus operators in context. Ms., Boston University.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. arXiv preprint [arXiv:1506.04967](https://arxiv.org/abs/1506.04967).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Blok, D. (2015). The semantics and pragmatics of directional numeral modifiers. *Semantics and Linguistic Theory (SALT)*, 25, 471–490.
- Blok, D. (2019). *Scope Oddity: On the semantic and pragmatic interactions of modified numerals, negative indefinites, focus operators, and modals*. Ph. D. thesis, Utrecht University.
- Bloom, P., Barss, A., Nicol, J., & Conway, L. (1994). Childrens knowledge of binding and coreference: Evidence from spontaneous speech. *Language*, 70(1), 53–71.
- Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17(3), 189–216.
- Boersma, P. (1997). How we learn variation, optionality, and probability. In R. van Son (Ed.), *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* (Vol. 21, pp. 43–58). Amsterdam: IFA University of Amsterdam.
- Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 53, 437–457.
- Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, 25(2), 93–139.
- Buccola, B., & Haida, A. (2019). Obligatory irrelevance and the computation of ignorance inferences. *Journal of Semantics*, 36(4), 583–616.
- Buccola, B., & Haida, A. (2020). How obligatory irrelevance, symmetric alternatives, and dense scales conspire: The case of modified numerals and ignorance. *Semantics and Linguistic Theory (SALT)*, 30, 464–484.
- Büring, D. (2008). The least *at least* can do. In C. B. Chang, & H. J. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics* (pp. 114–120). Somerville, MA: Cascadilla Press.
- Chien, Y.-C., & Wexler, K. (1990). Childrens knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, 1(3), 225–295.
- Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In P. Portner, C. Maienborn, & K. von Stechow (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2297–2331). Berlin: Mouton de Gruyter.
- Ciardelli, I., Coppock, E., & Roelofsen, F. (2018a). Implicatures of modified numerals: Quality or quantity? *Proceedings of Sinn und Bedeutung*, 21(1), 283–300.
- Ciardelli, I., Groenendijk, J., & Roelofsen, F. (2018b). *Inquisitive semantics*. Oxford: Oxford University Press.
- Coppock, E., & Brochhagen, T. (2013a). Diagnosing truth, interactive sincerity, and depictive sincerity. *Semantics and Linguistic Theory (SALT)*, 23, 358–375.

- Coppock, E., & Brochhagen, T. (2013b). Raising and resolving issues with scalar modifiers. *Semantics and Pragmatics*, 6(3), 1–57.
- Cummins, C. (2011). *The interpretation and use of numerically-quantified expressions*. Ph. D. thesis, University of Cambridge.
- Cummins, C. (2013). Modelling implicatures from modified numerals. *Lingua*, 132, 103–114.
- Cummins, C., & Katsos, N. (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics*, 27(3), 271–305.
- Cummins, C., Sauerland, U., & Solt, S. (2012). Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy*, 35(2), 135–169.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>.
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, 54(2), 128–133.
- De Villiers, J., Cahillane, J., & Altreuter, E. (2006). What can production reveal about Principle B? *Generative Approaches to Language Acquisition*, 1, 89–100.
- Degen, J., & Goodman, N. (2014). Lost your marbles? the puzzle of dependent measures in experimental pragmatics. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 397–402). Austin, TX: Cognitive Science Society.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Dieuleveut, A., Chemla, E., & Spector, B. (2019). Distinctions between primary and secondary scalar implicatures. *Journal of Memory and Language*, 106, 150–171.
- Enguehard, É. (2018). Comparative modified numerals revisited: scalar implicatures, granularity and blindness to context. *Semantics and Linguistic Theory (SALT)*, 28, 21–39.
- Farkas, D. F., & Roelofsen, F. (2017). Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics*, 34(2), 237–289.
- Farkas, D. F., & Roelofsen, F. (2019). Polarity particles revisited. *Semantics and Pragmatics*, 12(15), 1–16.
- Fox, D. (2007). Free Choice Disjunction and the Theory of Scalar Implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Basingstoke: Palgrave Macmillan.
- Fox, D., & Hackl, M. (2007). The universal density of measurement. *Linguistics and Philosophy*, 29(5), 537–586.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998.
- Geurts, B. (2006). Take ‘five’: The meaning and use of a number word. In S. Voegeleer, L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311–329). Amsterdam: John Benjamins Publishing.
- Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25(1), 130–148.
- Geurts, B., & Nouwen, R. (2007). At least *et al.*: the semantics of scalar modifiers. *Language*, 83(3), 533–559.
- Geurts, B., & van der Slik, F. (2005). Monotonicity and processing load. *Journal of Semantics*, 22(1), 97–117.
- Grice, H. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Groenendijk, J. (2009). Inquisitive semantics: Two possibilities for disjunction. In P. Bosch, D. Gabelaia, & J. Lang (Eds.), *Seventh International Tbilisi Symposium on Language, Logic, and Computation* (pp. 80–94). Berlin: Springer.
- Hendriks, P. (2014). *Asymmetries between language production and comprehension*. Berlin: Springer.
- Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120(8), 1887–1897.
- Hendriks, P., & Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, 13(4), 319–348.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Ph. D. thesis, UCLA.
- Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58(3), 376–415.
- Jansen, C. J., & Pollmann, M. M. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics*, 8(3), 187–201.

- Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph. D. thesis, University of Edinburgh, Edinburgh.
- Keller, F. (2006). Linear Optimality Theory as a model of gradience in grammar. In G. Fanselow, C. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in grammar: Generative perspectives* (pp. 270–287). Oxford: Oxford University Press.
- Kennedy, C. (2015). A de-Fregean semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics*, 8(10), 1–44.
- Klecha, P. (2018). On unidirectionality in precisification. *Linguistics and Philosophy*, 41(1), 87–124.
- Krifka, M. (1996). Pragmatic strengthening in plural predications and donkey sentences. *Semantics and Linguistic Theory (SALT)*, 6, 136–153.
- Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and evidence in semantics* (pp. 109–132). Stanford: CSLI Publications.
- Križ, M., & Spector, B. (2020). Interpreting plural predication: Homogeneity and non-maximality. *Linguistics and Philosophy*. <https://doi.org/10.1007/s10988-020-09311-w>.
- Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 302–323). Berlin: De Gruyter.
- Marty, P., & Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in Psychology*, 4, 403.
- Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, 133, 152–163.
- Matsumoto, Y. (1995). The conversational condition on Horn scales. *Linguistics and Philosophy*, 18(1), 21–60.
- Mayr, C. (2013). Downward monotonicity in questions. *Proceedings of Sinn und Bedeutung*, 17, 345–362.
- Mayr, C., & Meyer, M.-C. (2014). More than *at least*. Paper presented at the “Two days at least” workshop, Utrecht University.
- Mendia, J. A. (2016). Reasoning with partial orders: Restrictions on ignorance inferences of superlative modifiers. *Semantics and Linguistic Theory (SALT)*, 26, 489–505.
- Meyer, M.-C. (2013). *Ignorance and grammar*. Ph. D. thesis, MIT.
- Mihoc, T. (2019). *Decomposing logic: Modified numerals, polarity, and exhaustification*. Ph. D. thesis, Harvard University.
- Mihoc, T. & Davidson, K. (2019). Experiments on the comprehensibility of comparative- vs. superlative-modified numerals under downward-entailing operators. Ms., Harvard University.
- Moracchini, S. (2018). Evaluativity and structural competition. *Semantics and Linguistic Theory (SALT)*, 28, 727–746.
- Nilsen, Ø. (2007). *At least*: Free choice and lowest utility. In R. Nouwen and J. Dotlačil (Eds.), *Proceedings of the ESSLI workshop on Quantifier Modification* (pp. 33–35). Dublin: Association for Logic, Language and Information.
- Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics*, 3(1), 1–41.
- Nouwen, R. (2015). Modified numerals: The epistemic effect. In L. Alonso-Ovalle & P. Menendez-Benito (Eds.), *Epistemic Indefinites* (pp. 244–266). Oxford: Oxford University Press.
- Nouwen, R., Alexandropoulou, S., & McNabb, Y. (2019). Experimental work on the semantics and pragmatics of modified numerals. In C. Cummins & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198791768.013.15>.
- Panizza, D., Chierchia, G., & Clifton, C, Jr. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*, 61(4), 503–518.
- Papfragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, 86(3), 253–282.
- Penka, D. (2014). *At most at last*. *Proceedings of Sinn und Bedeutung*, 19, 463–480.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reinhart, T. (2006). *Interface strategies*. Cambridge, MA: MIT Press.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3), 367–391.

- Schulz, K., & van Rooij, R. (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, 29, 205–250.
- Schwarz, B. (2016a). *At least* and ignorance: A reply to Coppock & Brochhagen (2013). *Semantics and Pragmatics*, 9(4), 1–23.
- Schwarz, B. (2016b). Consistency preservation in quantity implicature: The case of *at least*. *Semantics and Pragmatics*, 9(1), 1–47.
- Spector, B. (2013). Bare numerals and scalar implicatures. *Language and Linguistics Compass*, 7(5), 273–294.
- Spector, B. (2017). The pragmatics of plural predication: Homogeneity and non-maximality within the rational speech act model. In A. Cremers, T. van Gessel, and F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium* (pp. 435–444). Amsterdam: ILLC.
- van Rij, J., van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, 37(3), 731–766.
- van Rij, J., van Rijn, H., & Hendriks, P. (2013). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. *Topics in Cognitive Science*, 5(3), 564–580.
- van Rooij, I., Kwisthout, J., Blokpoel, M., Szymanik, J., Wareham, T., & Toni, I. (2011). Intentional communication: Computationally easy or difficult? *Frontiers in Human Neuroscience*, 5, 52.
- Van Tiel, B., & Schaeken, W. (2017). Processing conversational implicatures: alternatives and counterfactual reasoning. *Cognitive science*, 41, 1119–1154.
- Westera, M., & Brasoveanu, A. (2014). Ignorance in context: The interaction of modified numerals and QUDs. *Semantics and Linguistic Theory (SALT)*, 24, 414–431.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.