Probability and processing speed of scalar inferences is context-dependent

The past two decades have seen a wealth of studies addressing the question of whether or not scalar inferences -- whereby a listener takes a sentence like *Alex ate some of the cookies* to mean that he did not eat all of them -- generally incur a processing cost, with conflicting results [1-5]. This has spurred the development of studies seeking to understand the contextual conditions that facilitate scalar inferences [6-10]. Here, we test a prediction made by [11]'s constraint-based account: that **the probability of an interpretation and the speed with which it is processed is a function of the contextual support it receives**. In contrast, if scalar inferences generally incur a processing cost, pragmatic responses reflecting that the scalar inference was drawn should be slower to process than literal responses regardless of context. To test the constraint-based versus the costly inference account, we manipulated two features of context between participants in a truth-value judgment task: one lexical (*presence of partitive "of"*) and one pragmatic (*implicit QUD*, see (1) and (2)). This allowed us to obtain estimates of inference rate and processing speed. We further considered a participant's *responder type* -- whether they have a preference to respond literally or pragmatically -- as a predictive feature for response times. While the partitive and the QUD have previously been shown to affect the probability of drawing a scalar inference [6,7,10,11], contextual and participant-specific effects on processing speed have remained under-explored.

Experiments. Participants' interpretations were probed in a 'gumball paradigm' [11]. On each trial, participants saw a gumball machine with an upper chamber with 13 gumballs and an empty lower chamber. On critical trials, all 13 gumballs dropped to the lower chamber and participants heard the partitive statement "You got some of the gumballs" **(Exp. 1, n=800)** or the non-partitive "You got some gumballs" **(Exp. 2, n=800)**. Participants were assigned to one of two QUD conditions (*all*-QUD, *any*-QUD), which differed in the experimental cover story (see [12] for description of cover stories). Participants were asked to indicate whether they agreed (indicating a *literal* interpretation) or disagreed (indicating a *pragmatic* interpretation) with the statement by pressing a button as quickly as possible.

Judgment results (Fig. 1). Pragmatic "disagree" responses were more likely in the partitive (β= 7.16, SE=0.69, p<.0001) and in the *all*-QUD (β= 2.85, SE=0.44, p<.0001) condition, replicating previous results. Participants with >4 pragmatic responses were categorized as pragmatic responders (38% in non-partitive, 74% in partitive experiment), participants with <4 pragmatic responses as literal responders (60%% in non-partitive, 23% in partitive experiment).

Log-transformed response time results (Fig. 2). There was an interaction between QUD and response (β=-0.11, SE=0.02, t=-4.67, p<.0001), such that pragmatic responses were faster under the *all*-QUD than under the *any*-QUD and literal responses were faster under the *any*-QUD than under the *all*-QUD. The largest observed effect was the interaction between responder type and response (β=-0.27, SE=0.02, t=-11.7, p<.0001), such that pragmatic responses were faster than literal responses for pragmatic responders and literal responses were faster than pragmatic responses for literal responders.

These results suggest that contextual factors affect listeners' overall contextual response strategy, which in turn impacts the speed with which they process the preferred interpretation. This is evidence against costly inference accounts and in support of constraint-based accounts.

Implicit QUDs (manipulated via cover stories as in [12]):
(1) Did I get all of the gumballs? *(all-*QUD*, more supportive of scalar inference)*
(2) Did I get any of the gumballs? *(any-*QUD*, less supportive of scalar inference)*

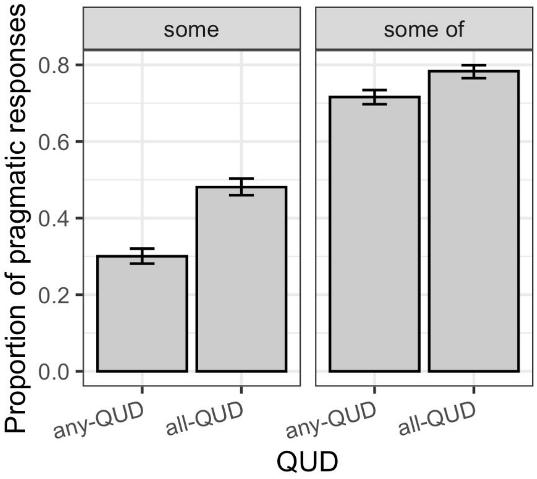

Fig. 1: Proportion of pragmatic responses on non-partitive “some” (left) and partitive “some of” (right) critical trials. Here and below, error bars indicate bootstrapped 95% confidence intervals.

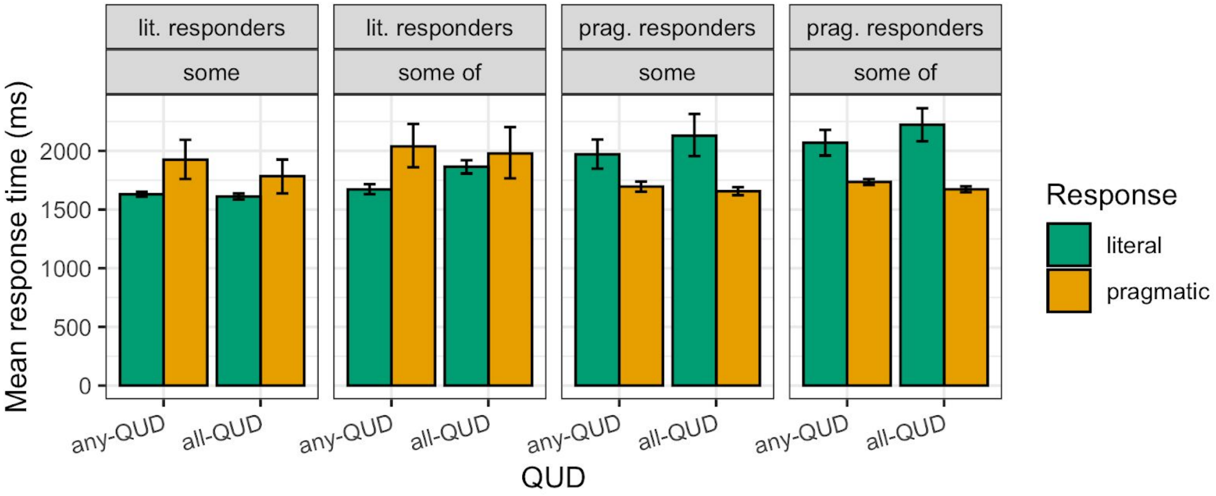

Fig. 2: Mean response times for literal (green) and pragmatic (orange) responses generated by literal (left panels) and pragmatic (right panels) responders on non-partitive “some” and partitive “some of” critical trials.

References

[1] Bott & Noveck (2004). [2] Huang & Snedeker (2009). [3] Grodner et al. (2010). [4] Breheny et al. (2013). [5] Degen & Tanenhaus (2016). [6] Zondervan (2010). [7] Degen (2015). [8] Augurzky et al. (2019). [9] Marty & Chemla (2013). [10] Degen & Goodman (2014). [11] Degen & Tanenhaus (2015). [12] Degen (2013).