Probability and processing speed of scalar inferences is context-dependent

names
{names}@school
Address

Abstract

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. This has spurred the development of studies seeking to understand the contextual conditions that facilitate scalar inferences. Here, we test a prediction made by Degen and Tanenhaus (2015)'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. We manipulated the contextual support for the scalar inference in two truth-value judgment experiments via the manipulation of a lexical feature (presence of partitive "of the") and a pragmatic feature (the implicit Question Under Discussion). Participants' responder type – whether they generally produced pragmatic responses reflecting the inference or literal responses reflecting its absence – was the main predictor of response times: pragmatic responses were faster than literal responses when generated by a pragmatic responder, whereas the reverse was true for literal responders. This suggests that, rather than generally incurring a processing cost, inferences are easy to process by listeners who take the context to generally support the inference, and hard to process by listeners who take the context not to support the inference. We interpret this as evidence against literal-first or costly-implicature accounts and in support of constraint-based accounts of pragmatic processing.

Keywords: psycholinguistics; experimental pragmatics; scalar implicature; Question Under Discussion

Introduction

[jd: this is still the elm abstract – need to flesh out each part in more detail]

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 (Bott & Noveck, 2004; Huang & Snedeker, 2008; Grodner, Klein, Carbary, & Tanenhaus, 2010; Breheny, Ferguson, & Katsos, 2013; Degen & Tanenhaus, 2016). This has spurred the development of studies seeking to understand the contextual conditions that facilitate scalar inferences (Zondervan, 2010; Degen, 2015; Augurzky, Franke, & Ulrich, 2019; Marty & Chemla, 2013; Degen & Goodman, 2014).

Here, we test a prediction made by Degen and Tanenhaus (2015)'s constaint-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 judgement task: one lexical (*presence of partititve "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 (Zondervan, 2010; Degen, 2015; Degen & Goodman, 2014; Degen & Tanenhaus, 2015), contextual and participant-specific effects on processing speed have remained under-explored.

Implicit QUDs (manipulated via cover stories as in Degen (2013)):

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*)

Experimental paradigm

In both of our experiments, participants' interpretations were probed using a 'gumball paradigm' introduced by Degen and Tanenhaus (2015). On each trial, participants were shown a display of a gumball machine with 13 gumballs in the upper chamber and an empty lower chamber. After 4 seconds, some number of gumballs moved to the lower chamber and the machine reported how many gumballs were distributed (Fig. 1). This pre-recorded statement was of the form "You got X gumballs", where X is a quantifier (*some*, *all*, *none*, or a number between 1 and 13). The number of gumballs that are dropped to the lower chamber and the quantifier reported by the gumball machine were varied and they did not always match (see Table 1).

Participants were assigned to one of the two conditions (*all*-QUD, *any*-QUD) which differed in the cover story they were presented with at the beginning of the experiment. These cover stories were designed to establish an implicit QUD (see Table 2).

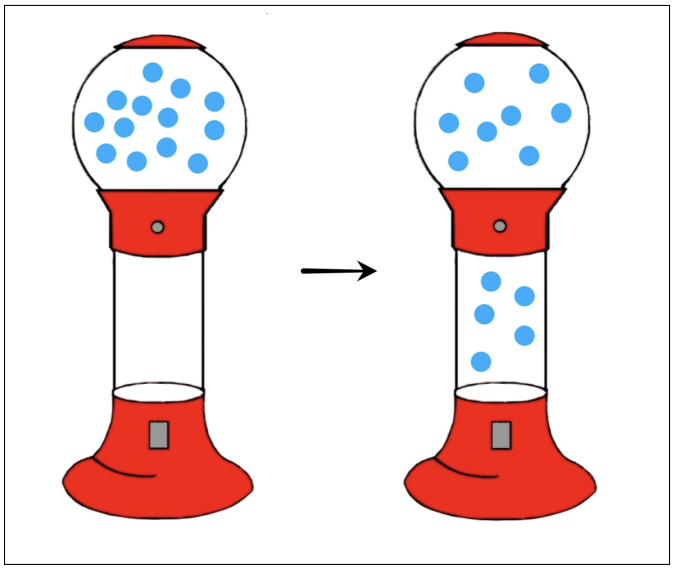

Figure 1: Example display from gumball paradigm. Left: initial display. Right: display with 5 gumballs dropped.

	Set size						
Quantifier	0	2	5	8	11	13	**Total**
some/some of	4	1	1	1	1	8	16
all of	2	1	2	1	2	8	16
none of	4	1	0	1	1	1	8
number	3	7	7	7	5	3	32
Total	13	10	10	10	9	20	72

Table 1: Distribution of experimental trials over quantifiers and set sizes.

Experiment 1: Partitive statement

In experiment 1 we tested whether the QUD, as a contextual feature of an utterance, could modulate the probability of a scalar implicature and the speed with which it is processed. The main prediction was that in the *all*-QUD condition, the implicit QUD "Did I get all of the gumballs", would be more relevant to the "You got all of the gumballs" interpretation. Thus, there would be more pragmatic "disagree" responses in the critical trials when participants hear "You got some of the gumballs" and get all 13 gumballs. We also predicted that the relevance of the QUD would increase the speed of pragmatic responses and slow down the literal responses.

Procedure, materials, analyses and exclusions were pre-registered on OSF and will be available upon publiciation along with data and experiment scripts.

Methods

Participants We recruited 800 participants on Amazon Mechanical Turk. Participants were required to have a US-based IP address and a minimal approval rating of 95%. They were paid $2.3.

Materials and procedure After reading the cover story of their QUD condition, participants went through a scripted demonstration that showed the consequences of store worker's responses to various scenarios. To ensure that they paid attention to the cover story, they were asked a multiple choice question about when the store worker will be fired. When participants answered this question incorrectly, they were presented with the cover story again and went through the same demonstration. Halfway through the experiment, participants were asked to answer this multiple-choice question again. This was done to prevent the decay of the implicit QUD over time.

There were 4 practice trials with *all* and *none*. On half of these trials, the statements were correct, and on the other half they were incorrect. After the practice trials, there were 72 experimental trials. On 32 of the trials, the expected answer was yes, and on 32 of the trials, the expected answer was no. The remaining 8 trials were occurances of the critical trial and the main focus of this experiment. On these trials, all 13 gumballs dropped to the lower chamber and participants heard the partitive statement "You got *some of* the gumballs". When participants press YES to agree with this statement, they interpret it semantically as "You got some, and possibly all, of the gumballs" and when they press NO to disagree, they interpret it pramatically as "You got some, but not all, of the gumballs".

Exclusions We excluded participants who were self-reported non-native English speakers (n=X), participants who got the second cover story comprehension questions wrong more than twice (n=Y) and participants with accuracy lower than 85% on non-critical trials (n=Z). Only responses to critical trials are reported below.

all-QUD	any-QUD
You are at a candy store and are testing a row of gumball machines. These are special gumball machines that say how many gumballs you got. However, this report is sometimes faulty.	
The store worker tells you that his boss has threatened to fire him if the gumball machines are left empty, and he really needs this job. He cannot see the machines from the register, but he can normally tell how full they are by the machines' statements.	The store worker tells you that machines sometimes jam and don't deliver any gumballs. His boss has threatened to fire him if the gumball machines stay jammed, and he really needs this job. He cannot see the machines from the register, but he can normally tell if they are working by the machines' statements.
He asks you to tell him if the statement is right or wrong, so that he will know if a machine is empty and needs to be refilled. After you hear the statement, you have 4 seconds to notify the store worker, so please make a decision as quickly as possible.	

Table 2: Cover stories for each QUD condition.

Results and discussion

Judgements

Proportion of pragmatic responses on critical trials are shown in Figure 2. 78% of responses given by the participants in the *all*-QUD condition were pragmatic compared to 71% pragmatic responses given by participants in the *any*-QUD condition. We ran a mixed effects logistic regression predicting response type with random by-participant intercepts from fixed effects of QUD and found a main effect of QUD such that there are more pragmatic responses for *all*-QUD compared to *any*-QUD (β=1.31, SE=0.52, $p<.05$). In the *all*-QUD condition, the cover story made the question participants were trying to answer more relevant to the "You got all of the gumballs" interpretation and participants were more likely to respond pragmatically.

Analysis of Variability in Judgements

In order to asses participants' response consistency, we [lk: plotted] the distribution of participants over number of pragmatic responses given on critical trials (see Figure 3). While there were people who were completely consistent, there was also some variability. For the purpose of the subsequent response time analyses, we divided participants into two groups. Following the procedure in Degen (2015), we categorized participants with more than 4 pragmatic responses as pragmatic responders (38%) and participants with less than 4 pragmatic responses as literal responders (60%). [lk: X] participants ([lk: X%]) gave an equal number of pragmatic and literal responses and were excluded from this analysis.

Response Times

There was an interaction between QUD and response (β=-0.12, SE=0.03, t=-3.70, $p<.0001$) such that pragmatic responses were faster under the *all*-QUD than under the *any*-QUD. This shows that the QUD, as a feature of the context, not only affects the robustness but also the speed of scalar inferences.

[lk: connect these two paragraphs] There was also an interaction between responder type and response (β=-0.28, SE=3.34, t=-8.37, $p<.0001$). Pragmatic responses were faster than literal responses for pragmatic responders and literal responses were faster than pragmatic responses for literal responders.

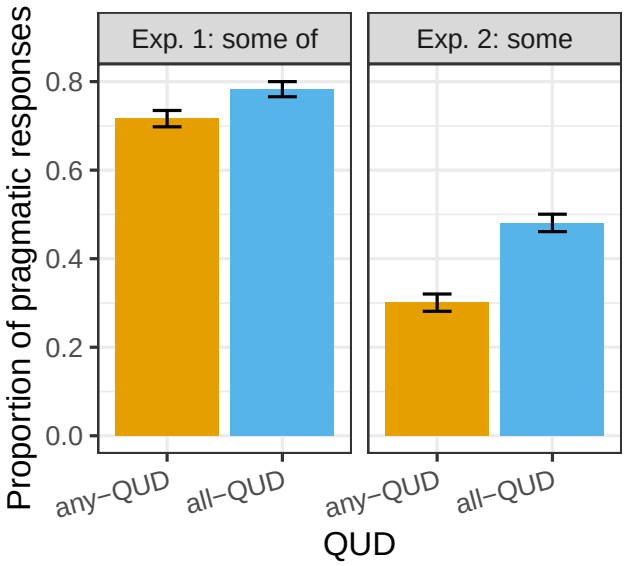

Figure 2: Proportion of pragmatic responses on partitive "some of" (left) and non-partitive "some" (right) critical trials. Error bars indicate bootstrapped 95% confidence intervals.

Experiment 2: Non-partitive statement

In Experiment 2 we tested whether the absence of the partitive, previously shown to decrease the contextual support for the inference, would decrease the inference strength and robustness of scalar implicatures,

Methods

Participants We recruited 800 participants on Amazon Mechanical Turk. Participants were required to have a US-based

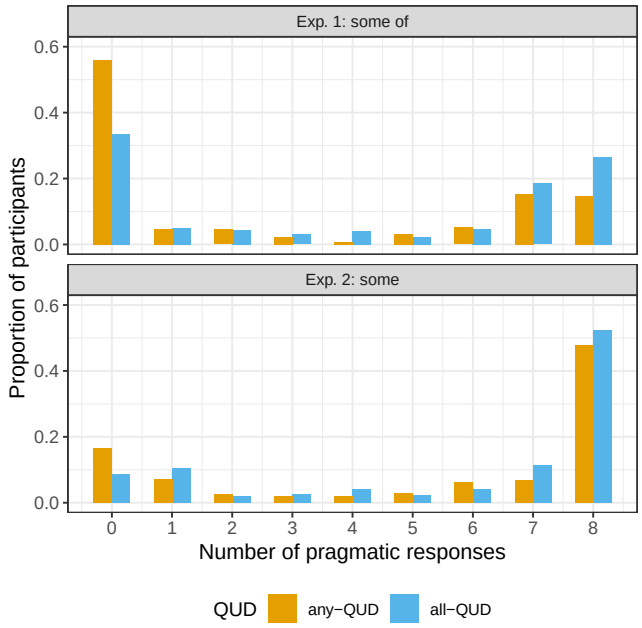

Figure 3: Distribution of participants over number of pragmatic responses given on critical trials.

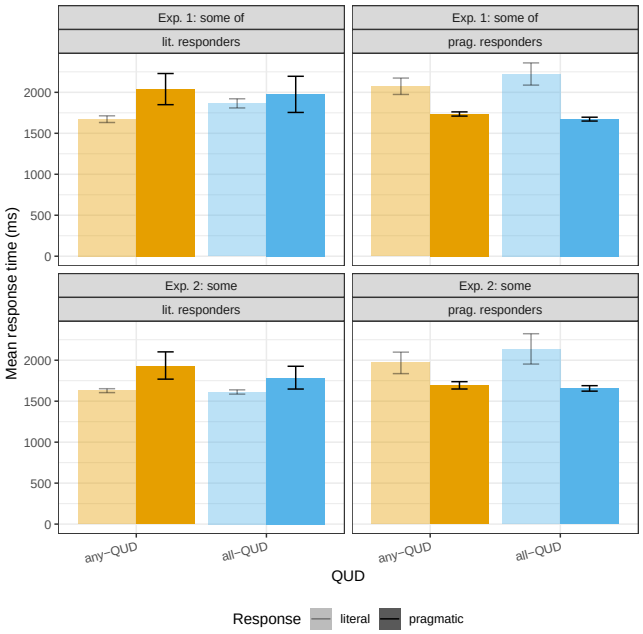

Figure 4: Mean response times for literal (green) and pragmatic (orange) responses generated by literal (upper panels) and pragmatic (lower panels) responders on partitive "some of" and non-partitive "some" critical trials.

IP address and a minimal approval rating of 95%. They were paid $2.3.

Materials and procedure The materials and procedures were the same as in Experiment 1 except on critical trials, when all 13 gumballs dropped to the lower chamber, participants heard the non-partitive statement "You got *some* gumballs".

Exclusions As in Experiment 1, we excluded non-native English speakers (n=X), participants who got the second comprehension question wrong more than twice (n=Y), and participants that had accuracy lower than 85% on non-critical trials (n=Z).

Results and discussion

Judgements

Replicating the QUD effect found in Experiment 1, we found that participants in the *all*-QUD condition gave more pragmatic "disagree" responses than participants in the *any*-QUD (β=4.69, SE=0.80, p<.0001). [lk: When data was pooled: pragmatic "disagree responses were more likely in the *all*-QUD (β=2.85, SE=0.44, p<.0001) condition]

We found that overall, participants in Experiment 2 who heard the non-partitive statement were less likely to respond pragmatically compared to participants in Experiment 1 who heard its partitive countepart (β=7.16, SE=0.69, p<.0001)(see Fig 2).

We also found an interaction of the lexical cue and QUD (β=-3.06, SE=0.90, p<.0001) but it was due to the qud effect being bigger for the non-partitive condition than the partitive condition. [lk: This interaction is perhaps due to a ceiling effect.]

Analysis of Variability in Judgements

[lk: As shown in Figure 3, plot skewed..]

Response Times

When the non-partitive form was used, the QUD didn't affect the speed of scalar implicatures significantly, however when data was pooled from both experiments, there was a main effect of QUD such that [lk: fill].

[lk: this analysis needs to be done separately for the two quantifiers because the stims differ in length?? - when done together: interaction of lexical cue and response (β=-4.07, SE=1.71, t=-2.38, p<.01)]

[lk: full model: interaction between qud and response (β=-1.07, SE=2.29, t=-4.67, p<.001), largest observed effect: 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.]

General discussion and conclusion

[lk: from ELM abstract]

Overall, 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.

This suggests that contextual factors affect listeners' overall contextual response strategy which in turn impacts the speed with which they process the preferred interpretation.

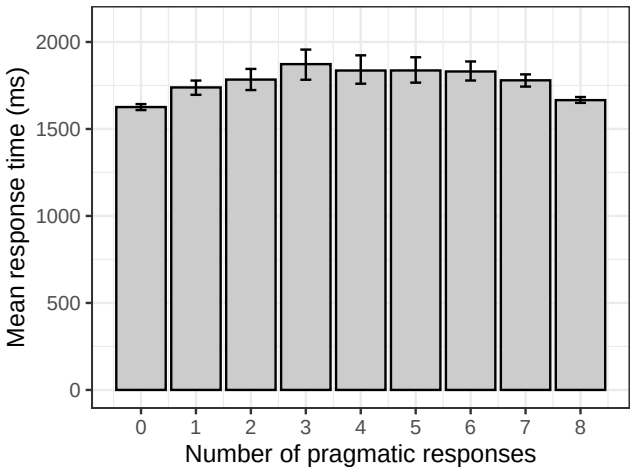

Figure 5: Mean response times for participants grouped based on the number of pragmatic responses they gave.

This is evidence against costly inferece accounts and in support of constraint-based accounts.

References

Augurzky, P., Franke, M., & Ulrich, R. (2019, 08). Gricean expectations in online sentence comprehension: An erp study on the processing of scalar inferences. *Cognitive Science*, *43*. doi: 10.1111/cogs.12776

Bott, L., & Noveck, I. (2004, 10). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437-457. doi: 10.1016/j.jml.2004.05.006

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 443-467. doi: 10.1080/01690965.2011.649040

Degen, J. (2013). Alternatives in pragmatic reasoning.

Degen, J. (2015, May). Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55. doi: 10.3765/sp.8.11

Degen, J., & Goodman, N. (2014). Lost your marbles? the puzzle of dependent measures in experimental pragmatics. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*(4), 667–710.

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive science*, *40*(1), 172–201.

Grodner, D., Klein, N., Carbary, K., & Tanenhaus, M. (2010, 07). "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42-55. doi: 10.1016/j.cognition.2010.03.014

Huang, Y.-t., & Snedeker, J. (2008, 11). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology*, *58*, 376-415. doi: 10.1016/j.cogpsych.2008.09.001

Marty, P., & Chemla, E. (2013, 07). Scalar implicatures: Working memory and a comparison with only. *Frontiers in psychology*, *4*, 403. doi: 10.3389/fpsyg.2013.00403

Zondervan, A. (2010, 01). Scalar implicatures or focus: an experimental approach.