What the inference task tells us about numbers, and what numbers can tell us about the inference task

It is widely agreed that scalar implicature (**SI**) is prevalent across constructions and languages. SIs are widely thought to be 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, a recently developed paradigm, the inference task, has revealed that SI emerges more robustly for some expression types (quantifiers, modals) than others (adjectives) [3]. This is the Scalar Diversity (**SD**) effect and its cause is 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 SI, the exclusion of (2c), or whether NNPs simply encode (2b) as an entailment of (2a) [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 secondary, less dominant meaning of the NNP [7]. To date, a limited amount of work provides some support for this non-SI view of NNPs, [8]. 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, >80%. Our claim is that, in the absence of the actual context, participants use the task question ('would you conclude not Alt?') to infer something about the context. By asking about Alt, the task question suggests Alt is relevant and this licences the SI, biasing a 'yes' response. Another way to probe for SI is to ask whether, according to the speaker, Alt could be true. A rejection here would be based on the participant assuming that the SI was conveyed. Experiments 1a and 1b differ in the probe (see Figs. 1a,b). For scalar expressions 'some' and 'possible', our hypothesis is that the 'could Alt' probe will yield lower SI responses than the standard 'not Alt' inference task because the stimulus does not itself bias the SI. As for NNPs, according to the SI theory of NNPs, we should get the same pattern as for other scalars: higher rates with not Alt probe than could Alt. However, if the two readings of NNPs result from some form of ambiguity and not from SI, then the probe in the not Alt inference task will not have the same effect as for scalars. Rates of 'yes' response in the not Alt case will reflect the extent to which participants only access the *exactly* reading. If they access only the *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 'implicature' response should increase in the could Alt case.

Methods: We examined the effect of probes across four types of scales (quantifier, modal, adjective, NNP) by manipulating probe type between subjects. In the 'not alt' condition, as shown in fig.1a, a 'Yes' response indicates that an SI reading has been drawn; whereas in the 'could alt' condition, as shown in fig.1b, a 'No' response is compatible with an SI reading. Filler items have the same structure as the experimental items, but the response is independent of whether an SI has been drawn. *Results*: See Fig.2. We constructed a mixed effects logistic regression model predicting response from probe type (could ALT or not ALT). For 'some', 'possible' and 'warm', the percentages of implicature responses were higher in the 'not alt' condition than in the 'could alt' condition (p=.002; p=.01; p=.02); Whereas for number items, this percentage was marginally higher in the 'could alt' condition (p=.08). These results suggest that SI is not involved in deriving readings of NNPs. They also point to an explanation for why some other scales do not have high rates of 'yes' response in the standard inference task. Though caution is needed here since 'warm' patterns with 'some' and 'modals' but has comparable not_Alt rates as NNPs.

1. a. Some of the students passed.
   b. Not all of the students passed.
   c. All of the students passed.

2. a. Three of the students passed.
   b. No more than three of the students passed.
   c. Four of the students passed.

(a)

Mary says:

Four chairs are in the room.

Would you conclude from this that, according to Mary,

it is not true that five chairs are in the room?

1:Yes 2:No

(b)

Mary says:

Four chairs are in the room.

Would you conclude from this that, according to Mary,

it could be that five chairs are in the room?

1:Yes 2:No

Fig.1 Example item

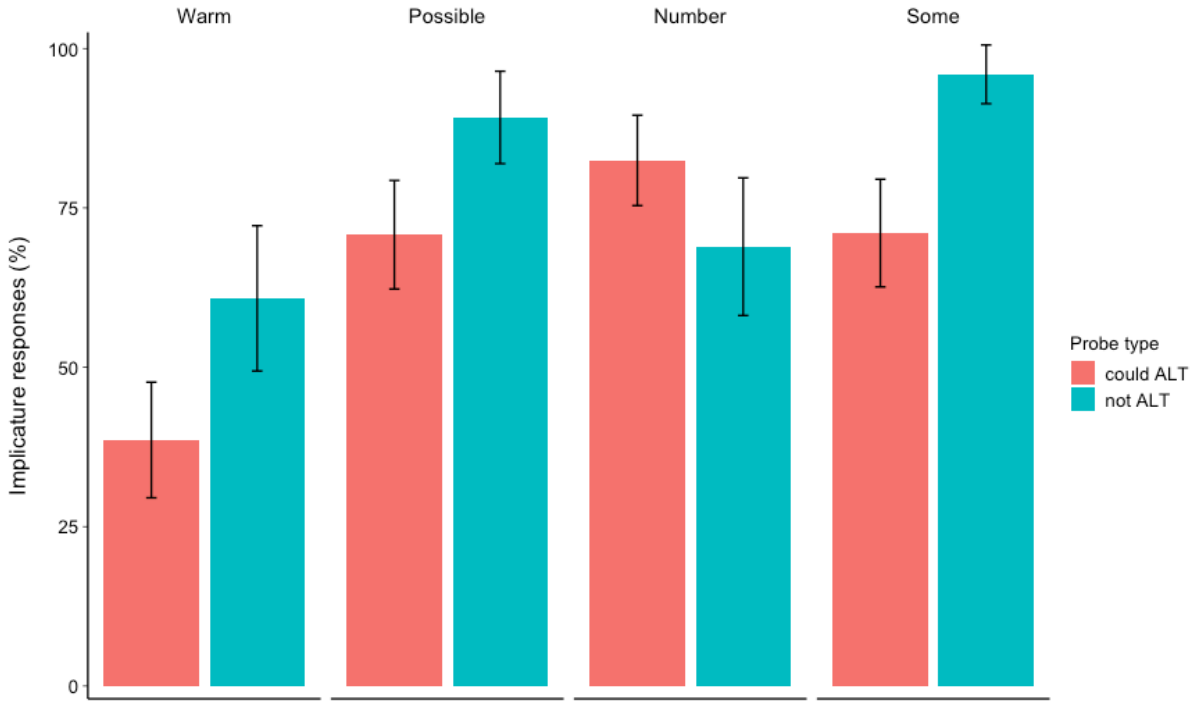

Fig.2 Percentage of implicature responses for each scalar word by probe type

References: **[1]** Geurts, (2010). *Quantity Implicataure*. OUP; **[2]** Chierchia Fox & Spector, (2012). In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *An international handbook of natural language meaning*. **[3]** van Tiel, B. Miltenburg, E. van, Zevakhina, N., Geurts, B. (2016). *J. of Semantics*, 33 **[4-6]** Horn, L. (1992). The said and the unsaid. *SALT II*; Geurts, B. (2006). Take five; Breheny, R. (2008). *J. of Semantics*. **[7]** Spector, B. (2013). *Language and Linguistics Compass*, *7*. **[8]** Marty P., Chemla, E., & Spector, B. (2013). *Lingua*, *133*