

What the inference task can tell us about the comprehension of scalars and numbers: An investigation of probe question and response bias

Chao Sun^{1,2} & Richard Breheny²

¹Leibniz Centre General Linguistics ZAS ²University College London

It is widely assumed that scalar implicature (**SI**) is derived by excluding Alternatives [1,2]. Thus an implication in (1b), of (1a), is understood to be the result of computing the Alternative (**Alt**) in (1c) and negating that. The class of expressions that routinely give rise to SI is thought to be broad, including quantifiers, modals, adjectives and many more. However, studies using the inference task (see Fig. 1a), has revealed that SI emerges more robustly for some expression types (quantifiers, modals) than others (adjectives) [3]. The causes of this Scalar Diversity (**SD**) effect are not yet fully understood. To date no study involving the inference task has included numerical noun phrases (**NNPs** – e.g. (2a)). There is controversy over whether the routinely available implication of (2a) shown in (2b) should be explained as the exclusion of (2c), or by other means [4-6]. The view where (2b) is not an SI of (2a) holds that, where (2b) is not available, this second *at least* reading is derived as a second, perhaps less dominant meaning of the NNP [7]. To date, a limited amount of experimental work provides some support for this non-SI view of NNPs, [8, 9]. Our aim is to test competing theories of NNP by developing some insights about the inference task, which can also shed light on SD.

In the standard inference task, participants are presented with a de-contextualised utterance involving a scalar term (e.g. ‘some’) and asked whether they would conclude that the Alt is excluded (e.g. ‘not all’). The results in [3] show that for quantifiers (‘some’, ‘sometimes’) and modals (‘might’, ‘possible’) the rates of ‘yes’ responses are very high (over 80%). Our claim is that, in the absence of the actual context, the task question suggests that the exclusion of the Alt is a relevant inference and this suggests a context in which Alt is relevant, biasing a ‘yes’ response. Another way to probe is to ask whether it could be that the speaker thinks Alt is true. Acceptance here would be based on the participant assuming that no SI was conveyed. And here the probe draws participants’ attention to contexts where the Alt is not relevant and no SI is conveyed.

Experiments 1a and 1b differ in the probe (see Figs. 1a,b). We note that for standard ‘not Alt’ probes (Exp. 1a), a ‘yes’ response unequivocally signals the participant is confident the SI would have been intended. Participants who say ‘no’ do so because they fail to see a clear SI meaning. This could be because they only access the un-enriched meaning, or because they can see SI and also other meanings and are not confident the SI is intended. For the ‘could Alt’ probe, a ‘no’ response can be given either because the participant is confident there is an SI, but also because they can see the un-enriched meaning (or other inferences) as possible but are not confident the unenriched meaning is intended. Thus if the probes in figures 1a and 1b had no influence in suggesting contexts to participants, rates of ‘no’ for ‘could Alt’ should be at least as high as rates of ‘yes’ for ‘not Alt’. However, if our hypothesis about the role of probe in suggesting context is correct, then rates of ‘no’ for ‘could Alt’ are not likely to be higher and in fact, rates of ‘yes’ in not Alt may be higher than ‘no’ for ‘could Alt’. This is our prediction for scalar expressions ‘some’ and ‘possible’ (used in Exp. 1).

As for NNPs, according to the SI theory of NNPs, we should get the same pattern as for other scalars. However, if the two readings of NNPs typically result from some form of ambiguity and not from exclusion of Alts, then the probe in the not-Alt inference task will not have the same effect as for ‘some’ and ‘possible’. Rates of ‘yes’ response in the not-Alt case will reflect the extent to which participants only access the anti-monotone, *exactly* reading. If they access only the monotone, *at least* reading or both readings, they should not feel that the conclusion can be drawn. By contrast, in the could-Alt study (Expt. 1b), participants should

accept if they access only the *at least* meaning. As it is widely assumed that the *exactly* meaning of NNPs is dominant, we predict that rates of target response should increase in the could Alt case.

Methods: We examined the effect of probes across three types of scales (quantifier, modal, NNP) in a within-subject design. 40 participants performed two blocks of 27 trials, with an unrelated experiment between them. In one block participants responded to the ‘not alt’ probe (Fig. 1a), and in the other block they responded to the ‘could alt’ probe (Fig. 1b). The order of the blocks was counterbalanced across participants. Each block contained 9 experimental items (3 per scale) and 18 control items. The controls had the same structure as the experimental items, but their responses were clearly ‘yes’ or clearly ‘no’. *Results:* See Fig. 2. A ‘yes’ response for ‘not alt’ or a ‘no’ response for ‘could alt’ was coded as the target response. We fitted a mixed effects logistic regression model predicting target response from probe type, scale and block order. There was a significant interaction between probe type and scale ($p < .001$). The rate of ‘possible’ was marginally higher for ‘not alt’ than for ‘could alt’ ($p = .059$); the same effect was numerically present for ‘some’. By contrast, the SI rate of NNPs was significantly higher for ‘could alt’ than for ‘not alt’ ($p < .001$). There was also a significant three-way interaction ($p = .02$). See Fig. 3. It revealed that the interaction between probe type and scale was stronger in the group of participants who faced the ‘could alt’ block of trials first (than that in the ‘not alt’ first group). The effect of block seems to be most marked for ‘possible’ where participants who understood ‘not-Alt’ probes first seem to have been primed to respond with SI in the second ‘could Alt’ block. While those who undergo ‘could-Alt’ trials first have lower target rates.

Discussion: 1. Our data adds to the body of evidence suggesting that inferences for NNPs are not the same as other scalars. Specifically, we provide new evidence that this difference is because exclusion of Alternatives is not involved, or less involved, in the case of NNPs.

2. The not-Alt task has been shown to have an effect on boosting the rates of SIs [10]. Here we provide one insight as to why, in terms of the probe suggesting a relevant context.

3. Taken together, these conclusions suggest a way of exploring the SD effect in [3], where many expressions give quite low target rates for not-Alt probes. Given that NNPs are not/less affected by not-ALT probes, it may be the case that some of the scalar terms in [3] are like NNPs in the way that they are less affected by contexts suggested from the probe and have a bias toward an anti-monotone or a monotone reading. For NNPs, the bias is strongly towards the anti-monotone reading, whereas for many low rating adjectives it might be towards the monotone reading. If more participants see only the monotone reading or both readings, the rates of ‘yes’ for these adjectives will be low in the not-ALT task in [3].

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. a. Some of the students passed. <li style="padding-left: 20px;">b. Not all of the students passed. <li style="padding-left: 20px;">c. All of the students passed. | <ol style="list-style-type: none"> 2. a. Three of the students passed. <li style="padding-left: 20px;">b. No more than three of the students passed. <li style="padding-left: 20px;">c. Four of the students passed. |
|--|---|

<div style="border: 1px solid black; padding: 10px;"> <p>Mary says: <i>Four chairs are in the room.</i></p> <p>Would you conclude from this that, according to Mary, <i>it is not true that five chairs are in the room?</i></p> <p style="text-align: center;">1: Yes 2: No</p> </div> <p>(a)</p>	<div style="border: 1px solid black; padding: 10px;"> <p>Mary says: <i>Four chairs are in the room.</i></p> <p>Would you conclude that, it could be that Mary thinks, <i>five chairs are in the room?</i></p> <p style="text-align: center;">1: Yes 2: No</p> </div> <p>(b)</p>
---	--

Fig. 1 Example items.

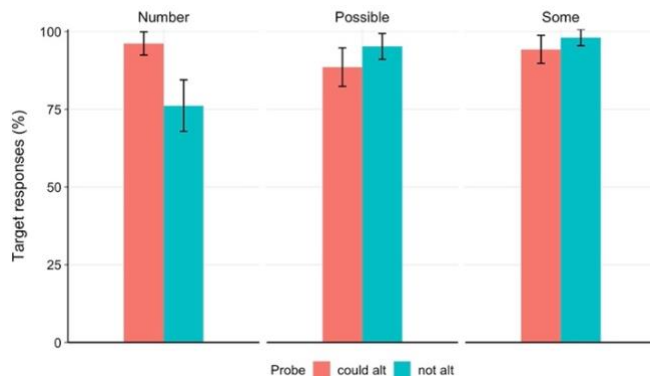


Fig. 2 Percentage of target responses for each scalar word by probe type

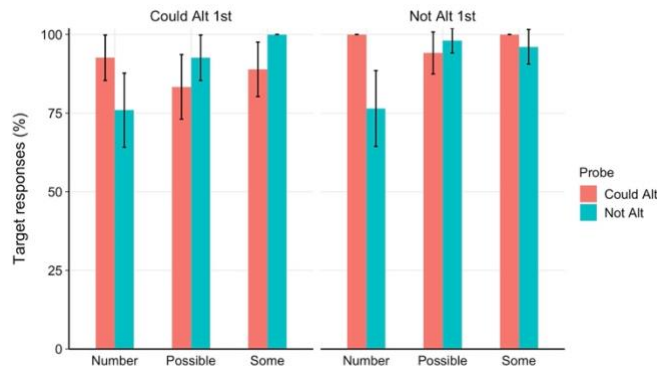


Fig. 3 Percentage of target responses for each scalar word by probe type and block order

[1] Geurts, B. (2010). *Quantity Implicature*. OUP; [2] Chierchia, G., Fox, D., & Spector, B. (2012). In P. Portner, C. Maienborn, & K. von Stechow (Eds.), *Semantics: An international handbook of natural language meaning*. [3] van Tiel, B. et al. (2016). *J. of Semantics* [4-6] Horn, L. (1992). The said and the unsaid. *SALT II*; Geurts, B. (2006). Take five; Kennedy, C. (2015) *S&P*; [7] Buccola, B. & Spector, B. (2016). *Ling. & Philos.* [8] Marty, P. et al. (2013). *Lingua* [9] Dieuleveut, A. et al. (2019). *JML* [10] Geurts, B. & Pouscoulous, N. (2009). *S&P*.