COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 172–201
Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12227

Availability of Alternatives and the Processing of Scalar Implicatures: A Visual World Eye-Tracking Study

Judith Degen,[a] Michael K. Tanenhaus[b]

[a]*Department of Psychology, Stanford University*
[b]*Department of Brain and Cognitive Sciences, University of Rochester*

Received 20 December 2012; received in revised form 9 October 2014; accepted 19 October 2014

Abstract

Two visual world experiments investigated the processing of the implicature associated with *some* using a "gumball paradigm." On each trial, participants saw an image of a gumball machine with an upper chamber with orange and blue gumballs and an empty lower chamber. Gumballs dropped to the lower chamber, creating a contrast between a partitioned set of gumballs of one color and an unpartitioned set of the other. Participants then evaluated spoken statements, such as "You got some of the blue gumballs." Experiment 1 investigated the time course of the pragmatic enrichment from *some* to *not all* when the only utterance alternatives available to refer to the different sets were *some* and *all*. In Experiment 2, the number terms *two, three, four, and five* were also included in the set of alternatives. Scalar implicatures were delayed relative to the interpretation of literal statements with *all* only when number terms were available. The results are interpreted as evidence for a constraint-based account of scalar implicature processing.

Keywords: Pragmatics; Scalar implicature; Quantifiers; Alternatives; Eye-tracking

1. Introduction

Understanding the remarkable speed and efficiency of language comprehension has been an important goal of research in psycholinguistics. One class of explanation attributes the speed of processing to encapsulated, autonomous subsystems or modules, which perform specialized computations, much like a worker on an assembly line (Fodor, 1983; Forster, 1979). The modularity hypothesis dovetailed nicely with evidence that rapid and effortless cognitive processes are automatic and inflexible, whereas more flexible strategic processing is slow and effortful (Neely, 1977; Posner & Snyder, 1975; Shiffrin & Schneider, 1977).

Correspondence should be sent to Judith Degen, Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305. E-mail: jdegen@stanford.edu

Throughout the 1980s and 1990s, the modularity hypothesis was contrasted with constraint-based approaches, which propose that comprehenders rapidly integrate multiple constraints within and across different types of linguistic representation. Syntactic ambiguity was an important test domain: Is ambiguity resolved by domain-specific processes or by appeal to multiple sources of information, including the general context of an utterance? Initial results seemed to provide striking evidence that listeners make initial commitments without taking into account context (Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983; for review see Frazier, 1987). However, as ideas about probabilistic constraints, including those provided by context, became more refined, evidence that listeners rapidly integrate multiple constraints emerged (MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994; for reviews, see Altmann, 1998; Gibson & Pearlmutter, 1998; Jurafsky, 1996; Levy, 2008; Tanenhaus & Trueswell, 1995).

A competing explanation for the speed and efficiency of language processing has its roots in information-theoretic approaches to language. The hypothesis is that rapid use of information from context allows language to efficiently use reduced and otherwise underspecified or ambiguous forms (Zipf, 1949). The general idea is that it is cognitively and articulatorily expensive to make information explicit in language, but relatively cheap for listeners to make use of readily available information from context (see also Horn's division of pragmatic labor; Horn, 1984). This idea has recently undergone a renaissance. For example, Jaeger and colleagues (Jaeger, 2010; Qian & Jaeger, 2012) demonstrate that aspects of language production follow principles that derive from communicative efficiency. Speakers maximize the degree to which information is distributed uniformly across an utterance, following the principle of Uniform Information Density (Levy & Jaeger, 2007).

Recently, Piantadosi, Tily, and Gibson (2012) have demonstrated that when context is informative, unambiguous utterances are partially redundant with the context and therefore inefficient. From this perspective, ambiguity in language is not a problem that complicates language processing. Instead, ambiguous systems are desirable because they allow for re-use of words and sounds that are easy to produce and understand.

The communicative efficiency approach depends on the assumption that listeners rapidly make use of information from context. Indeed, rapid use of context is not limited to syntactic and lexical ambiguity resolution. Addressees generate expectations about the domain of subsequent reference (Altmann & Kamide, 1999) and rapidly circumscribe referential domains using information including referential context (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus & Trueswell, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999), presuppositions and affordances relevant to intended actions (Chambers, 2002; Chambers, Tanenhaus, & Magnuson, 2004), and differences in their own perspective and that of their interlocutors (Barr, 2008; Brown-Schmidt, 2009a,b; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002; but cf. Keysar, Barr, Balin, & Brauner, 2000; Kronmüller & Barr, 2007; Wu & Keysar, 2007).

The processing of *scalar* inferences that arise in parts of discourse as in (1) would seem to be a prominent counter-example to the communicative efficiency approach.

(1) You got some of the gumballs.
↳ You got some, but not all, of the gumballs.

The semantic content of the utterance in (1) is that the listener got at least one gumball (lower-bound meaning of *some*). Under appropriate contextual conditions, however, the listener is also justified in taking the speaker to implicate that he, the listener, did not get all of the gumballs (upper-bound meaning of *some*). Arriving at the upper-bound meaning of *some,* also referred to as *pragmatic some*, is considered a scalar implicature. The term *scalar* is used because the inference that gives rise to the upper-bound interpretation is assumed to depend upon a commonly used, highly accessible set of alternatives (the scale) that the speaker could have selected but did not (e.g., *most*, *all*).

Relying on implicature to convey the upper-bound meaning, rather than making it explicit or having a lexicalized alternative such as “some-not-all,” seems consistent with the communicative efficiency hypothesis. However, there is a body of evidence suggesting that even the most basic implicatures might be slow and costly for listeners. Before turning to this evidence, we briefly discuss a proposal by Levinson (2000) that set the stage for much of the current research on scalar implicature.

Levinson (2000) assumed that most inference is relatively slow and costly. However, he proposed that making frequently occurring default inferences cognitively cost-free provides a solution to the problem of the articulatory bottleneck, that is, the problem that the amount of information that can be phonetically encoded per unit time is small, and yet communication proceeds rapidly. Levinson argued that inferences that are commonly assumed to arise virtually independently of context, namely *generalized conversational implicatures,* are an example of automatic, cost-free inference. Thus, in contrast to other context-based accounts which naturally treat context-supported inferences as part and parcel of real time interpretation, Levinson’s Default account (which assigns scalar implicatures the status of generalized conversational implicature) assumes that upper-bound *some* is precompiled.

Both the default and context-based accounts are challenged by a growing body of evidence showing that at least under some conditions, scalar inferences take longer to compute than literal interpretations (Bott, Bailey, & Grodner, 2012; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009, 2011). These results have been interpreted as evidence for a Semantics- or Literal-First model of scalar implicature processing (Huang & Snedeker, 2009), where the pragmatic interpretation of *some* is assumed to follow the semantic interpretation.

Most of these studies present sentences containing scalar items like *some* and *or* without strong supporting contexts (but c.f. Breheny et al., 2006). For example, many studies use sentence verification paradigms in which participants are presented with sentences of the type *Some elephants are mammals* and are asked to judge whether the sentence is true or false. If the sentence is interpreted semantically as *At least one elephant is a mammal*, it is true. If it is interpreted pragmatically as *At least one, but not all, elephants are mammals*, it is false. Pragmatic responses are generally slower than semantic responses. This is taken as evidence that scalar inferences are slow and costly. However, these results are not necessarily inconsistent with claims about rapid use of context and communicative efficiency. In these situations the main component of accounts that predict rapid contextual inference is absent: There is no (experimentally manipulated or controlled for) context that the sentence is interpreted relative to. Thus, it is possible that the inference is not in fact predictable in the same way it would be in more natural discourse.

Huang and Snedeker (2009, 2011) is a notable exception. Participants viewed a display with four quadrants, with the two left and the two right quadrants containing pictures of children of the same gender, with each child paired with objects. On a sample trial, the two left quadrants might each contain a boy: one with two socks and one with nothing. The two right quadrants might each contain a girl: one with two socks (pragmatic target) and one with three soccer balls (literal target). A preamble established a context for the characters in the display. In the example, the preamble might state that a coach gave two socks to one of the boys and two socks to one of the girls, three soccer balls to the other girl, who needed the most practice, and nothing to the other boy.

Participants followed instructions such as *Point to the girl who has some of the socks*. Huang and Snedeker (2009) reasoned that if the semantic (literal) interpretation is computed prior to the inference, then, upon hearing *some*, participants should initially fixate both the semantic and pragmatic targets equally because both are consistent with the literal interpretation. If, however, the pragmatic inference is immediate, then the literal target should be rejected as soon as *some* is recognized, resulting in rapid fixation of the pragmatic target. The results suggested that the inference was costly: For commands with *all* (e.g., *Point to the girl who has all of the soccer balls*) and commands using number (e.g., *Point to the girl who has two/three of the soccer balls*), participants converged on the correct referent 200–400 ms after the quantifier. In contrast, for commands with *some,* target identification did not occur until 1,000–1,200 ms after the quantifier onset. The authors concluded that “even the most robust pragmatic inferences take additional time to compute” (Huang & Snedeker, 2009, p. 408).

In a very similar study, however, Grodner, Klein, Carbary, and Tanenhaus (2010) found evidence for rapid interpretation of pragmatic *some*. Each trial began with three boys and three girls on opposite sides of the display and three groups of objects in the center. A pre-recorded statement described the total number and type of objects in the display. Objects were then distributed among the boys and girls. A pre-recorded instruction, of the form, *Click on the girl who has summa/nunna/alla/the balloons,* told the participant who to click on (using the computer mouse).

Convergence on the target for utterances with *some* was just as fast as for utterances with *all*. Moreover, for trials on which participants were looking at the character with all of the objects at the onset of the quantifier *summa*, participants began to shift fixations away from that character and to the character(s) with only some of the objects (e.g., the girl with the balls or the girl with the balloons) within 200–300 ms after the onset of the quantifier. Thus, participants were immediately rejecting the literal interpretation as soon as they heard *summa.*

If we compare the results for *all* and *some* in the Huang and Snedeker and Grodner et al. experiments, the time course of *all* is similar but the upper-bound interpretation of partitive *some* is computed 600–800 ms later in Huang and Snedeker. Why might two studies as superficially similar as Huang and Snedeker (2009) and Grodner et al. (2010) find such dramatically different results? The Grodner et al. (2010) design and analysis differed from Huang and Snedeker (2009) in several ways. First, the pre-recorded statement in Grodner et al. drew participants' attention to the total cardinality of each set of objects. Second, there were some potentially important differences the auditory stimuli used by Grodner et al.; participants were asked to click on the target items (rather than point) and "some of" was pronounced *summa* to avoid potential ambiguity with non-partitive "some." Grodner et al. also included a "none of "(*nunna*) condition. Third, Grodner et al. corrected for a baseline preference for participants to look at the character with more objects that appears in all of the experiments reported in Huang and Snedeker and Grodner et al., though the bias was only reliable in some of the Huang and Snedeker experiments. Although these factors might account for a small difference in timing between the Huang and Snedeker and the Grodner et al. experiments, it seems unlikely to us that even a combination of these factors could account for a difference as large as 600–800 ms.

Our proposal is based on the observation that utterances are always interpreted against a backdrop of alternative utterances that the speaker could have produced but did not. In the literature on scalar implicature processing, *some* is assumed to compete with *all*, but not with other alternative utterances. This assumption stems from the linguistic literature, where scalar implicatures have been argued to arise in virtue of competition between a weaker scalar item (e.g., *some*) and its stronger alternative (e.g., *all*). The idea that elements from the <all, some> scale might compete with elements from other scales (e.g., the number scale) is not considered, presumably because such considerations aren't necessary for deriving scalar implicatures. However, the fact that other scales aren't necessary for deriving basic scalar implicatures formally does not entail that they do not play a role in the *processing* of scalar implicatures. Indeed, if one considers the task that the listener is faced with—inferring the state of the world that a speaker is trying to convey from that speaker's utterance in a specific context—the role that alternatives might play begins to look very different. Speakers have many different ways of referring to approximate or absolute set sizes, for example by using vague quantifiers like *some* and *most*, context-dependent quantifiers like *few* and *many*, and number terms like *two* and *five*. Recent work in probabilistic pragmatics has provided evidence that in interpreting utterances (including quantifiers), listeners probabilistically reason about a model of the speaker that crucially includes alternative utterances the speaker could have produced, to infer that speaker's intended meaning (Bergen, Goodman, & Levy, 2012; Degen, Franke, & Jäger, 2013; Goodman & Stuhlmüller, 2013). These models have thus far not been extended to make predictions for online processing. However, the underlying idea that listeners simulate the utterances speakers might produce in a specific context provides a novel perspective on the processing of implicatures, which, in our view, provides a principled explanation for the differences between Huang and Snedeker (2009) and Grodner et al. (2010).

We propose that the differences between Huang and Snedeker (2009) and Grodner et al. (2010) in the time course of processing the implicature from *some* to *not all* is at least in part due to the different sets of alternatives that participants in those experiments were led to believe were available to speakers. In particular, there were two crucial differences. First, Huang and Snedeker included stimuli with numbers, whereas Grodner et al. did not. In Huang and Snedeker (2009) a display in which one of the girls had two socks was equally often paired with the instruction, *Point to the girl who has some of the socks* and *Point to the girl who has two of the socks*. Second, Huang and Snedeker always paired *some* with a smaller set (of size 2) and *all* with a larger set (of size 3). We elaborate on each of these in turn.

Why might instructions with exact number have such a large impact on upper-bound interpretations of partitive *some*? One possibility is that in the absence of number instructions, participants might have adopted a top–down strategy of associating characters with a subset of objects with the expression *some*. This strategy would be possible in Grodner et al. because there was a particular type of display that was predominantly associated with a *summa* instruction. Although we cannot rule out a pre-coding account, we are skeptical for two reasons. First, and most specifically, it assumes that Grodner et al.'s participants learned the mapping between the display and use of *summa*. However, Grodner et al. (2010) found no evidence that the patterns in their results changed across the experiment. Second, if participants had learned that certain displays were associated with *summa*, then one would expect anticipatory eye movements to the expected target (Altmann & Kamide, 1999; Revill, Tanenhaus, & Aslin, 2008). However, inspection of the proportion of fixation plots presented in Grodner et al. (2010) does not show any hint of anticipatory effects.

Another explanation for the differences between the two studies is that the delays found by Huang and Snedeker are due to listeners' expectations about quantifier use for different absolute set sizes rather than (or in addition to) whether or not they are partitioned. In Huang and Snedeker's stimuli, *all* was always used to refer to the larger set, while *some* was always used to refer to the smaller set. If, in the presence of number term alternatives, the expectation for *all* to be used with the larger set was greater than the expectation for *some* to be used with the smaller set, this relative asymmetry in expectation of use—and the resulting difference in the severity of the expectation violation—may have led to the delay for *some* relative to *all* observed by Huang and Snedeker. Indeed, previous studies have established that the naturalness or typicality—which can be interpreted as a measure of listeners' expectations of use—of *some* to refer to sets of different sizes varies, especially for very small sets (Degen & Tanenhaus, 2014; Van Tiel, 2013).

A final, not yet mentioned, explanation for the differences between the two studies may be that the populations of participants differed between the studies. Previous studies that measured participants' responses to underinformative sentences revealed variation in the proportion of participants who interpreted utterances like *Some elephants are mammals* pragmatically as *some but not all elephants are mammals* versus semantically as *Some and possibly all elephants are mammals* (Bott & Noveck, 2004). It is possible that participants who tend to respond semantically process the quantifier in its context differently from participants who tend to respond pragmatically. For example, pragmatic responders may be faster to enrich *some* to *some but not all* or be more susceptible to contextual factors like the availability of alternatives than semantic responders. Neither Huang and Snedeker nor Grodner et al. had an independent measure of participants' interpretation of *some*. It is thus possible that Grodner et al.'s study included a mostly pragmatic population while Huang and Snedeker's included a mostly semantic population, which may have resulted in faster processing of *some* in Grodner et al. than in Huang and Snedeker. This could have been due to differences in procedure, for example, Grodner et al. emphasized the upper bound by drawing attention to the cardinality might have encouraged pragmatic responses or simply due to individual differences.

In the current paper we directly test the hypothesis that instructions with exact number can delay processing of upper-bound *some,* thus providing a plausible account for the differences between the Huang and Snedeker (2009) and Grodner et al. (2010) results. Simple modifications of the Huang and Snedeker and Grodner procedures would not have allowed us to evaluate our hypotheses. Therefore, our goal was to provide a functional rather than an exact replication of Grodner et al. and Huang and Snedeker's conditions in a paradigm that would naturally let us manipulate set size, use of numbers, and responder type. To this end we modified the gumball paradigm introduced by Degen and Tanenhaus (2014) and described below, to test whether the precoding hypothesis or the expectation for number alternatives better accounts for whether or not processing of upper-bound *some* is delayed. The modified gumball paradigm is functionally similar to the setups used by Huang and Snedeker and Grodner et al. but avoids the confounding factors present in each. In particular, (a) quantifier use could not be predicted from the display (in contrast to Grodner) and (b) *some* and *all* were used with both larger and smaller sets (in contrast to Huang and Snedeker). In addition, (c) an independent measure of responder type was obtained by including trials on which participants responded to underinformative sentences.

The stimuli differed from Grodner's and Huang and Snedeker's in both object type and (in part) absolute set size. But the participant's task was functionally identical: to identify a set of objects by indicating a region of the display described by an utterance which used either *some* or *all* (or the number terms *two* to *five*) to refer to that set.

The functional replication of Grodner's setup was achieved by not including number alternatives to refer to the target set (Experiment 1). The replication of Huang and Snedeker's setup was achieved by including number terms and varying set size (Experiment 2). We also include exploratory analyses of potential response population differences.

2. Experiment 1

Experiment 1 tested the time course of listeners' scalar implicature processing in the absence of number terms. The main question of interest was whether computing the implicature from *some* to *not all* would incur a processing cost compared to the literal content of *all*. In order to avoid both the potential confounds in Grodner et al. (2010) and Huang and Snedeker (2009), each quantifier was used with different display types; and *some* and *all* were used to refer to both bigger sets of objects (four or five objects) as well as smaller ones (two or three objects).

2.1. Methods

2.1.1. Participants

Forty undergraduate students from the University of Rochester were paid $10 to participate. All were native speakers of English.

2.1.2. Procedure

On each trial, participants saw a display of a gumball machine with an upper chamber filled with 8 gumballs of one color of gumballs and either 2, 3, 4, or 5 gumballs of the other color. The lower chamber was empty. After 2 seconds, the button in the center of the machine started flashing. Participants then clicked on the button. Upon clicking, a gray mask was displayed for 200 ms. Then the gumball machine was redisplayed with a certain number of gumballs of each color having dropped to the lower chamber (see

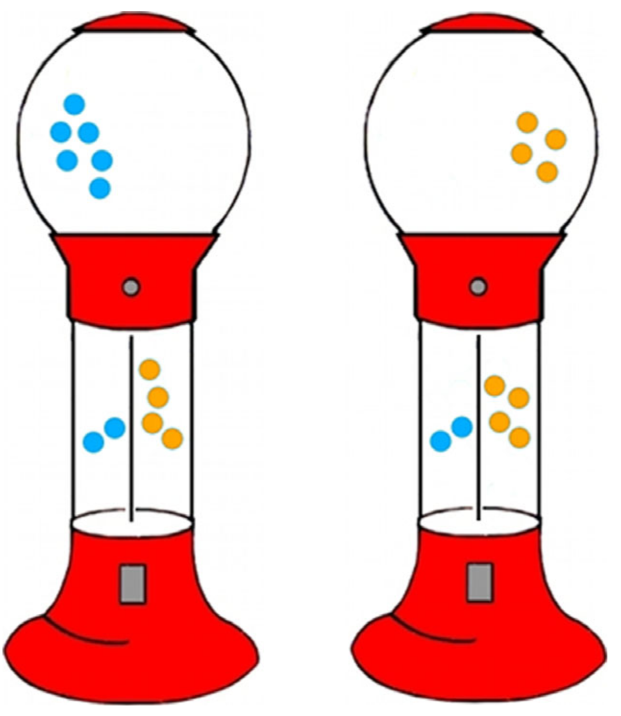

Fig. 1. Example displays that contain the same contrast between set sizes in the lower chamber but differ in whether the big or small set is partitioned. Display on the left could occur with sentences *You got some/two of the blue gumballs* or *You got all/four of the orange gumballs*, while the display on the right could occur with *You got all/two of the blue gumballs* or *You got some/four of the orange gumballs* in the *early* condition.

Fig. 1 for example displays). Clicking the central button ensured that all participants were looking at a central fixation point at the time of auditory stimulus onset. After 500 ms, participants heard an auditory stimulus of the form *You got some of the blue gumballs*. Their task was to click on the side of the lower chamber that contained the gumballs mentioned in the statement if they thought the statement was true, and click on the central button otherwise. Once they clicked either on a side of the lower chamber or on the central button, a gray screen was displayed for 1 s and the experiment advanced to the next trial.

Participants first received a short tutorial explaining the task. They then completed four practice trials before beginning the experimental trials. The total experiment consisted of 64 trials.

Participants' eye movements were recorded with an SR Eyelink II head-mounted eye-tracker with a sampling rate of 250 Hz. Drift correction was performed every five trials. Auditory stimuli were presented over Sennheiser HD 570 headphones at a comfortable listening level.

2.1.3 Materials

Each auditory stimulus was a sentence of the form *You got Q of the C gumballs*, where *Q* was one of the quantifiers *some* and *all* and *C* was one of the color adjectives *blue* and *orange*. Stimuli were cross-spliced in the following way. Each sentence was recorded individually. The quantifier and color adjective were subsequently spliced out of the original recording and onto a recording of the sentence *You got most of the green gumballs*, where *most* and *green* were replaced by the experimental items. This ensured that quantifier onset was equal across all stimuli (at 462 ms) and the onset of the color adjective was the same within each quantifier (*all*: 847 ms, *some*: 856 ms). Mean sentence length was 1,844 ms.

Visual stimuli were constructed to yield *early*, *late*, and *garden-path* trials. On *early* trials there was always a contrast between a larger and a smaller set in the lower chamber. Set size was manipulated to test whether there was a bias against the use of *all* and *some* with very small sets. If the smaller set was of size 2, the larger one was 4. If the smaller set was of size 3, the larger set was 5. One of the two sets was always partitioned. The set that was partitioned originally started out as a set of 8 in the upper chamber. We refer to trials where the smaller of the two sets of gumballs was the target as *small* trials and trials where the larger set was the target as *big* trials. The left panel of Fig. 1 shows a display that could have been used on an *early small some* trial or on an *early big all* trial. The right panel shows a display that could have been used on an *early big some* or *early small all* trial. The *early* trials were the trials of most interest because they were the ones on which the time course of the implicature from *some* to *not all* would emerge. The faster the implicature is computed, the earlier after the onset of *some* participants' looks should converge on the target set (the partitioned set). On *early all* trials, the quantifier is immediately disambiguating and participants should be able to quickly converge on the target set.

Late trials were included as a baseline against which to test whether participants used the information from the quantifier at all as a cue to the target before receiving the disambiguating information from the color adjective. On *late* trials, there were either two partitioned sets (for *some*) or two unpartitioned sets (for *all*) of the same size in the lower chamber. Each of the two quantifiers occurred with each set size (2, 3, 4, 5) twice. The quantifier thus did not disambiguate the target set of gumballs and participants had to wait until they heard the color adjective (*orange* or *blue*) further downstream to identify the target. If participants use the information from the quantifier to compute the implicature upon observing *some*, looks to the target should start to increase before observing the adjective in the *early* condition, where the quantifier can be used as a cue to the target, but not in the *late* condition. If, however, participants initially interpret *some* semantically, that is to mean, *some and possibly all*, then the point of disambiguation (the adjective) would be the same for the *early* and the *late* conditions.

The *garden-path* trials employed the same displays as *early* trials and were included in order to categorize participants as more pragmatic or more semantic responders. In the *garden-path some* condition, an underinformative statement like *You got some of the orange gumballs* occurred, for example, with the left display in Fig. 1. Such a statement is false if interpreted pragmatically as *You got some, but not all, of the orange gumballs,* but true if interpreted semantically as *You got at least one of the orange gumballs*. These trials also served as a further test for whether the inference from *some* to *not all* is delayed. An increase in looks to the partitioned set of gumballs before the switch to the gumballs compatible with the downstream adjective would constitute evidence for early implicature computation. We also included garden-path *all* trials, on which the statement was semantically false. That is, in a situation like the left panel of Fig. 1, participants might have heard *You got all of the blue gumballs*. In the instructions participants were told to click the central button on the machine if the statement they heard did not refer to a set of gumballs in the display. Thus, clicks of the central button served a dual analysis purpose: (a) as a basis for excluding participants who were not paying attention (as evidenced by a lack of clicks on the central button in the *garden-path all* condition, where the statement was semantically false); and (b) as a way of categorizing participants' responses to *some* as pragmatic (if they clicked on the central button) or semantic (if they clicked on the half of the chamber that contained the gumballs compatible with the color adjective).

Target set color (orange, blue) and location (left, right) were each counterbalanced within quantifier condition to discourage participants from forming target associations with either color or location cues. For the quantifiers *some* and *all,* target set size was counterbalanced; that is, *some* and *all* were used equally often to refer to a set of size 2, 3, 4 or 5. For each unique display type (i.e., for each unique combination of color and set size of left and right set of gumballs), two versions were created to further discourage participants from adopting display-type-based looking strategies. Importantly, each quantifier occurred with each unique display type. It was therefore not possible for participants to pre-code display-to-form mappings to predict whether they would hear an *all* or a *some* statement.

The gray shaded cells in Table 1 show the number of trials each participant saw in each condition in Experiment 1. Trial order was randomized for each participant.

Table 1
Distribution of trials over point of disambiguation (POD), target set size, and quantifier conditions

		Quantifier						
POD	Target	*Some*	*All*	*Two*	*Three*	*Four*	*Five*	Total
Early	Small (2/3)	8	8	4	4			24
	Big (4/5)	8	8			4	4	24
Late	Small (2/3)	4	4	2	2			12
	Big (4/5)	4	4			2	2	12
Garden-path	Small (2/3)	4	4	2	2			12
	Big (4/5)	4	4			2	2	12
Total		32	32	8	8	8	8	96

Note. Participants in the numbers *present* (Experiment 2) condition saw all 96 trials, in the numbers *absent* (Experiment 1) condition only the *some* and *all* trials.

2.2. Results and discussion

We report the results in the following order. First, we present eye movement analyses addressing the two main questions of interest. First, is the implicature from *some* to *not all* computed on the basis of information from the quantifier, before information from the adjective becomes available? If listeners compute the implicature based on the quantifier, we should observe more early looks to the target for *early some* compared to *late some*. Second, if the implicature is computed as the quantifier unfolds, does computing the implicature incur a processing cost compared to computing the literal content of *all*? Next, we present the distribution of semantic and pragmatic responders and the differences in the two sub-populations' eye movements. Finally, we present an analysis of response times from the *garden-path some* condition that addresses whether pragmatic responses incur a processing cost compared to semantic responses.

Two percent of the data were excluded because response time as measured from auditory stimulus onset was greater than three standard deviations from mean response time in each of the *early*, *late*, and *garden-path* conditions, respectively.[1] Exclusion of these trials here and in Experiment 2 was motivated by the possibility that on these trials, where participants took an inordinate amount of time to respond, they were not attending to the task in the same way as on trials where response times were closer to the mean, which may have also resulted in abnormal early eye movement patterns.

2.2.1. *Eye movement analyses: Is there an* early *versus* late *effect?*

Is the implicature computed on the basis of information from the quantifier, before information from the disambiguating noun becomes available? Visual inspection of Fig. 2[2] suggests that it is: For both *big* and *small* set *some* and *all*, looks to the target in the *early* condition begin to increase in the quantifier window compared to baseline, while in the *late* condition looks to the target do not increase until after adjective onset. Here and below, statistical results were obtained by fitting mixed-effects logistic

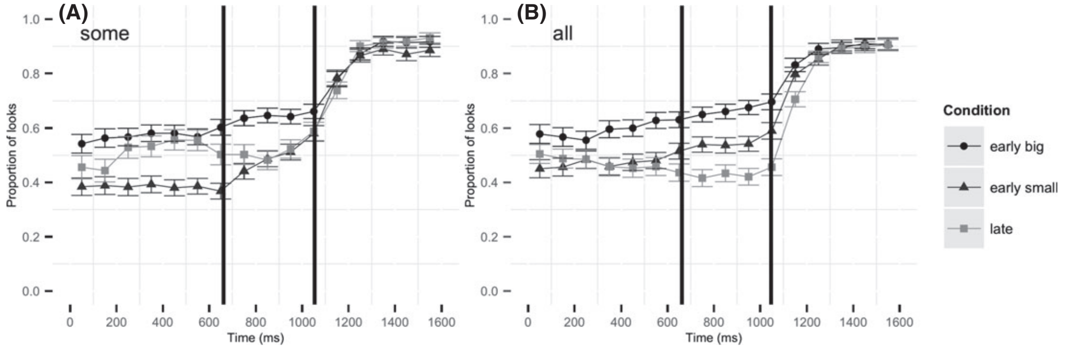

Fig. 2. Proportion of looks to the target by condition for *some* (A) and *all* (B) in the *early* (black) versus late (*gray*) conditions. Black vertical lines demarcate the quantifier window. *X*-axis shows time in ms relative to auditory stimulus onset. Here and below, error bars indicate 1 SE.

regression models predicting looks to the target over looks to the target and competitor. All models contained the maximal by-participant random effects structure, and all fixed effects predictors were centered before entering the analysis.

To determine whether participants were using information from the quantifier in the *early some* but not in the *late some* condition, we fit one model each to the *early/garden-path*[3] and *late* subsets of the data. Target looks were predicted from a predictor coding time window, where time window was either the *baseline* window (the 400 ms window ranging from quantifier onset minus 200 ms to quantifier onset plus 200 ms) or the *quantifier* window (the time window starting 200 ms after quantifier onset and lasting until 200 ms after adjective onset, the window in which looking patterns can plausibly first be driven by the observed quantifier). If participants were making use of the quantifier, there should be a significant effect of time window. The effect of time window was significant for the *early* condition ($\beta = 0.29$, $SE = 0.05$, $p < .0001$) but not for the *late* condition ($\beta = -0.10$, $SE = 0.08$, $p < .22$), suggesting that participants indeed made use of the information from the quantifier—that is, computed the implicature—before information from the adjective became available.

2.2.2. *Eye movement analyses: Is there a delay for* early some *compared to* early all*?*

Having established that the implicature is computed as the quantifier unfolds, we can ask whether participants' time to converge on the target set after hearing *some* (which requires computing the implicature to *not all*) is delayed compared to convergence time after *all*. In order to test this, we fit a mixed-effects logistic regression model to the *early* and *garden-path* quantifier window subset of the data, predicting target looks from fixed effects of quantifier type (*all* vs. *some*), target size (*big* vs. *small*), time (continuous, in 20 ms increments), and their interactions. If eye movement patterns differ over time for the two quantifiers, there should be a significant interaction of quantifier type and time. If they do not differ, there should only be a main effect of time. Target size was included as a control predictor and was of no theoretical interest in this experiment.

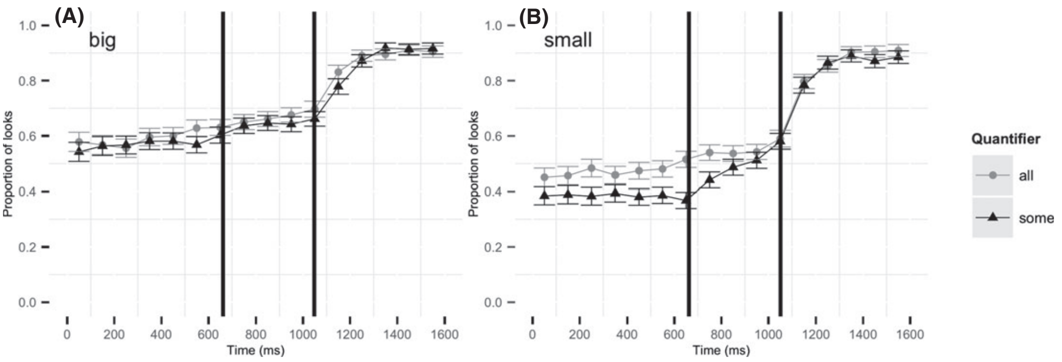

Fig. 3. Proportion of looks to the target by quantifier for *big* set (A) and *small* set (B) targets. Black vertical lines demarcate the quantifier window. *X*-axis shows time in ms relative to auditory stimulus onset.

Proportions of looks to the target in the *some* and *all* conditions are shown separately for *big* and *small* target sets in Fig. 3. There were only significant main effects of target size and time, such that participants were more likely to look at big sets rather than small sets in the quantifier window ($\beta = -0.59$, $SE = 0.11$, $p < .0001$) and as time increased the log odds of looking to the target increased ($\beta = 0.07$, $SE = 0.04$, $p < .05$).

Importantly, the main effect of time combined with the lack of effect for quantifier suggests that participants looked more toward the target in this window over time regardless of quantifier. That is, the information from the quantifier *some* was used before adjective onset to disambiguate the target, suggesting in turn that computing the implicature from *some* to *not all* in this experiment did not come at a higher cost or greater delay than integrating the information from the quantifier *all*.

This result replicates the finding of Grodner et al. (2010) and is inconsistent with the Literal-First hypothesis, which predicts a delay in the processing of *some* relative to *all.* Importantly, this result cannot be explained by appeal to display-based pre-coding in which partitioned sets are linked to *some* before the onset of the command. The same display types were used for both quantifiers. Therefore, participants could not use either the initial display in the upper chamber or the shifted display in the lower chamber to predict whether they would hear an *all* or *some* statement. Moreover, because of the garden-path trials, the partitioned set was sometimes paired with an instruction with *all* and the unpartitioned set was sometimes paired with an instruction with *some.* As a consequence, any strategy that associated the partitioned set with *some* would be ineffective.

A further check on the possibility that participants were adopting *some*-specific strategies is to test whether looking patterns change over the course of the experiment. We did this by adding a centered predictor coding *first* versus *second* half of the experiment to the model reported above. The experiment half predictor was allowed to interact with quantifier and target set size. A significant interaction between experiment half and quantifier would suggest that quantifier-specific strategies developed over the course of the experiment.

The main effects of target set size and time remained, and there was additionally a marginal main effect of half, such that participants were more likely to look at the target

Table 2
Distribution of participants over number of semantic responses given (out of 8 possible) in *garden-path some* condition

Number of semantic responses	0	1	2	3	4	5	6	7	8
Number of participants	20	6	1	3	6	0	1	1	2

in the second half than in the first half ($\beta = -0.28$, $SE = 0.16$, $p < .07$). However, there was no interaction of experiment half with quantifier, suggesting that participants did not adopt a *some*-specific strategy.

2.2.3. Participant population differences: Semantic versus pragmatic responders

The underinformative *garden-path some* trials were used to obtain information about participants' interpretation preferences. Recall that participants had the option of clicking the central button on the gumball machine if they believed the statement they heard was false. Responding to an underinformative utterance (*You got some of the orange gumballs* in a situation in which all the orange gumballs dropped to the lower chamber) by clicking on the central button thus constituted a *pragmatic* response, while responding by clicking on the unpartitioned set of gumballs constituted a semantic response.

There were 78.2% pragmatic responses and 21.8% semantic responses.[4] The distribution of participants over number of semantic responses given is shown in Table 2.

We categorized participants with fewer than four semantic responses as *pragmatic responders*, and participants with more than four semantic responses as *semantic responders.* There were 30 pragmatic responders but only four semantic responders. In addition, there were six *inconsistent* responders, that is, responders who gave 50% semantic and 50% pragmatic responses.

2.2.3.1. Looking behavior for semantic and pragmatic responders: Because of the preponderance of pragmatic responders, analyses comparing responder types are exploratory. Nonetheless they have the potential to be informative about possible differences between how pragmatic and semantic responders evaluate the description. Degen and Tanenhaus (2014) found that response times of participants who most often responded pragmatically on underinformative *some* trials were also more sensitive to the naturalness of alternatives than participants who most often responded semantically; in particular, the response time difference between responses to the more natural *some* versus the less natural *some of* on underinformative trials was greater for pragmatic than for semantic responders. This may indicate a heightened sensitivity of pragmatic responders to *alternative utterances,* to *pragmatic information* more generally, or to *linguistic information* even more generally. Since the current experiment did not contain alternatives to *some*, the first and second possibilities cannot be teased apart. However, if pragmatic responders are faster to integrate pragmatic information, they might be faster to derive the implicature than semantic responders but show no differences in how rapidly they converge on the target after observing *all*. If instead pragmatic responders are generally faster to integrate information from the linguistic signal, they should be faster than semantic responders to converge on the target after either quantifier.

To test for eye movement differences between semantic and pragmatic responders, we fit a model to the quantifier window data that was identical as the one reported in Section 2.2.1, except that a centered predictor coding responder type (*pragmatic* vs. *semantic*) was allowed to interact with the target size, quantifier, and time predictors. Inconsistent responders were excluded from the analysis.

The main effects of target size and time reported above remained significant. In addition, there was a main effect of responder type such that semantic responders were less likely to look to the target than pragmatic responders ($\beta = -0.39$, $SE = 0.19$, $p < .05$). In a model where responder type was additionally allowed to interact with experiment half, there was a marginally significant interaction between responder type and half, such that pragmatic responders looked more to the target in the second half than in the first, but there was no difference in looking behavior between experimental halves for semantic responders ($\beta = -0.77$, $SE = 0.43$, $p < .08$). Taken together, this suggests that the semantic responders were less apt to use the information from either quantifier to identify the target set of gumballs than the pragmatic responders. In addition, pragmatic responders became better at using the quantifier information as the experiment progressed, whereas semantic responders continued to rely on the adjective as the main cue to the target. This suggests that the pragmatic responders were generally faster than semantic responders to integrate information from the linguistic signal, regardless of whether that information was semantic or required pragmatic enrichment.

2.2.4. *Response time analysis*

Finally, we report an analysis of response times. While participants were not instructed to click on the target as quickly as possible (as is typically the case in studies that use response times to investigate the speed of scalar implicature processing, e.g., in truth-value judgment tasks), analyzing response times can nevertheless prove informative. In particular, if the "slow implicature" results found in many response times studies (Bott & Noveck, 2004; Bott et al., 2012) are replicated in this paradigm in the presence of early effects on eye movements, this would suggest that response times often reflect either greater uncertainty about the intended interpretation or increased verification time of the computed interpretation, rather than a delay in implicature computation.

For comparability with earlier studies, we analyzed response times only from the underinformative *garden-path some* condition. Of interest was whether pragmatic responses (clicks on the center of the gumball machine, reflecting that participants drew the implicature) were slower than semantic responses (clicks on the unpartitioned set in the lower chamber, reflecting that participants interpreted the utterance literally). Fig. 4 shows mean response times for semantic and pragmatic responses by responder type. Visual inspection of the graph suggests that pragmatic responders were slower to respond pragmatically than semantically, while there is a reverse trend for semantic responders.

This interaction is supported by a mixed-effects linear regression on the *garden-path some* subset of the data. The model predicted log-transformed response time from

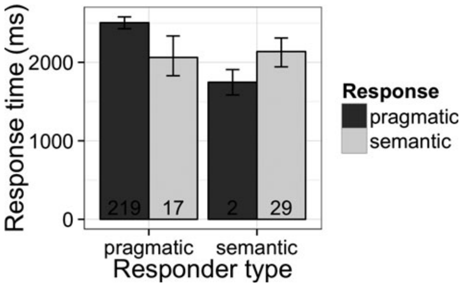

Fig. 4. Mean response times in the *garden-path some* condition for pragmatic versus semantic responses when given by pragmatic versus semantic responder types. Error bars indicate 95% bootstrapped confidence intervals. Numbers in bar indicate number of contributing observations.

centered fixed effects of response (*pragmatic* vs. *semantic*) and responder type (*pragmatic* vs. *semantic*, where entirely inconsistent responders were excluded from the analysis). The model included by-participant intercepts.

There was a significant interaction between response and responder type ($\beta = 0.32$, $SE = 0.17$, $t = 1.95$, $p < .05$) such that pragmatic responders' pragmatic responses were slower than their semantic responses, while there was no difference in semantic responders' response times. This replicates the "slow implicature" effects found in other response time studies, but only in pragmatic responders.

There are at least two alternative explanations for why pragmatic responses to *some* are slower than semantic responses. One explanation is that the interpretation of *some* as *some but not all* is associated with a more complex verification strategy than its literal interpretation because both the reference set (in the lower chamber) and the complement set (in the upper chamber) need to be verified and found to be non-empty, whereas the verification of the literal interpretation requires only a check that the reference set contains at least one gumball. Second, it is possible that the pragmatic interpretation is delayed due to a staged process of interpretation, consistent with the Literal-First hypothesis.

The eye movement data speak clearly against the second explanation: The increase in especially pragmatic responders' looks to the target in the quantifier window was similar after *some* and *all*. This suggests that participants began generating the implicature at the earliest moments after observing the quantifier. Therefore, the slower response times for pragmatic interpretations are likely due to increased verification time and not increased time to compute the implicature.

3. Experiment 2

Experiment 1 replicated the pattern of results reported by Grodner et al. (2010) under conditions where the design ruled out a display-specific pre-coding strategy. We found that pragmatic responders immediately computed an upper-bound interpretation of *some*. We also found that participants showed a strong bias to initially look at larger set sizes.

In Experiment 2, we introduce exact number descriptions to create a parallel to the conditions in Huang and Snedeker. This allows us to determine whether introducing explicit alternatives will slow down processing of upper-bound *some*. Crucially we can also compare processing of *some* and *all* in the subset of conditions where *some* refers to the smaller set and *all* refers to the larger set. These are the conditions under which Huang and Snedeker (2009) found delayed looks to the referent for *some* compared to *all*.

3.1. Methods

3.1.1. Participants

Forty undergraduate students from the University of Rochester were paid $10 to participate. All were native speakers of English who did not participate in the previous experiment.

3.1.2. Procedure and materials

The procedure was identical to Experiment 1 and the same materials were used, but Experiment 2 contained 32 additional number trials (shown in the non-shaded cells of Table 1). On these trials, the quantifier in the statement was either *two*, *three*, *four*, or *five*. Of these number trials, 16 were *early* trials, 8 were *late*, and 8 were semantically false *garden-path* trials. The same display types were used for each condition as in Experiment 1. For example, either of the displays in Fig. 1 could be used on early *two/four* trials. For number trials, target set partitioning was counterbalanced (half of the number trials had the number term referring to a partitioned, half to an unpartitioned set). This was intended to discourage participants from forming associations between number terms and whether or not a set of gumballs was partitioned or unpartitioned.

Auditory stimuli were cross-spliced as in Experiment 1. Quantifier onset was again 462 ms after sentence onset. Adjective onsets were the same within quantifier but differed between quantifiers: *two*: 815 ms, *three*: 894 ms, *four*: 902 ms, *five*: 909 ms. Trial order was pseudo-random, such that the number of *early*, *late*, and *garden-path* trials was counter-balanced within each quarter of the experiment, as was the number of *some*, *all*, and *number* trials and the number of *big* and *small* target set trials.

3.2. Results and discussion

We report the results in the following order. First, we again determine that the information from the quantifier was used in the *early*, but not in the *late* condition. Second, we present the eye movement analyses addressing the main question of interest: whether, in the presence of number terms, computing the implicature from *some* to *not all* incurred a processing cost compared to computing the literal content of *all*, specifically for big set *all* compared to small set *some*. This is exactly the condition under which Huang and Snedeker (2009) observed a delay. Next we present the distribution of semantic and pragmatic responders and the differences in the two sub-populations' eye movements. Finally, we include a response time analysis comparing the results of Experiments 1 and 2 to shed further light on how introducing number alternatives affects processing.

Data for three participants were excluded because they did not respond correctly on *all* and number term garden-path trials, where the statement was semantically false and required a click on the central button; instead of clicking the central button, these participants either clicked on the set of gumballs compatible with the quantifier or on the set of gumballs compatible with the adjective. This constituted 6.5% of the data. A further 1.5% of the data were excluded because response time as measured from auditory stimulus onset was greater than three standard deviations from mean response time in each of the *early*, *late*, and *garden-path* conditions, respectively.

3.2.1. Eye movement analyses: Is there an early *versus* late *effect?*

Proportion of looks to the target in the early and late conditions are shown for the different quantifiers and set sizes in Fig. 5. It is clear that while participants were very fast to converge on the target set after observing a number term in the *early* condition, the

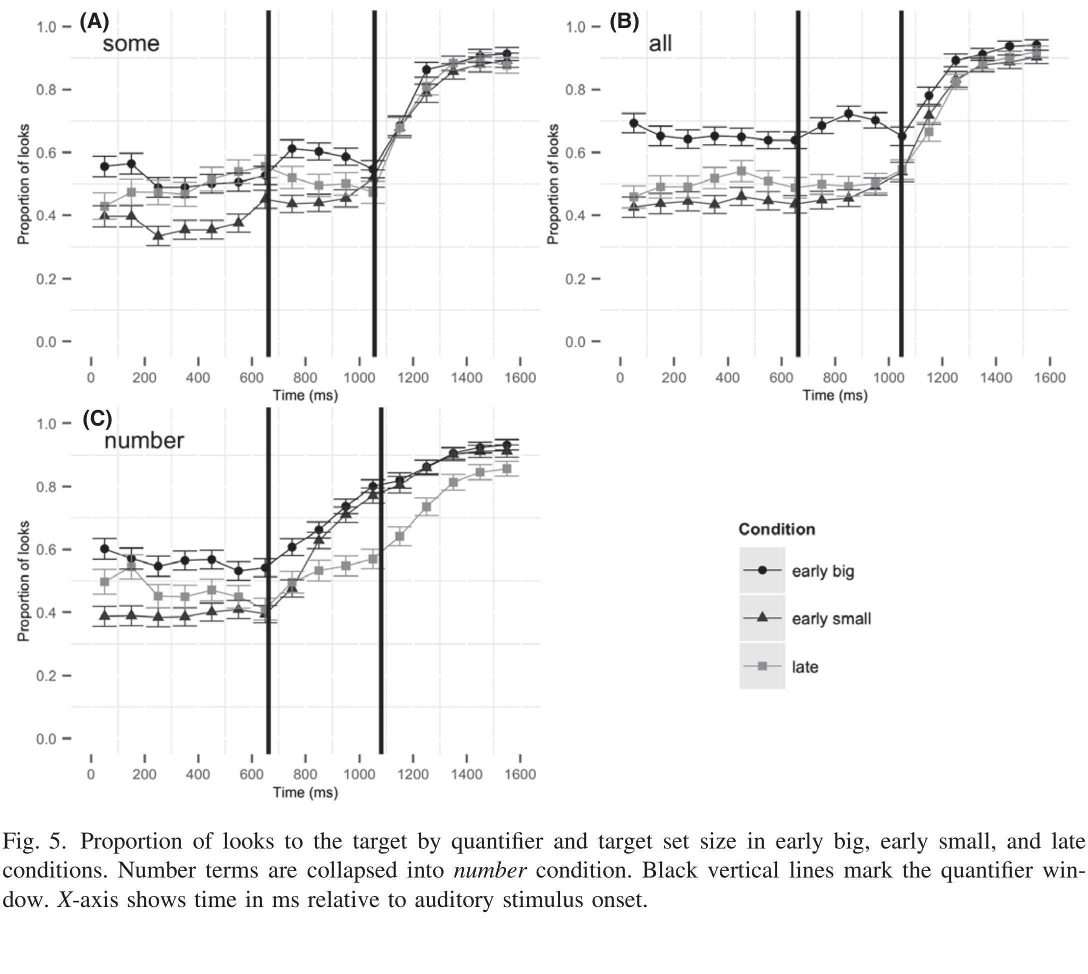

Fig. 5. Proportion of looks to the target by quantifier and target set size in early big, early small, and late conditions. Number terms are collapsed into *number* condition. Black vertical lines mark the quantifier window. *X*-axis shows time in ms relative to auditory stimulus onset.

difference between the *early* and *late* condition for both *some* and *all* is much smaller than on number trials.

To determine whether participants were using information from the quantifier in the *early* and *garden-path*[5] but not in the *late* condition, we fit one mixed-effects logistic regression model each to the *early/garden-path* and *late* subsets of the data. Target looks were predicted from a predictor coding time window, where time window was either the *baseline* window (the 400 ms window ranging from quantifier onset minus 200 ms to the start of the quantifier window) or the *quantifier* window. If participants were making use of the quantifier, there should be a significant effect of time window. As in Experiment 1, the effect of time window was significant for the *early* condition ($\beta = 0.42$, $SE = 0.05$, $p < .0001$) but not for the *late* condition ($\beta = 0.05$, $SE = 0.08$, $p < .53$), suggesting that participants were using the quantifier information when it could be used to disambiguate the target, but not when it could not.

3.2.2. Eye movement analyses: Is there a delay in computing the implicature?

Fig. 6 shows the functional replication of Huang and Snedeker (2009): Eye movements to the target are delayed after hearing *some* used with a *small* set (which requires implicature computation for target identification) compared to after hearing *all* used with a *big* set.

To investigate the difference in looking behavior between *some* and *all*, a mixed-effects logistic regression model was fit to the *early* and *garden-path* subset of the data in the quantifier window, predicting target looks from fixed effects of quantifier (*all* vs.

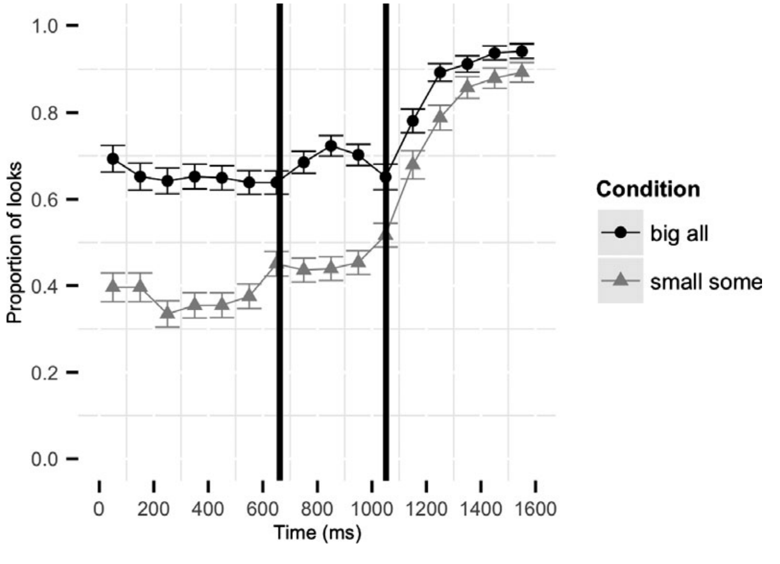

Fig. 6. Proportion of fixations to target in the *big* set *all* and *small* set *some* condition, the functional equivalent of Huang and Snedeker (2009)'s conditions. Black vertical lines mark the quantifier window. *X*-axis shows time in ms relative to auditory stimulus onset.

some), target size (*big* vs. *small*), time (continuous), and their interactions. As in the analyses of Experiment 1, this model and those reported below contained the maximal by-participant random effects structure.

There was a significant interaction between quantifier and target set size ($\beta = 0.19$, $SE = 0.07$, $p < .01$); inspecting the simple effects model revealed that there were overall more looks to the target after *all* than after *some* when the target set was big, but not when it was small. This is in contrast to Experiment 1, where only a main effect of target set size was observed. The target size main effect from Experiment 1 was replicated: Participants looked more to the target when it was a big set than when it was a small set ($\beta = -0.87$, $SE = 0.14$, $p < .0001$). Proportions of looks to the target in the early condition for big and small target sets by quantifier are shown in Fig. 7.

In sum, comparing these results to those from Experiment 1, we found that introducing number terms explicitly into the set of alternatives available to the speaker resulted in a delay for *small* set *some* compared to *big* set *all*. This replicates the pattern of results in Huang and Snedeker (2009) despite not nearly approaching the size of the delay found by those researchers (800 ms). However, in contrast to Huang and Snedeker, this study controlled for target set size, which allows us to determine whether there was a general delay for computing the implicature compared to computing the literal content of *all*, as the Literal-First hypothesis would predict. While there was a delay in computing the implicature compared to computing the literal content of *all* for the big set, no such delay was observed for the small set. Thus, there was no *general* delay in implicature computation.

We thus replicated Grodner et al.'s result—no delay in computing the implicature—in Experiment 1, even when pre-coding was not possible. We also replicated Huang and Snedeker's result—a delay in computing the implicature—in Experiment 2, but only when *all* was used with a big set. We defer further discussion of the implications of this result to the General Discussion and turn now to the exploratory analysis of the differences between semantic and pragmatic responders, in particular with respect to differences in the effect of introducing number terms into the set of experimentally available alternatives.

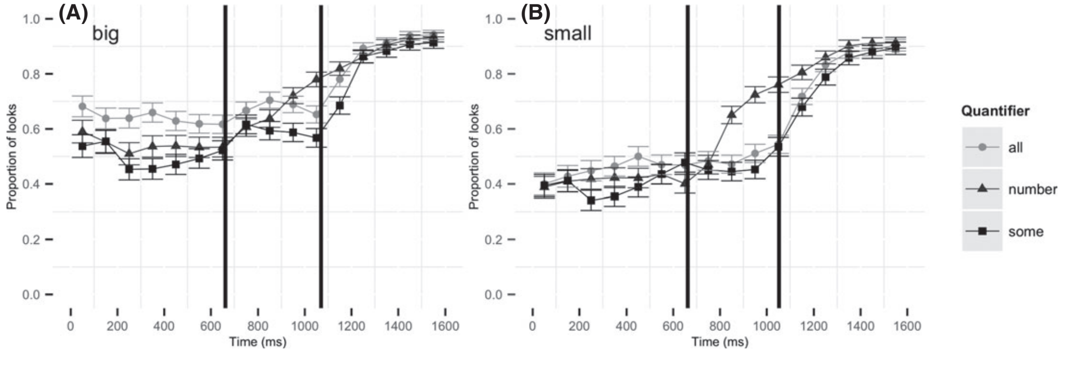

Fig. 7. Proportion of looks to target on *early* trials for the big set (A) and small set (B) trials, by quantifier condition. Vertical black lines indicate quantifier window.

Table 3
Distribution of participants over number of semantic responses given (out of 8 possible) in the *garden-path some* condition in Experiments 1 and 2

Number of semantic responses	0	1	2	3	4	5	6	7	8
Number of participants (Experiment 1)	20	6	1	3	6	0	1	1	2
Number of participants (Experiment 2)	14	6	4	1	1	1	0	3	7

3.2.3. Participant population differences: Semantic versus pragmatic responders

In the underinformative *garden-path some* trials, participants gave 64.9% pragmatic responses and 35.1% semantic responses (compared to 78.8% pragmatic responses and 21.2% semantic responses in Experiment 1). Table 3 shows how many participants made each number of semantic responses in Experiment 2, compared to Experiment 1.

As in Experiment 1, we categorized participants with fewer than four semantic responses as *pragmatic responders*, and participants with more than four semantic responses as *semantic responders.* A χ^2 test comparing Experiments 1 and 2 revealed that the distributions of responder types differed ($\chi^2(2) = 7.19$, $p < .05$): There are only twice as many pragmatic responders as semantic responders in Experiment 2 compared with seven times as many pragmatic responders in Experiment 1. That is, participants were overall more pragmatic in Experiment 1. This is an interesting result. It suggests that the availability of number terms decreases the probability of computing a scalar implicature. This dovetails with the predictions made by offline probabilistic models of scalar implicature and quantifier use that take into account the alternative utterances that a speaker could have produced (Frank & Goodman, 2012; Franke, 2014): If there are more (and more informative) alternatives the speaker could have used to refer to any one set size, the relative expectation for a speaker to use *some* is decreased for partitioned sets, where previously *some* was the only viable alternative. Thus, the expectation for *some* to be used with any set size (whether partitioned or not) is low in Experiment 2 compared to Experiment 1. One consequence of this is that the relative expectation violation upon observing an underinformative utterance of *some* used with an unpartitioned set is smaller in Experiment 2 compared to Experiment 1. Implicatures arise by virtue of there being an expectation for a stronger alternative to be used—since it did not, the speaker must mean something else. Under probabilistic accounts, the greater the expectation for the stronger alternative to be used, the stronger the implicature. Conversely, the lower the expectation for the stronger alternative to be used, the weaker the implicature. Introducing number terms does exactly this—by lowering the overall expectation for *some* use, the relative difference in expectation violation upon observing an underinformative utterance is reduced, leading to a lower rate of pragmatic enrichment. This highlights the importance that alternatives play in scalar implicature computation.

3.2.3.1. Eye movement behavior for semantic versus pragmatic responders: We next explored whether there were differences in looking behavior for pragmatic versus semantic responders both in the quantifier window and in the 200 ms window immediately following the quantifier window. In Experiment 1 we found that semantic responders were slower to integrate the information from the quantifier than pragmatic responders. Experiment 2 allows us to further explore whether there are differences in participants' sensitivity to the availability of alternatives.

To investigate specifically whether semantic and pragmatic responders display different sensitivity to the availability of alternatives, we analyzed the data from Experiments 1 and 2 simultaneously and included the interaction between number presence and responder type as a predictor in the model. A significant interaction is expected if semantic and pragmatic responders respond differently to the introduction of number terms.

The logistic regression model predicted target looks from fixed effects of quantifier (*all* vs. *some*), target set size (*big* vs. *small*), and number presence (*absent* vs. *present*) separately for both the quantifier window and the 200 ms time window immediately following it. In addition, all of the predictors were allowed to interact with responder type (*pragmatic* vs. *semantic*).

Unfortunately, there were very few semantic responders (14, of which only four were in the numbers *absent* condition), compared to 55 pragmatic responders. Seven inconsistent responders were excluded from the analysis. In the quantifier window, there was a robust main effect of responder type, such that semantic responders looked less to the target than pragmatic responders ($\beta = -0.37$, $SE = 0.16$, $p < .05$), replicating the result found for the Experiment 1 data only. With this small data set, the model was unable to detect any finer differences between semantic and pragmatic responders' behavior with different quantifiers or set sizes. The target set size effect remained robust, such that there were more looks to the target for the big set ($\beta = -0.70$, $SE = 0.08$, $p < .0001$).

In the 200 ms window directly following the quantifier window, there was both a main effect of number presence ($\beta = -0.39$, $SE = 0.13$, $p < .01$) and an interaction between number presence and responder type ($\beta = 0.70$, $SE = 0.33$, $p < .05$). Inspecting the simple effects model yielded an effect of number presence for pragmatic responders in the predicted direction ($\beta = -0.54$, $SE = 0.14$, $p < .001$) but not for semantic responders ($\beta = 0.16$, $SE = 0.30$, $p < .60$). The difference in number presence effect for semantic versus pragmatic responders is shown in Fig. 8.

Pragmatic responders initially made faster use of the quantifier as a cue to the target than semantic responders. However, in the 200 ms window directly following the quantifier window, pragmatic responders' increase in looks to the target is delayed when numbers are present compared to when they are absent, and delayed relative to semantic responders' increase in looks to the target. This suggests that semantic responders are employing a strategy of waiting for the disambiguating information from the adjective, but then use that information immediately without residual uncertainty as to the target. Pragmatic responders on the other hand initially use the information from the quantifier but are eventually delayed by the availability of alternatives.

3.2.4. Response time analysis

Mean response times in the underinformative *garden-path some* condition from Experiment 2 are shown alongside those from Experiment 1 in Fig. 9. Visual inspection of the graph suggests that (a) introducing number terms led to an overall slowdown in

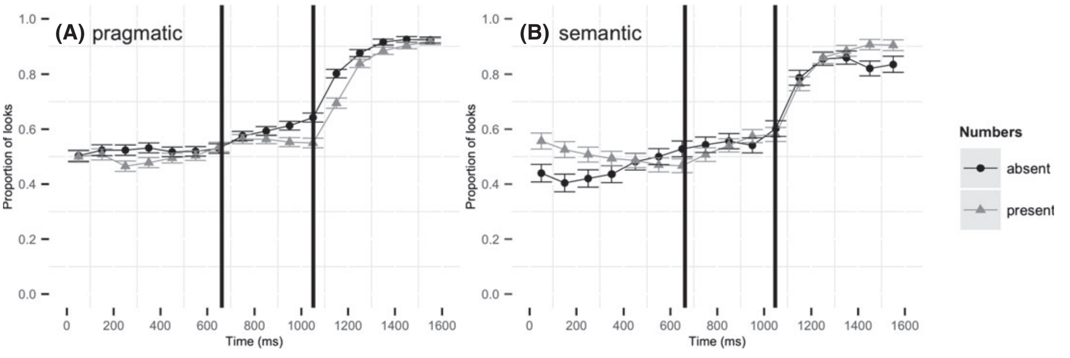

Fig. 8. Proportion of looks to the target for pragmatic (A) versus semantic (B) responders, collapsing over quantifier and target set size. Black lines mark the quantifier window. *X*-axis shows time in ms relative to auditory stimulus onset.

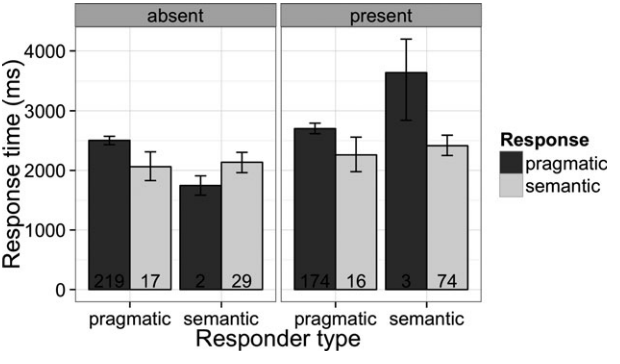

Fig. 9. Mean response times in the *garden-path some* condition for pragmatic versus semantic responses when given by pragmatic versus semantic responders when numbers were absent (Experiment 1, left panel) and present (Experiment 2, right panel). Error bars indicate 95% bootstrapped confidence intervals. Numbers in bars indicate number of contributing observations.

responding to underinformative utterances with *some*, and (b) pragmatic responses are even more clearly slower than semantic responses in Experiment 2, compared to Experiment 1.

To evaluate these questions, we performed a mixed-effects linear regression on the *garden-path some* subset of the data from both experiments. The model predicted log-transformed response time from centered fixed effects of number presence (*absent* vs. *present*), response (*pragmatic* vs. *semantic*), responder type (*pragmatic* vs. *semantic*), and their interactions. The model included by-participant intercepts.

The only effects that reached significance were the main effects of response type and number presence as well as the interaction between responder type and number presence, such that pragmatic responses were slower than semantic responses ($\beta = -0.18$, $SE = 0.04$, $t = -4.7$, $p < .0001$), responses were slower when numbers were present than when they were absent ($\beta = 0.17$, $SE = 0.05$, $t = 3.3$, $p < .001$), and introducing number led to a greater slowdown for semantic than for pragmatic responders ($\beta = 0.34$, $SE = 0.19$, $t = 1.85$, $p < .05$).

These results suggest the availability of number alternatives increases the difficulty of interpreting *some*, even in a coarse-grained measure like click response times, and even when *some* is interpreted literally and no inference is necessary.

4. General discussion

We explored an alternatives-based explanation for the time-course differences in the processing of scalar implicatures reported in recent studies by Huang and Snedeker (2009) and Grodner et al. (2010). In similar visual world experiments, Huang and Snedeker found that upper-bound *some* was delayed, whereas Grodner et al. found that it was computed without delay. We hypothesized that the availability of lexical alternatives to *some,* that do not lie on the <*all, some*> scale, in particular number terms, interfered with processing in Huang and Snedeker, but not Grodner et al. To test this hypothesis, we used a version of the gumball paradigm (Degen & Tanenhaus, 2014) that we modified for use with visual world eye-tracking. The paradigm allowed us to manipulate whether or not numbers were available as alternatives and also to equate the set size for *some* and *all*., using each quantifier to refer to the bigger and smaller of two set sizes, in instructions, such as *You got some/all of the blue gumballs.*

In the absence of instructions with number (Experiment 1), participants generated a scalar implicature from *some* to *not all* as the quantifier *some* unfolded, replicating the results of Grodner et al. (2010). When instructions with number terms were introduced (Experiment 2), there was a general advantage for *big* set *all* over *small* set *some* immediately after the quantifier, replicating the results of Huang and Snedeker (2009). Exploratory analyses of responder types suggested that contextual support for the implicature was lower in Experiment 2 than in Experiment 1 (as evidenced by a lower ratio of pragmatic to semantic responders) and that semantic responders' eye movement behavior differed in systematic ways from that of pragmatic responders. Semantic responders tended not to use the information from the quantifier (regardless of whether the quantifier was *some* or *all*) immediately after observing it, and instead waited for the disambiguating adjective. In contrast, pragmatic responders immediately began using the information from the quantifier after observing it. In contrast to semantic responders, they were additionally delayed in looking to the target after the quantifiers *all* and *some* when number terms were present, even after observing the disambiguating adjective.

These results suggest that the computation of scalar implicatures is affected by the availability of lexical alternatives from the earliest moments of processing. Importantly, the rapid implicatures observed in Experiment 1 cannot be due to pre-coding the implicature: Each display was used with both the quantifiers *all* and *some*. In addition, the inclusion of *garden-path* trials would have made pre-coding the implicature an inefficient strategy.

We also replicated previous response time findings that semantic responses to *some* are faster than pragmatic responses (e.g., Bott & Noveck, 2004). These response time differences were present even when numbers were absent and when there was clear eye movement evidence that participants generated the implicature without a delay. This strongly suggests that the response time differences stem from verification difficulty rather additional processing time to compute the implicature.

We began this article by noting that delayed processing of scalar implicature under conditions where the implicature is strongly supported by context would be a major counter-example to an emerging view of language comprehension in which inference plays a central role in guiding rapid and efficient communication. This inference-based approach contrasts with a view in which inference is a slow, costly, and strategic process. Fast, automatic, language processing is restricted to recovering aspects of linguistic form and meaning. Pragmatic inference, which would seem to be essential for determining speaker meaning, is achieved either by later, more costly, processing or is by-passed by using, simple heuristics.

We demonstrated that the time course of scalar implicature is affected by the availability of context-specific alternatives that cannot be considered members of context-independent scales. Moreover, factors such as the likelihood that different quantifiers are used to refer to big and small sets in the presence of numbers affects response times. We conclude that the availability of alternatives affects the interpretation of utterances containing quantifiers and the speed with which they are processed. In particular, the availability of number terms as potential alternatives with which to describe small sets of objects decreases the speed with which *some* is processed. Several recent formal models of Quantity implicature have incorporated ideas about probabilistic support for alternatives to model the outcome of the computation (Degen et al., 2013; Frank, Goodman, & Tenenbaum, 2009; Franke, 2014; Goodman & Stuhlmüller, 2013). Extending these models to make quantitative predictions for online processing is likely to be a productive avenue for future research.

Accounts of scalar implicature that do not allow for context to affect the early stages of processing, such as the Default and the Literal-First model, cannot explain the pattern of results we report here. In contrast, Degen and Tanenhaus (2014) proposed a constraint-based framework for processing implicatures that was inspired by both information-theoretic considerations and constraint-based approaches in other areas of language processing. This account provides an explanation for the speed at which communication proceeds that makes contact with similar accounts in phonetics, syntax, and lexical processing, and it has the potential for unifying conflicting time-course results.

Under the constraint-based account, the probability of a scalar implicature and the speed with which it is derived depend on the probabilistic support for the implicature in context: the greater the contextual support, the more likely the implicature, and the faster it should be derived. Under this account, the research program on processing of implicatures shifts away from trying to determine whether implicatures are fast versus slow in situations that are as decontextualized as possible, and shifts toward (a) determining the contextual factors that affect the speed and probability of computing an implicature and (b) spelling out how these contextual factors are expected to interact under the assumption that language processing is efficient (in the sense of Levy & Jaeger, 2007; Levy, 2008; Piantadosi et al., 2012).

The experiments reported here and in Degen and Tanenhaus (2014) focused narrowly on the scalar implicature associated with upper-bound *some* in a tightly controlled situation in order to address a debate about time course that has featured prominently in the literature. However, the results have more general theoretical and methodological implications. In particular, the studies we have presented demonstrate that context affects the earliest moments of processing for even the paradigmatic example of a generalized conversational implicature, the upper-bound interpretation of *some*. Two aspects of our data highlight this context dependence. First, even relatively minor changes in the task and the context can affect the probability that addressees will assign *some* an upper-bound interpretation. Second, the presence of alternatives had clear effects on the processing of the implicature. Since the putative context independence of generalized conversational implicatures is what distinguishes them from particularized conversational implicatures, these results suggest that there may not be a categorical difference between how the two are processed. Rather, each involves context-dependent pragmatic inference.

Most generally, as the relevant contextual constraints and the distribution of upper-bound *some* and other implicatures are better understood, we expect that pragmatic inference, like ambiguity resolution, will turn out to be consistent with approaches to language processing that are grounded in constraint-based and information-theoretic principles. Unconstrained inference might be slow and costly. But inference that is constrained by rich conversational context and natural use of linguistic forms might be remarkably easy. If so, then a unified account of the speed and efficiency of language processing might indeed be possible.

Acknowledgments

This work has benefited from conversations with Dan Grodner, Florian Jaeger, Richard Breheny, Jesse Snedeker, Yi Ting Huang, and members of the Tanenhaus lab. We thank Dana Subik for testing participants and Patricia Reeder for the audio recordings. This work was partially supported by NIH grant HD 27206.

Notes

1. Mean response time was calculated relative to these three conditions because especially the garden-path trials are expected to take longer, given that they are initially misleading. This was indeed the case. Mean response times in the *early*, *late*, and *garden-path* conditions were 1,894, 1,856, and 2,489 ms, respectively. Both here and in Experiment 2, the distribution of excluded trials over quantifier and *early/late/garden-path* conditions was relatively uniform, suggesting that the exclusion criterion did not introduce bias.
2. Note that in this graph as well as in all following graphs, data from the *garden-path* condition are included only up until 200 ms after adjective onset because

looking behavior after adjective onset diverges substantially from that on *early* trials (i.e., participants look to the set that is compatible with the adjective before typically looking to the central button on the machine that is clicked to indicate disagreement with the description). Data from *garden-path* trials are collapsed into the *early* category.

3. In order to maximize the power of our analyses, we collapsed over the e*arly* and *garden-path* trials up until adjective onset plus 200 ms because they are equivalent in this region and thus cannot be distinguished by participants. We use 200 ms because it is a conservative estimate of the lower bound for linguistically mediated saccades in task-based visual world experiments (Salverda, Kleinschmidt, & Tanenhaus, 2014). Before doing so, however, we confirmed that these trials indeed did not differ from one another. To test whether looking patterns differed between *early* and *garden-path* trials, we fit a model to the quantifier window predicting target looks from condition (*early* vs. *garden-path*), a continuous time predictor and their interaction. Only the main effect of time reached significance ($\beta = 0.07$, $SE = 0.03$, $p < .01$); neither the main effect of condition ($\beta = -0.01$, $SE = 0.10$, $p < .92$) nor the interaction ($\beta = 0.03$, $SE = 0.05$, $p < .54$) reached significance. This suggests that, while participants did use the information from the quantifier to identify the target (as evidenced in the main effect of time), there was indeed no difference in looking behavior between the *early* and *garden-path* condition. This remained true when controlling for target set size and quantifier. Therefore, combining the *early* and *garden-path* conditions for analyses on the quantifier window is justified. All analyses on the quantifier window here and below were thus performed on both *early* and *garden-path* data.
4. Note that the proportion of pragmatic responses is higher in this study than in many studies reported in the literature. For example, Bott and Noveck (2004, Experiment 3) report 60% pragmatic responses, and Degen and Tanenhaus (2014, Experiment 3), who also used a gumball paradigm, report 29% pragmatic responses. That is, the paradigm employed here seems to provide more probabilistic support for the implicature than these previous studies. One explanation for the difference is that the task in Experiment 1 is implicitly a referential task in which drawing the implicature is more relevant than in the truth-value judgment task of Degen and Tanenhaus (2014). In order to determine the target as quickly as possible, participants should make optimal use of the information provided by the quantifier. By interpreting *some* as *some but not all*, an observation of *some* is more informative than if it is interpreted as *some and possibly all*, since the former is compatible with only one set of gumballs, while the latter is compatible with both. This highlights the importance of the task relevance in scalar implicature computation.
5. As in Experiment 1, to make sure the *early* and *garden-path* conditions could not be distinguished in the quantifier window, we fit a model to the quantifier window predicting target looks from condition (*early* vs. *garden-path*), time, and their interaction. As in Experiment 1, only the main effect of time reached significance ($\beta = 0.18$, $SE = 0.02$, $p < .0001$); neither the main effect of condition ($\beta = -0.03$,

SE = 0.1, $p < .71$) nor the interaction ($\beta = 0.03$, *SE* = 0.05, $p < .54$) reached significance. This remained true when controlling for target set size and quantifier. This again provides justification for collapsing over *early* and *garden-path* trials when performing analyses on the quantifier window.

References

Altmann, G. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, *2*(4), 146–152. Available at http://www.ncbi.nlm.nih.gov/pubmed/21227111.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. Available at http://www.ncbi.nlm.nih.gov/pubmed/10585516.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. doi:10.1016/j.cognition.2008.07.005.

Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 120–125). Austin, TX: Cognitive Science Society.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. doi:10.1016/j.jml.2011.09.005.

Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. doi:10.1016/j.jml.2004.05.006.

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*(3), 434–463. doi:10.1016/j.cognition.2005.07.003.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. doi:10.1016/j.jml.2009.04.003.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. doi:10.3758/PBR.16.5.893.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. doi:10.1016/j.cognition.2007.11.005.

Chambers, C. G., Tanenhaus, M.K, Eberhard, K.M., Filip, H. & Carlson, G.N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49. doi:10.1006/jmla.2001.2832.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology*, *54*(2), 128–133. doi:10.1027/1618-3169.54.2.128.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 376–381). Austin, TX: Cognitive Science Society.

Degen, J., & Tanenhaus, M. K. (2014). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, doi:10.1111/cogs.12171.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. doi:10.1016/0749-596X(86)90006-9.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585. doi:10.1111/j.1467-9280.2009.02335.x.

Franke, M. (2014). Typical use of quantifiers: A probabilistic speaker model. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 487–492). Austin, TX: Cognitive Science Society.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). London: Erlbaum.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, *2*(7), 262–268. doi:10.1016/S1364-6613(98)01187-5.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184. doi:10.1111/tops.12007.

Grodner, D. J., 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*(1), 42–55. doi:10.1016/j.cognition.2010.03.014.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115. doi:10.1016/j.cogsci.2003.10.002.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. doi:10.1016/S0749-596X(03)00022-6.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. doi:10.1016/j.cognition.2008.04.008.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, DC: Georgetown University Press.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. doi:10.1016/j.cogpsych.2008.09.001.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. doi:10.1016/j.cogpsych.2010.02.002.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137–194. doi:10.1207/s15516709cog2002_1.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. doi:10.1111/1467-9280.00211

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, *56*(3), 436–455. doi:10.1016/j.jml.2006.05.002.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19, pp. 849–856). Cambridge, MA: MIT Press.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312. doi:10.1006/jmla.1997.2543.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. doi:10.1111/j.0956-7976.2002.00460.x.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254. doi:10.1037//0096-3445.106.3.226.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. doi:10.1016/j.cognition.2011.10.004.

Posner, M. I., & Snyder, C. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (pp. 669–682). New York: Academic Press.

Qian, T., & Jaeger, T. F. (2012). Cue effectiveness in communicatively efficient discourse production. *Cognitive Science*, *36*(7), 1312–1336. doi:10.1111/j.1551-6709.2012.01256.x.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374.

Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(5), 1207–1223. doi:10.1037/a0012796.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulatory information in spoken-word recognition. *Journal of Memory and Language*, *71*, 145–163.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. *Psychological Review*, *84*, 127–190.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. Available at http://www.ncbi.nlm.nih.gov/pubmed/12480476.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Elmas (Eds.), *Speech, language, and communication. Handbook of perception and cognition* (pp. 217–262). San Diego, CA: Academic Press. Available at http://linkinghub.elsevier.com/retrieve/pii/S0033295X0762012X.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 898–911.

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.

Van Tiel, B. (2013). Embedded scalars and typicality. *Journal of Semantics*, *31*(2), 147–177. doi:10.1093/jos/fft002.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600–606. doi:10.1111/j.1467-9280.2007.01946.x.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.