Semantics & Pragmatics Volume 15, Article 8, 2022
https://doi.org/10.3765/sp.15.8

This is an EARLY ACCESS version of

Sun, Chao & Richard Breheny. 2022. The role of alternatives in the interpretation of scalars and numbers: Insights insights from the inference task. *Semantics and Pragmatics* 15(8). https://doi.org/10.3765/sp.15.8.

This version will be replaced with the final typeset version in due course. Note that page numbers will change, so cite with caution.

EARLY ACCESS

The role of Alternatives in the interpretation of scalars and numbers – Insights from the inference task*

Chao Sun
Leibniz-Centre General Linguistics (ZAS)

Richard Breheny
University College London

Abstract Numerical noun phrases (NNPs) are ambiguous between *at least two* and *exactly two* interpretations. This ambiguity has been commonly discussed as a case of Scalar Implicature (SI), where meaning is enriched by exclusion of an alternative. But the SI approach to NNPs has also been widely challenged. We tested NNPs and scalar expressions in inference tasks and found that scalars were sensitive to a manipulation that altered the relevance of alternatives, whereas *exactly* readings for NNPs were not. Our findings provide a theory-critical challenge to the SI view of NNPs and support alternative views.

Keywords: numerical noun phrases, scalar implicature, alternative, inference task

1 Introduction

The ambiguity of noun phrases containing numerals has long been a focus of semantic/pragmatic research. Consider the following sentences with numerical noun phrases (NNPs). (1a) often carries the implication that Mary has no more than two children, whilst this implication does not always arise, as in the antecedent of the conditional in (1b). Instead, (1b) implies that if Mary has two or more children, she can claim tax relief. The interpretations of the NNP in (1a) and (1b) are often referred to as the *exactly* reading and the *at least* reading, respectively.

(1) a. Mary has two children.
    b. If Mary has two children, she can claim tax relief.

In the theoretical literature, different analyses have been proposed to account for the two interpretations of NNPs. According to exhaustification views, *at least*

* We thank the anonymous reviewer and the editor, Kristen Syrett, for valuable comments on an earlier draft; Blanche Gonzales de Linares, Erying Qin, Yike Fu, Peixin Yu, who worked in our lab. We also thank the audience at Sinn und Bedeutung 26 for helpful comments and discussion. This research was supported by the LeverhulmeTrust grant RPG-2018-425. Chao Sun was supported by Alexander von Humboldt Fellowship.

readings are basic readings, and *exactly* readings of NNPs include also scalar implicatures (SIs), which are derived by excluding Alternatives (Horn 1972, Singh 2019). Standard SI theory explains how (2a) often carries the implication in (2c), as a result of the exclusion of the Alternative in (2b) which is derived by lexical substitution of *some* with *all*. The conditions under which a given alternative is excluded are widely thought to involve contextual relevance (Fox & Katzir 2011, Geurts 2010). For example, when an alternative like (2b) would address a contextually salient question, intuition suggests it would normally be excluded. Alternatively, a context for (2a) may simply require an answer to the question whether some students passed, in which case, the sense in which (2a) implies (2c) is less keenly felt.

(2) a. Some of the students passed.
 b. All of the students passed.
 c. Not all of the students passed.

Extending the standard account of *some* to NNPs as in (1a), the Alternative, *Mary has three children*, is determined by lexical replacement of *two* with a higher value, *three* (Fox & Katzir 2011). The *exactly two* reading is then explained by conjoining the basic, *at least two* reading and the negation of the Alternative.

Many researchers argue against the Scalar Implicature (SI) view of NNPs and suggest that their interpretation should be explained by other means (Breheny 2008, Horn 1992). For instance, it has been proposed that the ambiguity in NNPs is explained by the optional application of a maximality operator (Buccola & Spector 2016, Kennedy 2015). Alternatively, it has been proposed that the basic meaning of NNPs is the upper-bounded, *exactly* reading, while the lower-bounded, *at least* reading is derived from mechanisms such as semantic (Geurts 2006) or pragmatic (Breheny 2008) coercion. Taken together we refer to these analyses as the non-SI view of NNPs. These views have in common the idea that *exactly* readings are not derived by excluding alternatives. Moreover, at least some of these views propose that the *exactly* reading is the primary, or biased reading. If that is correct, then the relevance of a higher number in context should have less impact on which meaning is derived.

The standard SI view of NNPs suggests that they should give rise to the same experimental outcomes as other scalar terms in some standard tests. However, evidence from different paradigms shows that NNPs behave differently from scalar terms. Using a dual task methodology, De Neys & Schaeken (2007) had demonstrated that fewer SIs for *some* are derived in a verification task when a secondary task places participants under memory load (e.g., memorising geometric patterns). Marty et al. (2013) tested how NNPs and scalar term *some* are interpreted using the same approach. They replicated De Neys & Schaeken's finding for *some* but found that the effect of cognitive load for NNPs is the opposite of that for *some*. More recently, using a truth-value judgment task, Dieuleveut et al. (2019) investigated 'primary scalar implicatures' which theorists argue are a stage in the derivation of SIs. They found evidence for primary implicatures being derived in the case of *some*, but not NNPs. Child language data has also been understood to challenge the SI view of NNPs. Children interpret NNPs with *exactly* readings more reliably than other scalar expressions (Huang et al. 2013, Hurewitz et al. 2006, Papafragou & Musolino 2003).

In short, there is already some evidence that NNPs and scalar terms exhibit different characteristics, and this evidence runs contrary to the view that *exactly* readings of NNPs are derived via SI. However, the empirical record does not directly bear on the main points of theoretical difference, which is whether ambiguity in NNPs results from optional exclusion of Alternatives. As recognised in Marty et al. 2013 and Dieuleveut et al. 2019, an alternative SI view of NNPs could account for the existing data by suggesting that numerals are lexically focused and trigger *exactly* readings by default, explaining the results in Dieuleveut et al. 2019. Also in that case, it could be that cancelling *exactly* readings is cognitively demanding, and this could explain why fewer *at least* readings are provided under working memory load in Marty et al. 2013. With respect to language development, Barner & Bachrach (2010) showed that children can derive inferences similar to SIs when alternatives are specified contextually. The developmental difference in interpreting NNPs and *some* could then be attributed to differences in the accessibility of alternatives.

In this paper, we present experimental research which addresses this question. The paradigm we use has been developed out of insights that have been reported about the inference task. Unlike verification tasks, the inference task elicits judgements from participants about whether certain implications follow from a statement. In particular, inference task research to date has targeted implications like those in (2c). Before we outline the main features of our study, we summarise in the next section key features of previous inference task research.

2 Inference task paradigm

Figure 1 shows an example of a standard inference task (adapted from Van Tiel et al. 2016). Participants read about a de-contextualised utterance involving a scalar term and have to decide whether they could conclude that the speaker implied the negation of the Alternative. A 'Yes' response indicates that an SI is drawn. Geurts & Pouscoulous (2009) measured the rates of SIs triggered by *some* using both the inference task and a sentence-picture verification task. In their verification task, participants were presented with a picture in which the sentence containing *some* was false when interpreted with SIs. They found that the inference paradigm yielded higher rates of SIs than the sentence-picture verification paradigm (62% vs. 34%).

Mary says:

Some of the questions are easy.

Would you conclude from this that, according to Mary,

not all of the questions are easy?

☐ Yes ☐ No

Figure 1 Experimental item used in the 'not Alt' condition. 'Yes' response is indicative of an SI response.

The derivation of SIs we described above provides several explanations for the inflated rate in the standard inference task. One is simply that the probe in the inference task mentions the alternative, and this must raise its salience for implicature derivation (Rees & Bott 2018). In addition, the probe question can change the context in which the sentence containing the scalar term is understood (Geurts & Pouscoulous 2009). In particular, by asking whether you would conclude that the Alternative is false, the stimulus is suggestive that the question of whether or not the Alternative is true might be relevant to the speaker's utterance. In the absence of any other information about the context of the speaker's utterance, this may bias participants to respond 'Yes'. If this latter explanation is on the right track, we hypothesise that if the probe question suggests that the Alternative is not contextually relevant, this should affect the rates of SIs in an inference task. Also, to the extent that such a manipulation is effective, it will allow us to test different approaches to NNPs. The SI view of NNPs predicts that inferences for NNPs should pattern like SIs under each probe. By contrast, the non-SI view of NNPs predicts that inferences for NNPs should be less susceptible to the manipulation of the probe than SIs.

3 The current study

Using a block design, we introduced two types of probe question. The 'not Alt' probe is the probe question used in the standard inference task, and a new 'could Alt' probe, which is shown in Figure 2 . The 'could Alt' probe asks participants whether they would conclude that it could be that the speaker thinks the Alternative is true. In other words, for items with *some*, the probe asks if they could attribute to the speaker the intention to communicate the literal 'some and possibly all' meaning. According to theoretical accounts of when scalar implicature is actually derived (mentioned above), a speaker uses *some* in this literal way only when the alternative is not relevant. Thus, in contrast to the 'not Alt' probe, if anything, the 'could Alt' probe would provide a cue to contexts in which the Alternative is not relevant and may reduce target responses as a result. Note that under the 'could Alt' probe, participants who derive the SI as part of the sentence meaning should give a 'No' response. In the following, responses associated with SI inferences are referred to as target responses. As mentioned above, the target response for the 'not Alt' probe is 'Yes'.

Mary says:

Some of the questions are easy.

Would you conclude that, it could be that Mary thinks,

all of the questions are easy?

☐ Yes ☐ No

Figure 2 Experimental item used in the 'could Alt' condition. 'No' response is indicative of an SI response.

In the current study, we tested two representative SI expressions, *some* and *possible*, and NNPs including *three*, *four* and *five*. The two SI expressions have shown near-ceiling rates of target responses in previous 'not Alt' inference tasks (e.g., 89% and 93% respectively, from Van Tiel et al. 2016). By contrast, NNPs have not been tested yet in the 'not Alt' paradigm. To present our predictions on response patterns, let's first consider how participants will respond to a sentence with a scalar expression like *some*. For either probe, we expect there to be three kinds of participant. We expect there to be participants who give target responses because they consider the SI relevant and are confident that the speaker's intended meaning includes the SI. There could also be participants who give non-target responses because they consider the SI irrelevant and are also confident about not inferring the SI as part of the intended meaning. A third group could infer both the relevant and irrelevant contexts for the SI and be uncertain about the intended meaning. How will this latter group respond in a binary judgment task? For these participants, the sentence with the SI expression is ambiguous and they would not feel in a position to conclude anything about what the speaker thinks. Given such uncertainty, we expect participants to be more likely to give 'No' response, irrespective of whether the probe is 'not Alt' or 'could Alt'. Notice that 'No' responses are target for the 'could Alt' probe and non-target for the 'not Alt'. This means that, if probe questions provide no hint about whether the implication is intended, if anything, we expect more target responses in the 'could Alt' condition compared to the 'not Alt' condition. However, if the probe questions suggest different contexts, then this information would help to resolve ambiguities. Particularly, in the 'not Alt' condition, participants would be more likely to accept SIs. Therefore, we may even find more target responses in the 'not Alt' condition compared to the 'could Alt' condition for standard scalar terms like *some* and *possible*. According to the standard SI theory of NNPs, the ambiguity between the *at least* reading and the *exactly* reading can be resolved by the relevance of alternatives. It predicts that we should get the same response pattern for NNPs as for other scalars.

4 Experiment

4.1 Participants

40 adult participants (29 females and 11 males) were recruited from Prolific Academic and were paid £1.4 for their participation. All participants reported English as a native language and were naïve to the purpose of the experiment. This experiment was approved by the local research ethics committee. Participants were provided with an electronic version of informed consent before taking part.

4.2 Materials and procedure

We tested three types of scales: the quantifier scale <some, all>, the modal scale <possible, certain>, and the numerical scale. Table 1 shows examples of target items for different scales.[1] For each scale, we constructed 6 target items and 12 control items.[2] Control items had the same structure as target items, but the responses of controls were either clearly 'Yes' or clearly 'No'.

We employed a within-subjects block design.[3] In one block participants responded to the 'not Alt' probe, and then in the other block they responded to the

1 The full list of items can be found in the supplementary file.

2 We followed Marty et al. and Dieuleveut et al. in using the non-partitive form for NNPs and the partitive form for *some*. One motivation for this choice is that using a partitive form for NNP items (e.g., Four of the chairs are in the room), as well as *some* items, may increase the chance that participants view numerals as more specific alternatives for *some* (Degen & Tanenhaus 2015). Nevertheless, a follow-up study, which did use the partitive form in NNPs, found the same results as the experiment reported below. Figures and analyses for this work are available at https://github.com/sunchaosc/number-inference

3 A pilot study with a between-subjects design gave qualitatively similar results.

'could Alt' probe. Each block contained 27 items, including 3 target items and 6 control items per scale. Four lists were created, each contained two blocks. Each

			not Alt probe	could Alt probe
<some, all>	Target	Some of the questions are easy.	not all of the questions are easy?	all of the questions are easy?
	Control-Yes	Some of the files have been deleted.	not none of the files have been deleted?	
	Control-No	Some of the plants have flowers.	none of the plants have flowers?	
<possible, certain >	Target	It is possible that the train will arrive on time.	it is not certain that the train will arrive on time?	it is certain that the train will arrive on time?
	Control-Yes	It is possible that Mary will attend the lecture.	it is not impossible that Mary will attend the lecture?	
	Control-No	It is possible that Peter will pass the exam.	It is impossible that Peter will pass the exam?	
<four, five >	Target	Four chairs are in the room.	it is not true that five chairs are in the room?	five chairs are in the room?
	Control-Yes	Three girls are dancing.	more than two girls are dancing?	
	Control-No	Five books are new.	no more than four books are new?	

Table 1 Examples of target and control items for different scales.

item only appeared once in each list, in one of the two blocks. The order of the blocks was counterbalanced over the lists.

Instructions and four practice trials were provided at the beginning of each block. The order of items was randomized for each participant in each block. An unrelated study, which took approximately 5 minutes, was inserted between the 'not Alt' and the 'could Alt' block to serve as a filler task.

4.3 Results

Five participants were removed because more than 20% of their answers to control items were incorrect. Figure 3 shows the percentages of 'Yes' responses for each scale and condition by probe type. The mean accuracy of the control items was 92% (control yes: 89%, control no: 96%).

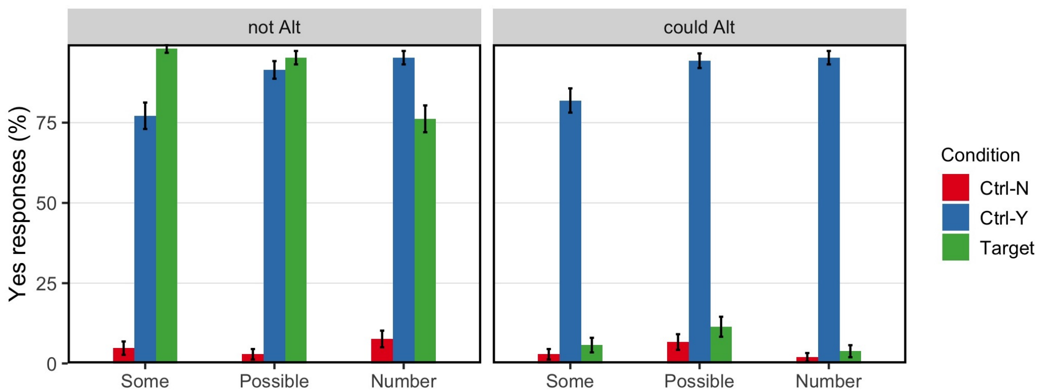

Figure 3 Percentages of 'Yes' responses for each scale and condition by probe type. Error bars represent standard errors of the mean.

We coded the 'Yes' response in the 'not Alt' condition and the 'No' response in the 'could Alt' condition as the target response. Figure 4 shows the percentages of target responses for each scale and probe type. We fitted a mixed effects logistic regression model predicting response (target or non-target) from probe type, scale, block order ('not Alt' first or 'could Alt' first), and their interactions, including random intercepts for participants.[4] Random slopes were dropped due to non-convergence or singularity.

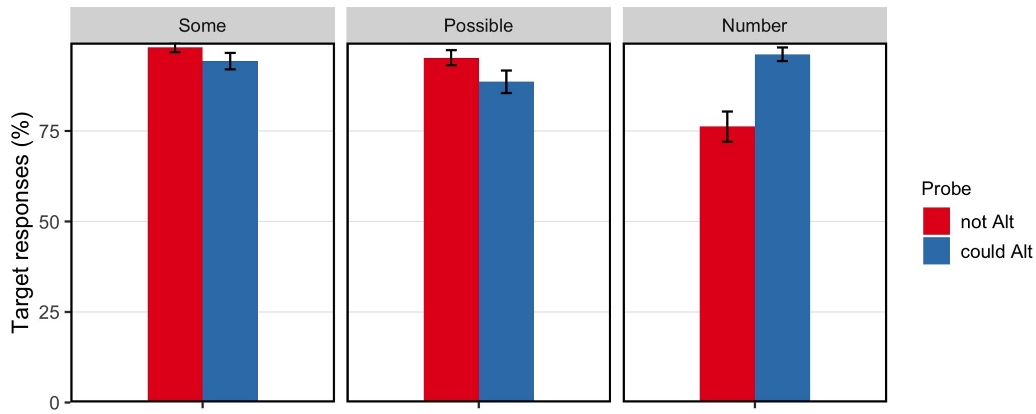

Figure 4 Percentages of target responses for each scale and probe type. Error bars represent standard errors.

There was a main effect of scale ($\chi^2(2) = 14.91$, $p < .001$) and a significant interaction between scale and probe type ($\chi^2(2) = 22.64$, $p < .001$). Planned comparisons within the levels of scale revealed that the probability of target responses was higher in the 'not Alt' condition compared to the 'could Alt' condition for *possible* ($\beta = -1.12$, SE = 0.63, $p = .06$), the same effect was also numerically present for *some* ($p = .16$), but it was reversed for NNPs: the probability of target responses was significantly higher in the 'could Alt' condition compared to the 'not Alt' condition ($\beta = 2.58$, SE = 0.66, $p < .001$). Planned comparisons within the levels of probe type revealed that the probability of target responses was higher for scalars than for NNPs in the 'not Alt' condition (*some*: $\beta = -3.17$, SE = 0.81, $p < .001$; *possible*: $\beta = -2.16$, SE = 0.57, $p < .001$). In contrast, in the 'could Alt' condition, the probability of target responses was higher for NNPs than for *possible* ($\beta = 1.35$, SE = 0.64, $p = .03$) and there was no difference between NNPs and *some* ($p = .50$).

There was also a significant three-way interaction between scale, probe type and block order ($\chi^2(2) = 8.51$, $p = .01$). Post-hoc analyses revealed that the interaction

4 Mixed-effect analyses were conducted in R (R Core Team 2020) using the 'lme4' package (Bates et al. 2015) and the 'lmerTest' package (Kuznetsova et al. 2017). Scale was dummy-coded, probe type and block order were deviation coded. Model comparisons were conducted to test the significance of fixed effects with more than two levels, using likelihood ratio tests. Significant interactions were followed up by conducting analyses on subsets of data defined by the levels of relevant factors.

between probe type and scale was stronger in the ‘could Alt’ first group ($\chi^2(2)$ = 17.76, $p < .001$) than that in the ‘not Alt’ first group ($\chi^2(2)$ = 13.44, p = .001).[5]

5 Discussion

Contrary to the predictions of the standard SI view, we found different outcomes for scalars and NNPs. Focusing on *some* and *possible*, the rates of target responses were high across conditions, which may be due to the presence of alternatives (*all*, *certain*) in both types of probes. Nevertheless, rates for ‘not Alt’ were greater than ‘could Alt’. This response pattern is predicted by both views, since the probes differ in the extent to which they suggest that the Alternative is relevant. Turning to NNPs, the reverse happened. Target response for ‘could Alt’ was greater than ‘not Alt’. This pattern was predicted if the probe type had no effect on ambiguity resolution. As this was the finding for NNPs, the results suggest that the relevance of putative alternatives did not help to resolve the ambiguity of NNPs. This finding is incompatible with the standard SI view of NNPs. However, it is in line with non-SI view that exclusion of Alternatives is not involved, or less involved, in the case of NNPs.[6]

As this is the first inference task study to report results for NNPs and other scalars together, it is worthwhile commenting further on the patterns of responses here to highlight how the results challenge SI views, even those beyond the standard one. Previous verification tasks show that *exactly* readings of NNPs are clearly more robust than SI readings for *some* (Papafragou & Musolino 2003, Dieuleveut et al. 2019, Huang et al. 2013). An alternative account of these results, mentioned above, suggests that NNPs do trigger scalar implicatures but do so in virtue of Alternatives

5 The interaction with block order may be due to ‘not Alt’ trials potentially priming participants to respond with more SIs in the following ‘could Alt’ block.

6 This discussion does not take into account the possibility that participants derive what is referred to a ‘Primary Implicature’ but not the scalar implicature itself (see Dieuleveut et al. 2019 for an explanation of Primary Implicatures and a discussion of when they can occur). Both intuition and experimental evidence suggests that a PI-not-SI inference is accessed in a very small minority of cases, unless there are fairly clear cues to speaker ignorance about the truth of a relevant Alternative and these are not present here (see Dieuleveut et al. 2019). In any case, the logic of our design allows us to set this aside as an option because a participant who only derives the PI will give the same ‘No’ response to reflect uncertainty in both ‘not Alt’ and ‘could Alt’ trials and because any cues to speaker’s knowledge about the Alternative are the same across all expression types. Thus, according to our null hypothesis where probe type has no influence on computing readings, PI-only participants will advantage target response rates in ‘could Alt’ trials over ‘not Alt’. If the proposal that ‘not Alt’ probes provide a better cue to a relevant context for PI-only or SI is correct, then any PI-only respondents will make this hypothesis more difficult to substantiate since they give non-target responses, despite employing Alternatives in their reading for the sentence. In light of our findings below, it makes the test for the effect of probe on response a conservative one and considerations.

being lexically focused and the SI being default (Dieuleveut et al. 2019, Singh 2019). More generally, the counterproposal could be that alternatives are always highly salient regardless of context. But here, in what is supposed to be the standard ('not Alt') inference task, NNPs gave rise to a clearly lower rate of target responses than *some* or *possible*. If the exclusion of the NNP Alternative is default, it is difficult to see how a probe which encourages exclusion would lower the target response.

We turn now to our account of the results for NNPs in the 'not Alt' condition. We explain the observed rate of non-target responses for NNPs as being mostly driven by uncertainty. To see this, note that, in the 'could Alt' condition, a non-target response would be based on the participant seeing the *at least* reading as clearly intended, and this rate was as low as 4%. In the 'not Alt' condition, we are assuming that the non-target response rate was the outcome of participants who only derive the *at least* reading and also those who give a 'No' response due to uncertainty. Given the apparent infrequency of *at least*-only readings, 'No' responses in 'not Alt' trials mainly result from uncertainty about which reading may have been intended. Therefore, for NNPs, *exactly* readings are available for most trials and are strongly biased compared to *at least* readings. However, a higher rate of unresolved ambiguity led to the lower rate of target responses in the 'not Alt' condition.

Finally, our results raise a question about the status of *at least* readings of NNPs on the non-SI view. The question is why deriving an *at least* reading does not lead to a scalar implicature via normal processes of exclusion. In other words, suppose that one parse of a target sentence has an *at least* reading. We are asking why a participant who considers that parse does not also consider it valid to derive an SI using a sentence with a higher number as alternative. One answer to this question is based on accounts which hold that the *exactly* reading is the single grammatically encoded meaning while the *at least* reading is a result of some subsequent coercion (semantic or pragmatic). For reasons that would need to be made clearer, it could be that only grammatically encoded meanings can enter into computations of exclusion by SI, precluding *at least* readings of NNPs from such a process.

6 Summary

The current study goes beyond previous findings by showing a theory-critical difference between NNPs and scalar expressions. The responses for *some* and *possible* are affected by the manipulation of probes in a way that is to be expected if their upper-bounded meaning is derived by exclusion of Alternatives. Responses to NNPs show a pattern that would be expected if the different probes had no effect on which reading becomes available for the participant. Our findings are consistent with the view *exactly* readings of NNPs are strongly biased but not derived as Scalar Implicature. Finally, our results substantiate previous insights about why standard inference tasks yield inflated rates for scalar expressions.

References

Barner, David & Asaf Bachrach. 2010. Inference and exact numerical representation in early language development. *Cognitive Psychology* 60(1). 40–62. https://doi.org/10.1016/j.cogpsych.2009.06.002.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2). 93–139. https://doi.org/10.1093/jos/ffm016.

Buccola, Brian & Benjamin Spector. 2016. Modified numerals and maximality. *Linguistics and Philosophy* 39(3). 151–199. https://doi.org/10.1007/s10988-016-9187-2.

De Neys, Wim & Walter Schaeken. 2007. When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental psychology* 54(2). 128–133. https://doi.org/10.1027/1618-3169.54.2.128.

Degen, Judith & Michael K. Tanenhaus. 2015. Processing scalar implicature: A constraint-based approach. *Cognitive Science* 39(4). 667–710. https://doi.org/10.1111/cogs.12171.

Dieuleveut, Anouk, Emmanuel Chemla & Benjamin Spector. 2019. Distinctions between primary and secondary scalar implicatures. *Journal of Memory and Language* 106. 150–171. https://doi.org/10.1016/j.jml.2019.02.008.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107. https://doi.org/10.1007/s11050-010-9065-3.

Geurts, Bart. 2006. Take 'Five': The meaning and use of a number word. *Non-definiteness and Plurality* 95. 311–329. https://doi.org/10.1075/la.95.16geu.

Geurts, Bart. 2010. *Quantity implicatures*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511975158.

Geurts, Bart & Nausicaa Pouscoulous. 2009. Embedded implicatures?!? *Semantics and Pragmatics* 2(4). 1–34. https://doi.org/10.3765/sp.2.4.

Horn, Laurence R. 1972. *On the Semantic Properties of Logical Operators in English*. Los Angeles, CA: University of California Los Angeles Ph. d. dissertation.

Horn, Laurence R. 1992. The Said and the Unsaid. In *Semantics and linguistic theory 2*, vol. 2, 163–192. Columbus. https://doi.org/10.3765/salt.v2i0.3039.

Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What Exactly do Numbers Mean? *Language Learning and Development* 9(2). 105–129. https://doi.org/10.

1080/15475441.2012.658731.

Hurewitz, Felicia, Anna Papafragou, Lila Gleitman & Rochel Gelman. 2006. Asymmetries in the Acquisition of Numbers and Quantifiers. *Language Learning and Development* 2(2). 77–96. https://doi.org/10.1207/s15473341lld0202_1.

Kennedy, Christopher. 2015. A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8(10). 1–44. https://doi.org/10.3765/sp.8.10.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. https://doi.org/10.18637/jss.v082.i13.

Marty, Paul, Emmanuel Chemla & Benjamin Spector. 2013. Interpreting numerals and scalar items under memory load. *Lingua* 133. 152–163. https://doi.org/10.1016/j.lingua.2013.03.006.

Papafragou, Anna & Julien Musolino. 2003. Scalar implicatures: experiments at the semantics– pragmatics interface. *Cognition* 86(3). 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. https://www.r-project.org/.

Rees, Alice & Lewis Bott. 2018. The role of alternative salience in the derivation of scalar implicatures. *Cognition* 176. 1–14. https://doi.org/10.1016/j.cognition.2018.02.024.

Singh, Raj. 2019. Context, Content, and the Occasional Costs of Implicature Computation. *Frontiers in Psychology* 10. 2214. https://doi.org/10.3389/fpsyg.2019.02214. https://www.frontiersin.org/article/10.3389/fpsyg.2019.02214/full.

Van Tiel, Bob, Emiel Van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175. https://doi.org/10.1093/jos/ffu017.

Chao Sun
Schützenstr. 18
10117, Berlin
c.sun@leibniz-zas.de

Richard Breheny
Chandler House, 2 Wakefield Street
London WC1N 1PF
r.breheny@ucl.ac.uk