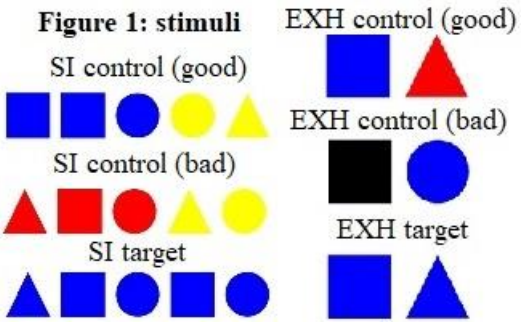
Computing implicatures under QUDs

Eszter Ronai & Ming Xiang (The University of Chicago)

Introduction: Implicatures serve as an important testing ground for examining the time-course of integrating semantic and pragmatic information. Starting with Bott & Noveck (2004), several studies have found that the computation of scalar inferences is costly, i.e. it causes longer reaction times. Recently, however, van Tiel & Schaeken (2017) found processing cost for computing scalar inference, but not for *it*-cleft exhaustivity inference. They propose that the time it takes to compute inferences is not uniform, and it depends on the structural characteristics of the required alternatives (following Chemla & Bott, 2014; Katzir, 2007). Processing costs occur only if the lexicon needs to be accessed to construct the relevant alternatives. According to van Tiel & Schaeken, in scalar inference, *all* needs to be retrieved to compute *some but not all*, but in *it*-cleft exhaustivity, there are no (lexical) alternatives.

However, a potential problem is that van Tiel & Schaeken presented target sentences to participants in the absence of any context. Previous research (i.a. Zondervan, et al., 2008) has shown that context affects how likely a scalar inference is to arise, and we might predict that this effect extends to other inferences and speed of computation as well. In this paper we compare scalar inferences (SI) to *it*-cleft exhaustivity (EXH), embedded under Questions Under Discussion (QUDs) that were empirically elicited. Our data show that QUDs modulate calculation rates and processing speed. Importantly, we find similar patterns across the two inference types: for both SI and EXH, under QUDs that bias towards calculating the inference, there is no increase in reaction times, but under QUDs that bias against calculating the inference we observe longer reaction times. Our results, therefore, found strong effect of QUDs, instead of the characteristics of alternatives, on the computation of SI and EXH inferences.

Experiments: The target sentences investigated were e.g. SI: *Some of the shapes are blue* (intended inference: *Not all of the shapes are blue*) and EXH: *It is the square that is blue* (inference: *Only the square is blue*). In **Exp.1**, we elicited QUDs in the following way. Participants were given a background story: Anne is asking questions from Bob, about pictures that only Bob can see. Participants then saw SI and EXH target sentences paired with pictures (see Fig.1). They were told the sentences they saw were Bob's answers to Anne's questions, and they need to guess what Anne's questions were. **Exp.2** is a sentence-picture verification task. Participants saw a dialogue between Anne and Bob, together with a picture (see Fig.1). They had to make a binary judgment about whether Bob gave a good answer to Anne's question, given the picture he saw. Anne's questions are always the most frequent questions elicited from Exp.1. There are three kinds for the SI stimuli: What color are the shapes?/Are any shapes blue?/Are all shapes blue?; and three kinds for the EXH stimuli: Which shape is blue?/Are there any blue shapes?/Are both shapes blue?. Bob's answers (e.g. *Some of the shapes are blue*, *It is the square that is blue*) were either unambiguously good/bad descriptions of the Control pictures, or were good descriptions on their literal, but not on their inference-enriched reading of the Target pictures. Thus if a participant says Bob gave a "good" description of a Target picture, she has not calculated the SI/EXH inference. Saying "not good" indicates inference calculation. The Picture manipulation was done within-, and the Question manipulation between-participants. Participants' responses, as well as reaction time (RT) were recorded.



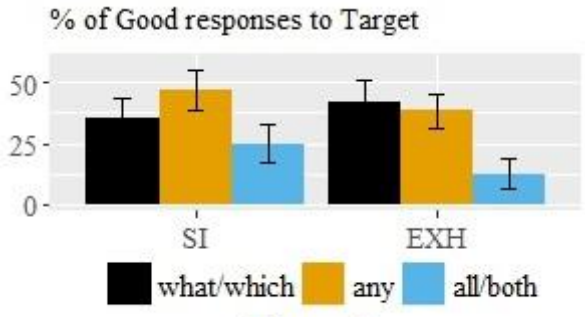


Figure 2

Results: Fig.2 plots the percentage of "Good" responses to Target: higher percentages indicate a lower rate of inference calculation. For SI, any QUDs resulted in fewer inferences calculated than all

($p<0.001$) or what ($p<0.05$). For EXH, both QUDs resulted in more inferences calculated than any ($p<0.001$) or which ($p<0.001$), suggesting that any and which are Literal-biasing, while all, what and both are Inference-biasing QUDs. These differences in question type predict speed of processing. Fig.3-4 plot reaction times ("good": G, "not good": NG), broken down by inference type and question type. In line with previous literature, we take longer RT when responding **NG to Target**, relative to the RT when responding **NG to Control**, to be what indexes the cost of implicature calculation. In SI, we find a significant interaction of Question and Response ($p<0.01$) such that with Literal-biasing QUDs (*any* questions), making an SI-enriched judgement takes longer than responding to the relevant literal control. I.e., responding NG to Target takes longer than responding NG to Control. With Inference-biasing QUDs (*all* and *what*), there is no such difference. These findings strongly suggest that SI computation is only costly when preceded by non-supportive QUDs. For EXH, the results are more nuanced. For the Literal-biasing *any*-QUD, we also see a cost for inference computation (response RT for NG, target vs. control, $p<0.05$). But for the seemingly equally Literal-biasing *which*-QUD, we saw similar but not exactly the same pattern. For the Inference-biasing *both*-QUD, there was an unexpected cost for responding NG to the Control picture, which is likely a side-effect of the picture stimuli, and we are conducting a follow up to address that. Overall, for both of SI/EXH, QUDs that bias towards literal interpretation incur a cost for inference computation. But inference computation is not costly under a congruent QUD.

Figure 3 SI: mean reaction times

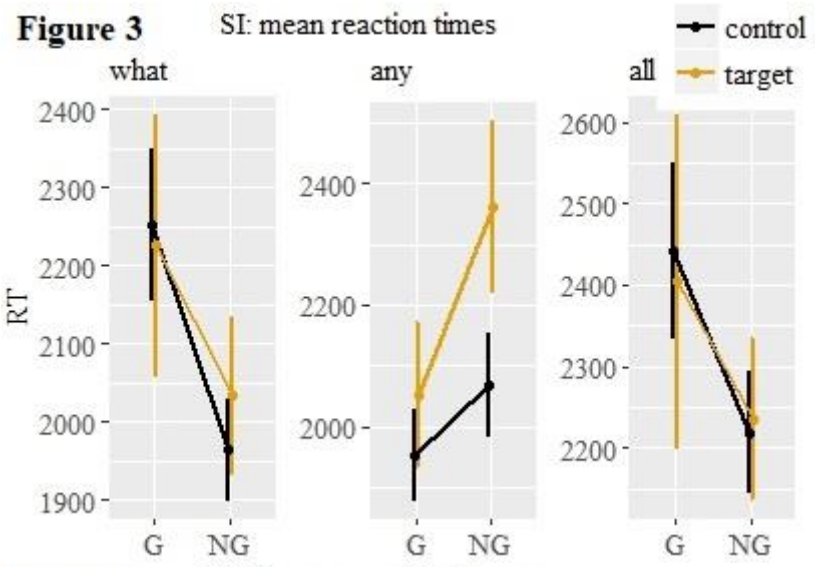
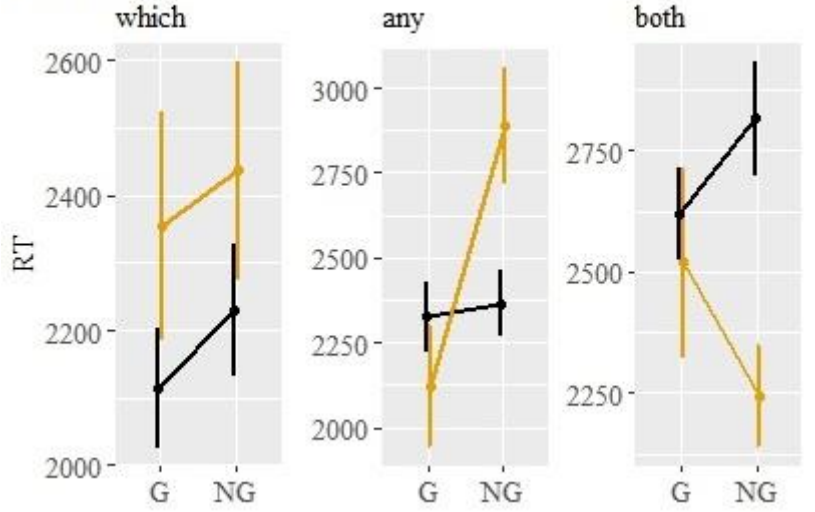


Figure 4 EXH: mean reaction times



Discussion and Conclusion: Van Tiel & Schaeken show that SI calculation incurs a cost (as has been known from earlier literature as well), but EXH calculation does not. Both pragmatic inferences involve the hearer's counterfactual reasoning about what the speaker could have said, but did not say. If it was this reasoning process that was costly, then both SI and EXH should incur a cost. However, only SI involves accessing the lexicon during the reasoning process, to retrieve *all* and construct the relevant alternative the speaker could have said (*All of the shapes are blue*). Therefore, the observed asymmetry in processing cost was argued to be in line with the hypothesis that the cost of computing pragmatic inferences is largely triggered by the computation involved in constructing alternatives. However, we present novel evidence showing that for both SI and EXH inferences, calculation rates and processing cost are strongly modulated by QUDs. We find not that SI is always costly and EXH is always cost-free; rather that both can be costly or cost-free depending on the context they occur in. We suggest that a QUD-based account better explains the current findings than an alternative-based account.