Making Inferences: The Case of Scalar Implicature Processing

Judith Degen (jdegen@bcs.rochester.edu)
Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain and Cognitive Sciences, University of Rochester
Rochester, NY 14627-0268, USA

Abstract

Scalar implicature has served as a test case for investigating the nature of inference processes in language comprehension. Specifically, the question of whether or not scalar implicatures are computed by default has been extensively investigated in recent years. We argue that the question of default is overly simplistic and propose instead to think of scalar implicature computation as a problem of optimal cue combination within a constraint-based framework. We provide evidence from three experiments supporting the view that multiple constraints of differing strength operate in parallel to provide probabilistic support for or against an implicature.
Keywords: experimental pragmatics; scalar implicature; eye-tracking; subitizing

Introduction

Successful communication requires comprehenders to infer a speaker's intended meaning from an underspecified utterance. While some information is transmitted via an utterance's semantic content, additional meaning is computed by taking into account pragmatic information about the discourse context, as in the sample discourse in (1).

(1) Alex: Did you submit your paper?
Thomas: Some of the sections are written.
⇝ Thomas didn't submit his paper.
⇝ Some, but not all, of the sections are written.

Here, Alex might infer that Thomas, being a cooperative speaker, intends to convey both that he hasn't yet submitted his paper and that some, but not all, of the sections are written. The former of these inferences is what Grice (1975) termed a *particularized conversational implicature* (PCI), while the latter is a prototypical example of a *generalized conversational implicature* (GCI). According to Grice, both of these inferences arise because comprehenders assume that speakers conform to certain conventions of rational communicative behavior. The crucial difference between GCIs and PCIs lies in the role that context plays: PCIs are assumed to arise in virtue of special features of the context, while GCIs are assumed to arise unless blocked by context. Applied to (1): the inference that Thomas did not submit his paper is tied tightly to Alex's question and would not have arisen if Thomas's utterance was, for example, an answer to the question whether Thomas had written the introduction yet. However, the inference that some but not all of the sections are written is taken to arise independently of the context. This particular kind of inference, that arises in virtue of a speaker not uttering a relevant stronger alternative, is called a *scalar implicature*.

In the example there is a stronger statement the speaker could have made but didn't (e.g. *All of the sections are written*). Under the assumption that a speaker is being as informative as possible, a weak implicature, that the speaker doesn't know whether all of the sections are written, is licensed. If the hearer further assumes that the speaker is competent with respect to the truth of the stronger statement, the implicature that the speaker believes that some, but not all of the sections are written, is licensed.

In recent years the representation and processing of scalar implicatures has emerged as perhaps the defining problem in experimental pragmatics – a subfield of cognitive science that seeks to combine theoretical proposals from linguistics, computer science and philosophy with state-of-the-art experimental methods.

Importantly for the experimental investigation of the phenomenon, scalar implicatures are cancelable, that is, they are defeasible inferences. There are cases in which the scalar implicature does not contribute to achieving the discourse goal (Horn, 1984, Sperber & Wilson, 1995, Levinson, 2000). In such cases, where all that is relevant or can be known, is the lower bound (in our example that *at least some* of the sections are written), the implicature does not arise. Following Katsos, Breheny, and Williams (2005), we will call these contexts *lower-bound* contexts, and contexts in which pragmatic inference is required to achieve the discourse goal *upper-bound* contexts.

Emphasising the GCI-PCI distinction, Levinson (2000) argues that GCIs are pragmatic default inferences that have evolved to maximize the speed of communication. It is only in special contexts that the inference has to be cancelled, where cancellation proceeds in a second, effortful step.

In contrast to this *default* approach, a variety of approaches have viewed scalar implicature as a *context-driven* process (e.g. Hirschberg, 1991, Sperber & Wilson, 1995). Under these accounts, scalar implicatures are generated as part of the same process as PCIs. The strongest formulation of such accounts is a modular one, whereby pragmatic processing begins only once semantic processing is under way (Huang & Snedeker, 2009). That is, generating the implicature requires computing the literal meaning first. A notion of implicit cancellation is not necessary under the context-driven account since the implicature does not arise in lower-bound contexts in the first place.

The default and context-driven approach make different empirical predictions. Under the default model, generating a scalar implicature should be very rapid, indeed as rapid as computing an expression's literal meaning. An increase in processing effort is predicted only for cases where the implicature is cancelled. In contrast, the modular context-driven account predicts increased processing effort for implicature generation as compared to literal meaning.

Support for the context-driven account comes from reaction-time experiments (Noveck & Posada, 2003, Bott & Noveck, 2004), reading-time experiments (Breheny, Katsos, & Williams, 2006), and visual-world eye-tracking (Huang & Snedeker, 2009). Response times associated with the pragmatic inference are longer than both response times to the scalar item's literal meaning and to other literal controls (typically statements including *all* or number terms instead of the scalar item *some*). For example, Huang and Snedeker (2009) found that participants' gaze was slower to converge on a target that required pragmatic inference for disambiguation than one that did not. Participants saw displays with a contrast between e.g. a girl with a proper subset of all socks in the display and a girl with all of the soccer balls in the display, and heard auditory stimuli such as *Point to the girl with some of the socks*. If *some* was automatically enriched to *some but not all*, disambiguation should be just as rapid for *some* as for *all*. However, since both girls are compatible referents of lower-bound *some*, increased looks to the target should be delayed if the enrichment is not automatic. This is indeed what Huang and Snedeker (2009) find.

However, Grodner et al. (2010) found evidence of rapid implicature generation. Their design differed from Huang and Snedeker (2009) in three major ways: first, they did not use instructions with exact number. Second, each trial began with a prerecorded statement describing the total number and type of objects in the display, to draw participants' attention to the total cardinality of each type of item. And finally, they included the quantifier *none*. In this setup, convergence on the target in the *some* and *all* condition was equally fast.

Under an approach to scalar implicatures as either default GCI or context-driven PCI (in the modular sense), these results are puzzling. If scalar implicatures associated with *some* are defaults, there should be no delays. If, however, implicature computation only occurs after an initial stage of semantic processing, the fast implicatures found by Grodner et al. (2010) cannot be explained.

In other domains of language processing initial arguments for an encapsulated early stage (Frazier, 1987) of context-free processing (akin to computing the literal interpretation) followed by a second stage of evaluation and re-analysis (akin to a second stage of pragmatic inference) have been superceded by models in which evidence from multiple constraints is evaluated from the earliest moments of processing (e.g. Trueswell, Tanenhaus, & Garnsey, 1994, MacDonald, Pearlmutter, & Seidenberg, 1994). A crucial step in developing these accounts has been to understand and quantify the relevant constraints.

We propose that the robustness of a scalar implicature and the speed with which it is computed depends on the strength and reliability of multiple probabilistic cues that are integrated as soon as they become available; the stronger and the more reliable the cues to the implicature are, the faster it should become available and the more robust it should be.

This account raises two related questions. First, what are the cues that contribute to or inhibit the implicature? And second, do we see the predicted effect of cue strength and reliability reflected in measures of processing difficulty?

The cues we investigate are a) the **syntactic form** of the *some*-NP, specifically whether partitive *some of the* is a stronger cue to the implicature than simple *some*; b) the availability of **lexical alternatives** to *some*; and c) the **set size** relative to which *some* is interpreted. The investigation of the tradeoff between b) and c) was directly motivated by features of the studies reported in Huang and Snedeker (2009) and Grodner et al. (2010).

In both of these studies the referent of the *some*-NP was a girl or boy that had a set of objects. The size of this set was always within the subitizing range (1 - 4 items). In this range judgments of number are rapid, accurate, and confident (Kaufman, Lord, Reese, & Volkmann, 1949). Number judgments for larger set sizes require counting, with response times dramatically increasing for each added item. The crucial difference between the two studies is that number terms (*two, three*) were present in Huang and Snedeker (2009) but not in Grodner et al. (2010). Our studies are motivated by the following argument: Given that subitizing is accompanied by a strong sense of number "popping out", it is conceivable that number terms are more natural labels for set sizes in this range. From a Gricean perspective, speakers should use the most natural label available to establish reference. If instead a vague quantifier like *some* is used, it is likely that this will result in increased processing effort on the comprehender's part. This is one potential explanation for the different results found by Huang and Snedeker (2009) and Grodner et al. (2010). In the former study, number terms were available to refer to the intended referent, while this was not the case in the latter study. Thus, the presence of number terms as a more natural cue to the referent might have led to the delay in processing of *some*. In addition, the presence of *none* in Grodner et al. (2010) but not in Huang and Snedeker (2009) might have made the ⟨all, some, none⟩ scale more salient, potentially resulting in even faster processing of *some*. Thus, if it is the case that number terms are preferred over *some* for referring to sets in the subitizing range, use of *some* should lead to processing delays in the subitizing range but not outside of it. Similarly, if *all* is preferred over *some* for referring to unpartitioned sets, use of *some* for referring to an unpartitioned set should lead to processing delays.

Experiment 1 asks whether *some* is indeed dispreferred in the predicted ranges via naturalness ratings. Experiment 2 asks whether the preferences found in Experiment 1 are reflected in response times. Experiment 3 replicates the finding from Experiment 2 in an eye-tracking paradigm to ensure comparability with the studies reported by Huang and Snedeker (2009) and Grodner et al. (2010). Experiments 1 and 2 additionally test the effect of the partitive as a cue to scalar implicature.

The gumball paradigm

All experiments were conducted in a "gumball paradigm". Participants see a display of a gumball machine that initially has a full upper chamber and an empty lower chamber. After 2.5 seconds a "ka-ching" sound is followed by a new display in which some number of gumballs has "moved" to the lower chamber. Participants then hear a pre-recorded statement of the form *You got X gumballs*, where *X* is a quantifier or number term. Depending on the task, they are asked to rate the naturalness of the statement as a description of the scene (Experiment 1), press a button indicating whether or not they agree with the statement (Experiment 2), or click on the mentioned gumballs if they agree with the statement while eye movements are recorded (Experiment 3). Then the next trial begins.

The flexibility of this paradigm thus allows for manipulating a number of key variables, e.g. the size of the set in the lower chamber, the quantifier employed in the statement, and the behavioral measure.

Experiment 1

If number terms are more natural labels than *some* in the subitizing range, naturalness ratings should be lower for *some* than for number terms. Moreover, naturalness of *some* should be lower than that of *all* for the unpartitioned set of 13 gumballs. In addition, if partitive *some of the* is a better cue to an implicature than simple *some*, naturalness ratings for simple *some* should be higher than for the partitive when used with the unpartitioned set but not with partitioned sets of 1 - 12 gumballs.

Methods

PARTICIPANTS. Twenty undergraduate students from the University of Rochester were paid to participate. All were native speakers of English who were naïve as to the purpose of the experiment.

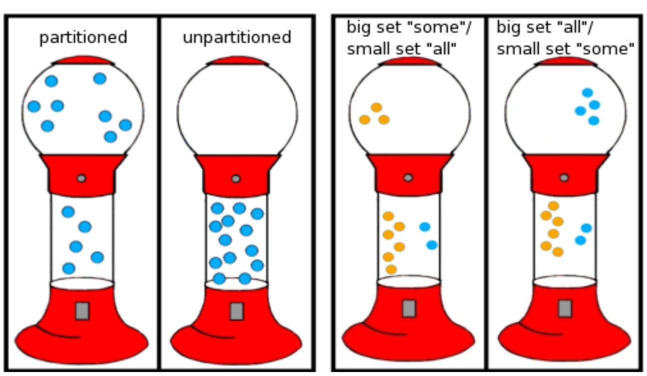

Figure 1: Display types in Experiments 1 and 2 (left) and contrast display types in Experiment 3 (right).

PROCEDURE AND MATERIALS. Participants rated the naturalness of the statement they heard as a description of the scene using a seven point Leikert scale on a total of 112 trials. We varied both the size of the set in the lower chamber (0, 1, 2, 3, 5, 7, 8, 11, 12, or 13 gumballs) and the quantifier construction in the statement (*some, some of the, all of the, none of the, one of the, two of the, three of the, seven of the, eleven of the*). Table 1 shows how often each quantifier occurred with each set size. To limit the number of trials, only a subset of the total range was sampled, and not every quantifier occurred with every set size. See Figure 1 for sample displays.

Table 1: Number of trials with each set size/quantifier. *some* column contains number of stimuli for both simple and partitive *some*. All other quantifiers occurred only in the partitive.

	some	all	none	one	two	three	seven	eleven
0	4	2	4	1	1	1	1	1
1	2		4					
2	2	2	2		4			
3						4		
5	2	2	2	1	1	1	1	1
7	2	2					4	
8	2			1	1	1	1	1
11	2	2						4
12	2	2		1	1	1	1	1
13	4	4	4					

No number of gumballs occurred more often with any other quantifier than with its correct exact number. Table 1 shows how often each number of gumballs occurred with each quantifier or number term. There were two versions of every visual display to prevent participants from learning the number of displayed gumballs based on the visual pattern.

Results and Discussion We fitted a series of mixed effects linear models predicting mean naturalness rating from quantifier (*some*, *some of the*, correct exact quantifier or number term) with random subject intercepts for each range of gumballs (subitizing range: 1-4 gumballs, mid range: 5-8, high range: 9-12, unpartitioned set: 13). Ratings for *some*/*some of the* were lower in each comparison (see Figure 2), but the difference was largest in the subitizing range ($\beta = 3.68$, $SE = 0.48$, $p < .001$) and for the unpartitioned set ($\beta = 3.98$, $SE = 0.47$, $p < .001$), where we hypothesized the greatest effect of salient lexical alternatives interfering with the interpretation of *some*. As predicted, ratings for partitive *some* were lower than for simple *some* when used with the unpartitioned set ($\beta = 1.55$, $SE = 0.39$, $p < .01$).

Experiment 2

Using the same paradigm and stimuli, we asked participants to respond to the statements by pressing one of two buttons depending on whether they agree with the statement or not, and measured their response times. We predicted that the availability of more natural alternatives to *some* as established in Experiment 1 should result in a delay in response time, regardless of the response itself. This stands in contrast to a simple modular view of implicature, where a delay is only

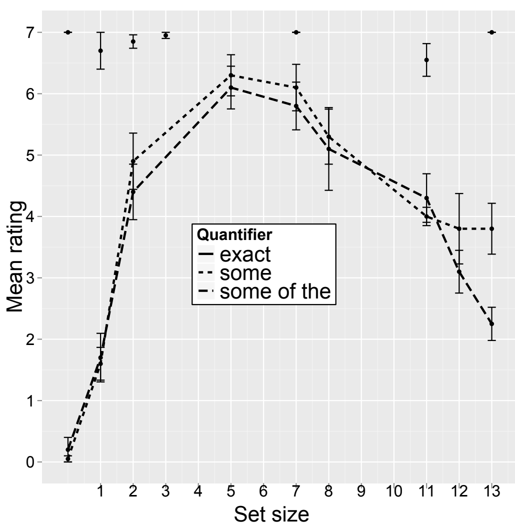

Figure 2: Mean naturalness ratings. For *none*, *all*, and number terms only the rating for the correct number of gumballs is shown (*exact*). Dotted lines show mean ratings for simple *some* and partitive *some of the*.

expected for the pragmatic response to the unpartitioned set.

Methods

PARTICIPANTS. Twenty-seven undergraduate students from the University of Rochester were paid to participate. All were native speakers of English who did not participate in Experiment 1.

PROCEDURE AND MATERIALS. The procedure and materials were the same as in Experiment 1, except participants responded by pressing one of two buttons to indicate that YES, they agreed with the heard statement, or NO, they disagreed. Participants were asked to respond as quickly as possible. If they did not respond within four seconds of stimulus onset, the trial timed out and the next trial began. Both participants' judgments and reaction times were measured.

Results and Discussion

Judgments. The condition that replicates the underinformative conditions from Bott and Noveck (2004) is when participants received the unpartitioned set and heard *You got some (of the) gumballs.* Under a semantic interpretation of *some*, participants should respond YES, while a pragmatic interpretation yields a NO response. The judgment data replicates the findings from these earlier studies: 99% of participants responded YES to *all*, while only 58% responded YES to *some of the*. Interestingly, 91% of participants gave a semantic response to simple *some*, which is significantly more than for partitive *some* (β = 2.39, *SE* = 0.4, $p < .0001$). We conclude that the partitive is a better cue to the implicature than is simple *some*.

Response times. Response times ranged from 567 - 3961ms (mean: 1461ms, SD: 512ms). Response time results for *some* and *some of the* are not reported separately, as they never differed. We fitted a series of mixed effects linear models predicting log reaction times, with random effects of speaker and item.

Replicating earlier reaction time results (Bott and Noveck (2004)), semantic YES responses to *some* for the unpartitioned set are faster than pragmatic NO responses (β = .14, *SE* = .06, $p < .01$). In addition, pragmatic responses to *some*, reflecting the implicature, are slower than YES responses to *all* for the unpartitioned set (β = 0.18, *SE* = 0.0004, $p < 0.001$).

However, it is illustrative to compare mean response times of YES responses to *some* in the mid range (5 - 12 gumballs) vs. the unpartitioned set (where it constitutes a semantic response) and vs. the subitizing range (1 - 4 gumballs). Recall that there are more natural lexical alternatives both in the subitizing range (number terms) and for the unpartitioned set (*all*). YES responses to *some* are slower both in the subitizing range (β = 0.08, *SE* = 0.02, $p < 0.001$) and for the unpartitioned set (β = 0.13, *SE* = 0.02, $p < 0.001$) than in the mid range. In addition, responses to *some* in the subitizing range are slower than to number terms (β = 0.16, *SE* = 0.02, $p < 0.001$). These effects are demonstrated in Figure 3.

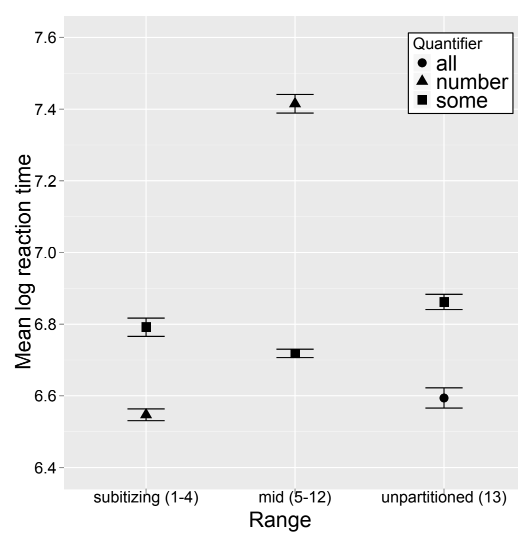

Figure 3: Mean log reaction times for YES responses.

The availability of salient lexical alternatives thus intrudes even on the semantic, lower-bound interpretation of *some*. This result cannot be accounted for by a view in which pragmatic information is not integrated until after an initial stage of semantic processing.

Experiment 3

This experiment monitored eye movements. Critical trials contained a contrast between a partitioned and an unpartitioned set in the lower chamber (see Figure 1). Importantly, we varied whether the target set size was in the subitizing range. In contrast to the previous eye-tracking studies (where *some* always referred to a set that was smaller than the set that *all* referred to), this allowed for direct comparison of quantifiers while controlling for size-driven baseline differences that have been observed in previous visual world studies (for discussion see Grodner et al., 2010).

We predict delayed convergence on the target after *some* relative to *all* and number terms within the subitizing range, where more natural lexical alternatives to *some* are available.

Methods

PARTICIPANTS. Twenty-four undergraduate students from the University of Rochester were paid to participate. All were native speakers of English who had not participated in either of the two previously reported experiments.

PROCEDURE AND MATERIALS. This setup differed from Experiments 1 and 2 in four ways. First, two colors of gumballs were available to create the contrast between partitioned and unpartitioned set. This led to a slight modification in the statement that participants heard, which was modified to contain a color adjective: *You got Q of the C gumballs*, where *Q* was one of *some, all, two*, or *six* and *C* was either *blue* or *orange*. Second, to ensure that participants' gaze location at stimulus onset was comparable across trials, the "kaching" sound was replaced with a visual cue: after viewing the first display for 2 seconds, the central button on the gumball machine began to flash. Once participants clicked the button the display changed (the gumballs "moved"). The auditory stimulus began 500ms later.

The third difference is that the upper chamber did not initially contain the same number of gumballs on every trial. Instead, it either contained six gumballs of each color or two gumballs of one color and nine of the other. This ensured that on critical trials the machine's lower chamber in the second display contained a contrast between two gumballs of one color and six of the other that was crucially the same for both display types (see Figure 1). *Some* was used to refer either to the small (subitizable) or the big (non-subitizable) set, and vice versa for *all*.

Finally, the participants' task was to respond to the statement by clicking on the set of gumballs in the lower chamber that had been mentioned if they believed that the person uttering the statement was right. If they thought the person was wrong, they were instructed to click on the button in the center of the machine.

We manipulated four variables. First, one of four quantifiers occurred in the statement. Second, the lower chamber either contained a contrast between a set of two and a set of six gumballs or not, resulting in a hypothesized early (in the contrast condition) vs. late (no contrast) point of disambiguation. In the no-contrast condition, set size was the same for both sets of gumballs in the lower chamber (either two for *some/two* or six for *all/six*). The no-contrast condition served as a baseline: there should be no difference in convergence time for the different quantifiers when there is no contrast cue to the intended referent. Third, within the contrast condition, we manipulated whether *some* and *all* referred to the big or small set and analogously, whether *two* and *six* referred to the partitioned or unpartitioned set. Finally, we included garden path trials that were either semantically false (e.g. *You got all of the blue gumballs* with the symmetric display in Figure 1) or underinformative (e.g. *You got some of the orange gumballs* with the same display). The former served as a test that participants were in fact doing the intended task and were not just following a strategy of clicking on the gumballs whose color was mentioned in the statement. The latter served to establish whether implicature rates are comparable to Experiment 2.

How often the target set occurred on either side of the chamber was counterbalanced, as was the color of the target set and whether the target set was the big set or the small set (for *some* and *all*) and the partitioned or unpartitioned set (for *two* and *six*), yielding a total of 120 trials. Two versions of each display type were created to prevent an association of particular displays with particular quantifiers. Table 2 shows the distribution of trials over conditions.

Table 2: Number of trials in each condition. Target sets for *some/all* were either big or small, for *two/six* either unpartitioned or partitioned.

Trial type	Contrast	Target set	Quantifier			
			some	all	two	six
regular	present	big	8	8	8	8
		small	8	8	8	8
	absent	big/small	4	4	4	4
garden path	present	big	4	4	4	4
		small	4	4	4	4
	absent	big/small	2	2	2	2

Participants' eye movements were recorded with an SR Eyelink II head mounted eye-tracker with a sampling rate of 250 Hz. Auditory stimuli were presented over Sennheiser HD 570 head phones at a comfortable listening level.

Results and Discussion In the underinformative garden path trials there were 54% pragmatic responses, which is comparable to the implicature rates found in previous studies.

We fitted a series of mixed logit models (with random intercepts for subjects and items) to the subset of the data in the contrast condition on which participants were not already looking at the target at quantifier onset, in order to assess the probability of switching to the target after hearing the quantifier, given that the target set was either in the subitizing range or not. We further included control variables coding whether participants were looking at the target on the previous sample; trial number (to control for learning effects); participants' predominant response type (pragmatic or semantic); and non-linear effects of time.

Because we only had two set sizes in the lower chamber, participants quickly learned that *two* referred to the small set and *six* referred to the big set, resulting in a rapid increase in looks to the target regardless of whether number was used in the subitizing range.

Participants were faster to converge on the target for *all*

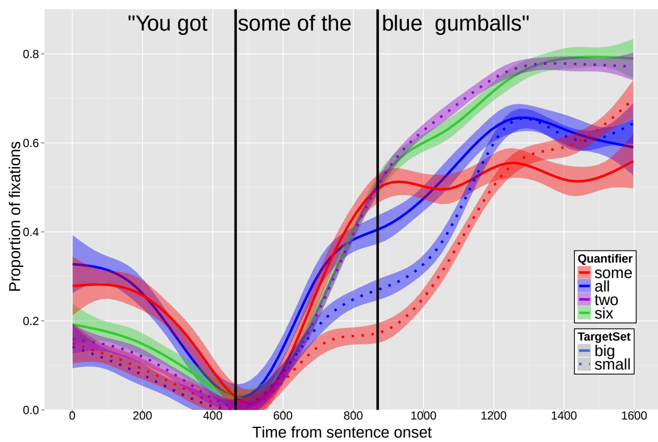

Figure 4: Looks to target for contrast trials on which there were no looks to target at quantifier onset. Ribbons represent 95% confidence intervals. Solid lines indicate big set, dotted lines small set target conditions.

than for *some* when the quantifier referred to the small set, but not when it referred to the big set (β = 1.12, *SE* = 0.56, $p < .05$, see Figure 4). Similarly, number was faster than *some* in the subitizing range, but not when *some* referred to the big set (β = 1.13, *SE* = 0.49, $p < .05$). Further, looks to big set *all* increased faster than to small set *some* (β = 0.75, *SE* = 0.36, $p < .05$), replicating the finding from Huang and Snedeker (2009).

We conclude that *some* may give rise to a rapid implicature outside the subitizing range (indeed, as fast as number terms within the subitizing range), but is delayed in the presence of salient lexical alternatives within the subitizing range.

General Discussion

We identified three linguistic and extra-linguistic constraints that comprehenders are sensitive to in processing scalar implicatures. The syntactic form of the *some*-NP affects implicature rate but not response times. The availability of more natural lexical alternatives inhibits computation of both the upper-bound and lower-bound meaning of *some*. And subitizing directly affects the naturalness of *some*.

These results have both methodological and theoretical consequences. Methodologically, it is important to either control for or explicitly model the effect of set size when using visual-world eye-tracking to study scalar implicature. This sheds new light on the findings from Huang and Snedeker (2009) and Grodner et al. (2010).

Theoretically, our results show that scalar implicatures may be computed with the speed associated with default inferences if strongly supported by probabilistic constraints, but may require the processing effort associated with typical PCIs if constraints are weak or in competition. Thus scalar implicature does not fit neatly into the GCI-PCI dichotomy.

Acknowledgments

This work was partially supported by NIH grant HD 27206. We thank Dana Subik for running participants, Daniel Pontillo for help with stimulus creation and Patricia Reeder for being a competent voice.

References

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434 - 463.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance xii: The psychology of reading*. Lawrence Erlbaum Associates.

Grice, P. (1975). Logic and conversation. *Syntax and Semantics*, *3*, 41-58. (Reprinted in Grice 1989, pp. 22-40)

Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "some", and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42 - 55.

Hirschberg, J. (1991). *A theory of scalar implicature*. New York: Garland.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference. In D. Schiffrin (Ed.), *Meaning, form and use in context: Linguistic applications, proceedings of gurt 84* (p. 11 - 42). Washington D.C.: Georgetown University Press.

Huang, Y., & Snedeker, J. (2009). On-line interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*, 376-415.

Katsos, N., Breheny, R., & Williams, J. (2005). The interaction of structural and contextual constraints during the on-line generation of scalar inferences. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society.*

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*(4), 498 - 525.

Levinson, S. C. (2000). *Presumptive meanings - the theory of generalized conversational implicature*. The MIT Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676 - 703.

Noveck, I., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language*, *85*, 203-210.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285 - 318.