Probability and processing speed of scalar inferences is context-dependent

Leyla Kursat and **Judith Degen**
{lkursat, jdegen}@stanford.edu
Department of Linguistics, Stanford University
Stanford, CA 94305, USA

Abstract

need to change. Speakers exhibit variability in their choice between uncertainty expressions such as *might* and *probably*. Recent work has found that listeners cope with such variability by updating their expectations about how a specific speaker uses uncertainty expressions when interacting with a *single speaker*. However, it is still unclear to what extent listeners form speaker-specific expectations for *multiple speakers* and to what extent listeners are adapting to a situation independent of the speakers. Here, we take a first step towards answering these questions. In Experiment 1, listeners formed speaker-specific expectations after being exposed to two speakers whose use of uncertainty expressions differed. In Experiment 2, listeners who were exposed to two speakers with identical use of uncertainty expressions formed considerably stronger expectations than in Experiment 1. This suggests that listeners form both speaker-specific and situation-specific expectations. We discuss the implications of these results for theories of adaptation.

Keywords: psycholinguistics; semantics; pragmatics; adaptation; uncertainty expressions

Introduction

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.

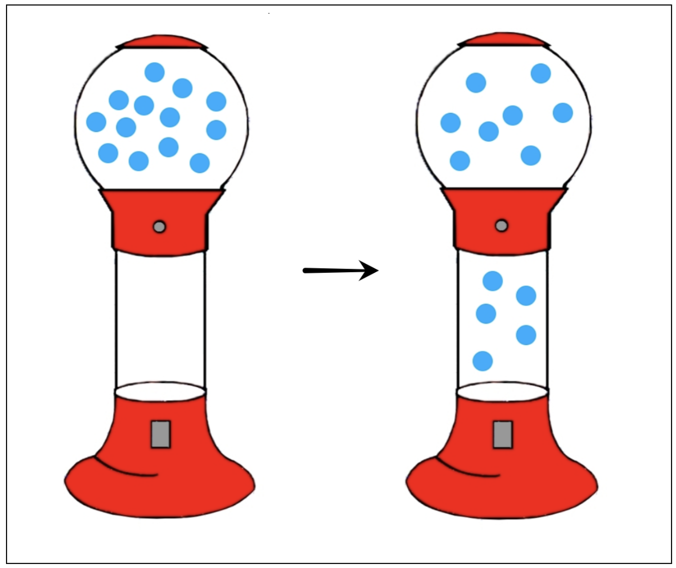

Figure 1: Example display from gumball paradigm. Left: initial display. Right: display with 5 gumballs dropped.

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). This paradigm was developed to test XXXX. On each trial, participants were shown a display of a gumball machine with 13 gumballs in the upper chamber

	Set size						
Quantifier	0	2	5	8	11	13	**Total**
some/some of	x	x	x	x	x	x	x
all of	x	x	x	x	x	x	x
none of	x	x	x	x	x	x	x
number	x	x	x	x	x	x	x
Total	x	x	x	x	x	x	x

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

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.

***all*-QUD** Participants assigned to the *all*-QUD condition read a cover story explaining them that they are at a candy store and testing a row of special gumball machines. These gumball machines report how many gumballs they distribute but are sometimes faulty in their report. The store worker's boss has threatened to fire him if the gumabll machines are left empty and he cannot see the machines from the register and only tell how full they are by their statements. Participants were asked to help the store worker by telling him if this statement is right or wrong by pressing the yes or no key. They were also told that they have 4 seconds to notify the sotre worker and they should make a decision as quickly as possible.

Is the machine empty? $\rightarrow$ Did I get all of the gumballs?

***any*-QUD** Participants in the *any*-QUD condition read the same cover story except in their story, the gumball machines sometimes jam and don't deliver gumballs. The store worker's boss has threatened to fire him if the gumball machines stay jammed.

Is the machine jammed? $\rightarrow$ Did I get any of the gumballs?

Experiment 1: Partitive statement

In Experiment 1, the partitive form was XXX. Procedure, materials, analyses and exclusions were pre-registered on OSF (https://osf.io/suwtr).

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 scenerios. 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 2 of these trials, the statement was correct, and on 2 of them it was 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).

Analysis and predictions Only responses to critical trials are reported below. (?)...

Results and discussion

Judgements As shown in Figure 2, participants in the *all*-QUD condition gave more pragmatic ”disagree” responses than participants in the *any*-QUD. or A mixed effects logistic regression predicting response type with random by-participant intercepts from fixed effects of QUD found a main effect of QUD such that there are more pragmatic responses for *all*-QUD compared to *any*-QUD (β=0, SE=0, t=0, p$<$.0001). This was previously shown....

Participants were divided into two groups based on their responses. Participants with more than 4 pragmatic responses were categorized as pragmatic responders (38%) and participants with less than 4 pragmatic responses were categorized as literal responders (60%).

Response Times There was an interaction between QUD and response (β=0, SE=0, t=0, p$<$.0001) such that pragmatic responses were faster under the *all*-QUD than under the *any*-QUD. The largest observed effect was the itneraction between

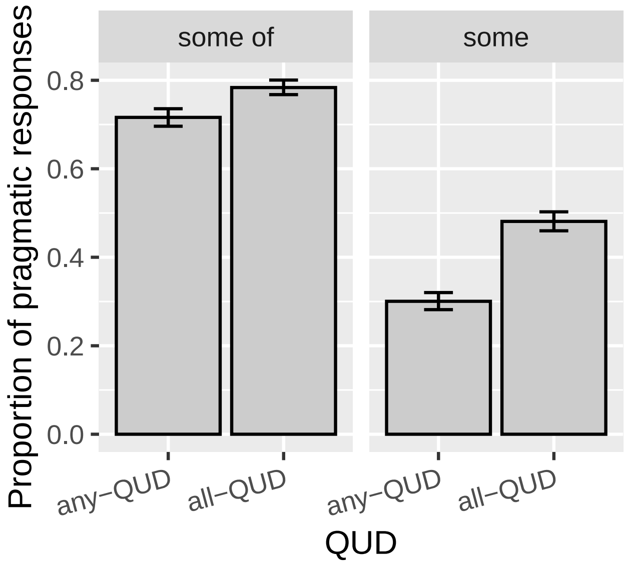

Figure 2: Proportion of pragmatic responses on partitive ”some of” (left) and non-partitive ”some” (right) critical trials. Error bars indicate bootstrapped 95% confidence intervals.

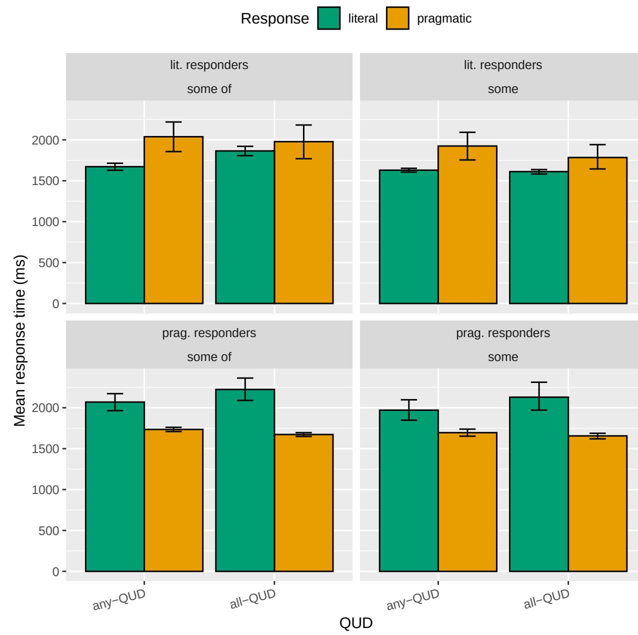

Figure 3: Mean response times for literal (green) and pragmatic (orange) responses generated by literal (upper panels) and pragmatic (lower panels) responders on partitive ”some of” and non-partitive ”some” critical trials.

responder type and response (β=0, SE=0, t=0, 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.

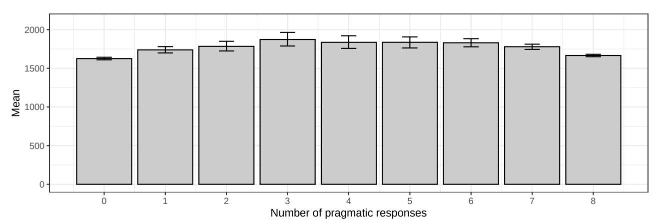

Figure 4: Mean response times for participants grouped based on the number of pragmatic responses they gave.

Experiment 2: Non-partitive statement

In Exp. 2,

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

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).

Analysis and predictions

Results and discussion

Lastly, the effect size (Cohen's *d*: 1.68) was larger in this experiment than in Experiment 1 and the single-speaker

General discussion and conclusion

In two experiments, we found that.This suggests 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 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.