The speed of inference: Evidence against rapid use of context in calculation of scalar implicatures

Joshua K. Hartshorne
Jesse Snedeker
Harvard University

Running title: The speed of inference

Key words: pragmatics; scalar implicature, entailment, self-paced reading

Word count: 8,551 + 191 (abstract) + tables and references

Send Correspondence to:

Joshua Hartshorne
33 Kirkland Street
WJH 1180
Cambridge, MA 02138
Email: jharts@wjh.harvard.edu
Tel: 617-496-4486
Fax: 617-495-3728

Acknowledgements:

The authors wish to thank Gennaro Chierchia and Yi Ting Huang for comments and discussion, as well as Athena Xie and Stephanie Liem for assistance in data collection. Many thanks also to the audience at the 2009 AMLaP conference in Barcelona, Spain. We are grateful to Richard Breheney and Napoleon Katsos for piquing our curiosity and sharing their stimuli. This work was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program and by the Gordon W. Allport Memorial Fund grant to JKH and by a National Science Foundation grant to JS (0921012).

Abstract

Are pragmatic inferences necessarily the result of complex, contextually-dependent computations? Or are some computed in a heuristic, context-free manner? If context plays any role, when and how is contextual information integrated? A central test case for exploring pragmatic processing addressing these questions has been scalar implicature (e.g., the inference that *some* means *some but not all*). Prior studies of the effects of context on online implicature processing have concluded that context guides processing from the earliest moments with implicatures being computed only when contextually-supported (Breheny, Katsos & Williams, 2006; see also Panizza, Chierchia & Clifton, 2009). These experiments, however, lack critical controls for the lexical and syntactic differences between the different context conditions raising the possibility that the data patterns reflect processes other than implicature. In two self-paced reading studies employing a tightly controlled contextual manipulation, we find that while context does affect scalar implicature generation (Experiment 1), this effect only appears if there is a delay of more than roughly 900 ms between the scalar implicature trigger and the probe word (Experiment 2). Thus we conclude that while context affects pragmatic inference, it does so less rapidly than previously thought.

On most theories, understanding language involves two broad but distinct kinds of processes: decoding processes that allow us to convert phonological representations into semantic representations and pragmatic processes that allow us to draw additional inferences about the interpretation of an utterance in a particular context (Morris, 1938/1971; Grice, 1989; see Bach, 1999 for review). In many cases, the distinction between these two types of knowledge seems uncontroversial. For example, given sentence (1), the fact that Gabe is the agent of the drinking event would typically be attributed to semantic decoding, while the inference that he is an inconsiderate lout who has annoyed the speaker would generally be construed as pragmatic.

(1) Gabe drank all of the milk and put the carton back in the fridge.

There is, however, considerable controversy about where semantics ends and pragmatics begins, and further uncertainty about how pragmatic inferences are calculated. Pragmatics has been defined largely by what it is not (semantics) and thus pragmatic inferences are a diverse class of phenomena that may stem from equally diverse mechanisms. For this reason, research on pragmatic processing necessarily focuses on specific subclasses of inferences such as presuppositions (Chemla, 2009; Gazdar, 1977; Stalnaker, 1973), indirect requests (Brown & Levinson, 1987; Clark, 1979; Holtgraves & Yang, 1990; Pinker, Nowak & Lee, 2008), and, most notably, scalar implicatures (Breheny, Katsos & Williams, 2006; Bott & Noveck, 2004; Hirschberg, 1985; Horn, 1972; Huang & Snedeker, 2009; Noveck, 2001; Noveck & Posada, 2003; Noveck, Chierchia, Chevaux, Guelminger, & Sylvestre, 2002; Noveck & Reboul, 2008). Scalar implicatures provide a convenient entry point for the psycholinguistic investigation of pragmatic inferences because they are ubiquitous, reasonably well-defined and central to theories of pragmatics and its interface with semantics.

Example 2 illustrates a case of a scalar implicature. When we hear a sentence like (2a), we typically assume that (2b) is true as well.

(2) a. John ate some of the cookies.

b. John did not eat *all* of the cookies.

Although this inference is robust, theorists have long argued that it is not part of the semantic (or truth conditional) content of the utterance (Hirschberg, 1985; Horn, 1972). Critically, the inference that *some* implies *not all* is defeasible and thus can be cancelled (3a) unlike a semantic entailment (3b) (Hirschberg, 1985; Horn, 1989; Horn, 1972).

(3) a. John ate some of the cookies. In fact, he ate all of them.

b. *John ate some of the cookies. In fact, he ate none of them.

While theories of scalar implicature vary on many dimensions, all accounts are rooted in Grice's observation (quoted in Strawson, 1952) that these inferences are attributable to the communicative alternatives available to the speaker. In interpreting a sentence like (2a) we draw, implicitly or explicitly, on our knowledge of other contextually-relevant things that the speaker might have said. If John had eaten all the cookies, then (4) would have been a more informative statement, thus the fact that the speaker said (2a) instead suggests that (4) is not true.

(4) John ate all of the cookies

Thus scalar implicatures arise when there are two expressions which form a scale with respect to their strength (*or/and, few/none, warm/hot, OK/excellent*). In all of these cases, the use of the weaker, more inclusive, term (e.g., *or*) can imply that the stronger term (*e.g., and*) does not hold (Horn, 1972). While implicatures are not part of the meaning of the utterance, they can be an essential part of the communicative act. For example, if after hearing (2a) we learned that John had actually eaten all the cookies, and the speaker was aware of this, we would probably feel that we had been misled.

Scalar Implicature, Contextual Dependence & the Speed of Processing

A central challenge for theories of pragmatic inference is specifying the kinds of information that are used to make the inference and the algorithm by which it is made. In the case of scalar implicature, theorists have focused on the role of scalar term itself and the context in which it appears.

For example, in Presumptive Meanings (2000), Levinson posits that comprehenders automatize the process of calculating scalar implicatures and apply it immediately when they encounter a scalar trigger, prior to any combinatorial processing and without consideration of any contextual factors. This theory is rooted in the intuition that the scalar implicature for terms like *some* is a robust part of the meaning of the utterance which is difficult to override. This default implicature can be explicitly cancelled by subsequent information in the discourse (e.g., 3a), but it cannot be suppressed.

In contrast, most other theories posit that scalar implicatures are only computed when contextually supported. For example, on the Gricean analysis (Grice, 1989; Hirschberg, 1985; Horn, 1989), a listener hearing (2a) derives the scalar implicature by reasoning that: a) if the speaker knew that John had eaten *all* the cookies, then it would have been more informative for her to say (3); b) this additional information would be relevant; c) the speaker is sufficiently knowledgeable about the incident to determine whether all of the cookies were eaten; and, d) the speaker is a cooperative speaker and thus she would produce the more informative utterance if she was in a position to do so. If any of these assumptions do not hold, then the scalar implicature is not warranted and should not be calculated. Relevance Theorists also posit that scalar implicature is context dependant (Noveck & Sperber, 2007). On this proposal, the construction of meaning is a continuous process with no clear division between semantic interpretation and pragmatic inference. Context affects which inferences are most readily accessible and which inferences would result in a relevant enrichment of meaning.

The Grammatical Theory of implicature (Chierchia, Fox and Spector, *in press*) provides a different construal of context. This account posits that there is a covert operator in the grammar which has similar properties to the word “only”. “Only” is a focus particle which takes two arguments: one that is its scope (e.g., *Jan* in 5a) and one that is its restrictor (*had braces* in 5a). The interpretation of “only” involves a process called exhaustification in which a contrast set for the scope element is constructed (perhaps in 5 the Brady siblings). A statement with “only” is true if the restrictor is true for the scope element (Jan had braces) but false for all other members of the contrast set (5b).

(5) a. Only Jan had braces.

b. Marsha, Cindy, Greg, Peter and Bobby did not have braces.

In the case of a scalar implicature, the silent exhaustification operator takes the scalar term as it scope turning 6a into 6b and generating the inference in 6c.

(6) a. I ate some of the cookies.

b. I ate only some of the cookies.

c. I did not eat all, most or none of the cookies.

As this is a grammatical process, which can operate at the pre-propositional level, it is presumably initially insensitive to extra-linguistic considerations of the sort explored by Griceans or Relevance Theorists. However, under this theory, the insertion of the covert *only* does depends on the *semantic* context in which the scalar trigger appears. Specifically, exhaustification is more likely in upward-entailing contexts, where it would result in a stronger statement, and less likely in downward-entailing contexts where it would result in a weaker statement.[1] Thus on this hypothesis implicature is sensitive to one form of context .

The present paper focuses on two questions that are central to a theory of scalar implicature (and pragmatic processing more broadly): (a) are scalar implicatures initially calculated by default in a context-independent matter or are these inferences only calculated when the context licenses them, and (b) how do the effects of context on implicature emerge over time? At first these may appear to be the same question, but they are not. Default inferences could be quickly overturned, resulting in a rapid sensitivity to context. Or scalar implicature processing could be gated by contextual factors from the beginning, but the process of integrating this information could be slow resulting in a global delay in scalar inferences. While our first question has been a focus of linguistic theorizing, the second question is rarely addressed. Because such theories are primarily account of linguistic representations rather than account of language processing, they make no explicit predictions about *when* contextual information becomes available for implicature computation. For instance, the Grammatical Theory (Chierchia et al., *in press*) specifies that downward entailing contexts block scalar implicature. However, the theory does not address how and when these contexts are identified during comprehension.

In fact, researchers working with similar linguistic theories often arrive at very different predictions about time course. For example, many experimentalists have adopted the central

[1] Downward-entailing contexts are those in which normal entailment scales are reversed. *John has a black dog* entails that *John has a dog*. In contrast, *If John has a black dog, then X* does not entail *If John has a dog, then X*. Rather, entailment runs the opposite direction: *If John has a dog, then X* entails *If John has a black dog, then X*. Reversals of the entailment scale also arise in negations, imperatives, and other contexts. Thus, the relationship between scalar implicature and entailment context is principled: it leads to the negation of logically stronger alternatives, and what is logically stronger depends on entailment context. Indeed judgments studies suggest that scalar implicatures are less frequently computed in downward-entailing contexts (Chierchia, 2006; Chierchia, Crain, Guasti, Gualmini & Luisa, 2001; Chierchia et al., *in press*; Geis & Zwicky, 1971; Noveck et al., 2002).

tenets of Relevance Theory: that *meaning* emerges gradually with no principled distinction between semantic and pragmatic processes and that inferences are made only when they are relevant in the context. From this starting point Noveck and Sperber conclude that the enriched meanings of scalar terms (*some but not all*) should be calculated slowly relative to unenriched meanings (*some and possibly all*) because enrichment depends on fine-grained contextual processes. In contrast, Grodner, Klein, Carbary & Tanenhaus (2010) predict that enriched meanings should be available as rapidly as unenriched meanings, because there is no purely semantic level of interpretation to traverse. Nevertheless, it is critical that these issues be resolved. A complete theory of pragmatic inference must specify the information and algorithms that are used. Understanding the time course of different processes will place critical constraints on our theory of these mechanisms.

Scalar Implicature and Contextual Dependence: Evidence

There are several empirical studies which suggest that the ultimate interpretation of a scalar sentence depends on context (Bonnefon, Feeney & Villejoubert, 2009; Chierchia et al., 2001; Noveck et al., 2002), but only one study which addresses the time course of context sensitivity: Breheny, Katsos and Williams (2006; henceforth BKW).[2] In three self-paced reading experiments, BKW manipulated the context in which a scalar trigger appeared so that it either supported the implicature or did not. The results suggested that implicatures were only calculated in supportive contexts: participants read continuations which presupposed the implicature more rapidly in the supportive contexts than in unsupportive ones. Interestingly, BKW also found slower reading times on segments containing the scalar implicature trigger (e.g., *some*) in the

[2] Panizza et al. (2009) found early contextual dependence for numeral comprehension, but the relationship between numerals and scalar implicature is complex and hotly contested (Breheny, 2008; Horn, 1989; Panizzza, Huang, Hartshorne, Snedeker & Chierchia, *in prep*).

implicature-supporting contexts, indicating that not only is the outcome of scalar implicature processing context-sensitive, but that the earliest stages of processing are context sensitive.

These results are widely regarded as evidence that scalar implicatures are immediate, context-sensitive, and immediately context sensitive (Benz & van Jooij, 2007; Bonnefon et al., 2009; Chevallier, Noveck, Nazir, Bott, Lanzetti & Sperber, 2008; Chevallier, Wilson, Happe & Noveck, 2010; Frazier, Clifton & Stolterfoht, 2008; Grodner et al., 2010; Guerts & Pouscoulous, 2009; Katsos, 2008; Nieuwland, Ditman & Kuperberg, 2010; Noveck & Reboul, 2008; Noveck & Sperber, 2007; Panizza et al., 2009; Winterstein, 2009). As such, the paper has quickly become one of the most-cited empirical papers on scalar implicature.

Problems with Interpreting Breheny, Katsos & Williams (2006)

Unfortunately, all three of the experiments in BKW contain confounds that make it impossible to link the critical effects to scalar implicature per se. Consequently, we know less about scalar implicature than we believe we do. We consider each experiment in turn.

Experiment 1 contrasted sentences in which, by hypothesis, scalar implicatures were supported (7a) or nonsupported (7b). The sentences were presented in segments, marked here by slashes. The original stimuli were in Greek and thus the examples here are translations.

(7) a. SI-supporting: The day's offer usually is:/You can have a full menu/for one person/and the second person/can have the plat de jour/for free./Today,/customers could have for free/meat or fish./

b. SI-nonsupporting: The dietician that visited the school/explained to children/how useful for our body/protein can be./He also told them/that we can find protein in/meat or fish./

Participants read the critical segment (*meat or fish*) more slowly in the scalar implicature-supporting context (7a), which the authors took as evidence that scalar implicature processing is modulated by context. However, the two conditions also differed substantially in their lexical and propositional content, raising the possibility that context was confounded with factors such that are known to affect reading speed, such as cloze probability, syntactic complexity, lexical repetition and semantic priming (*inter alia*: Gordon, Grosz, & Gilliom, 1993; Frisson, Rayner & Pickering, 2005; McNamara & Healy, 1988; Trueswell, Tanenhaus & Kello, 1993; Zola, 1984). In the absence of controls for these factors, there is no way of knowing whether the reading time difference was due to scalar implicature processing, particularly since the authors offer no evidence that participants in fact calculated scalar implicatures in (7a) and not (7b).

Experiment 2 develops a method of detecting online whether a scalar implicature has been calculated. The authors predicted that scalar implicatures are more likely in focused positions (8a) than non-focused positions (8b):

(8) a. SI-supporting: Some of the consultants/had a meeting/with the director./The rest/did not manage/to attend.

b. SI-nonsupporting: The director/had a meeting/with some of the consultants./The rest/did not manage/to attend.

Note that in these sentences *the rest* makes reference to the compliment set (the rest of the consultants) and is thus only felicitous if the director did not meet with all of the consultants – that is, if the scalar implicature has been made. Indeed, BKW found slower reading time for *the rest* in the non-focused condition (8b) relative to the focused condition (8a). Importantly, this delay was not present in a third condition in which *some* in 8b was replaced with *only some.* As we noted earlier “only some” has as a part of its meaning, *some-but-not-all,* thus forcing the partition of the set.

While this control is clever and informative, it does not eliminate a critical confound between 8a and 8b. In 8a, where the implicature is calculated, the scalar trigger occurs three regions before the critical probe ("the rest"). In 8b, it occurs immediately before the trigger. Thus it is unclear whether the lack of a scalar implicature in 6b reflects context-sensitivity or the inability to make an SI in the time allotted. BKW were aware of this concern and included a 1000 ms pause before *the rest* in all of the conditions. However, if we take the studies which show that scalar implicature calculation is slow at face value (Bott & Noveck, 2004; Feeney, Scrafton, Duckworth, & Handley, 2004; Huang & Snedeker, 2009a, *in press*; Noveck & Posada, 2003), then it is not clear that the additional 1000 ms would provide sufficient time for completing this process.

Experiment 3 combined the best features of Experiments 1 and 2, again comparing a scalar implicature-supporting context (9a) with a scalar implicature-nonsupporting condition (9b):

(9) a. SI-supporting: Mary/asked John/whether he intended to host/all of his relatives/in his tiny apartment./John replied/that he intended to host/some of his relatives./The rest/would stay/in a nearby hotel.

b. SI-nonsupporting: Mary was surprised/to see John/cleaning his apartment/and she asked/the reason why./John told her/that he intended to host/some of his relatives./The rest/would stay/in a nearby hotel.

As in Experiment 1, participants read the trigger segment (*some of his relatives*) more slowly in the scalar implicature supporting context (9a) relative to the nonsupporting context(9b), which the authors again interpreted as evidence that scalar implicature computation is gated by context. Second, participants read *the rest* more slowly in (9b) relative to (9a), as would be expected if scalar implicatures were calculated more often for the latter.

However, as in Experiment 1, these conclusions depend on the assumption that it was scalar implicature calculation that produced the differences in reading times, not other confounds present in the broad contextual manipulation. Unfortunately, there is good reason to question this assumption. In Experiment 3's scalar implicature-supporting context (9a), the critical quantified phrase (*some of his relatives*) was always preceded by a reference to *all* of the objects in question (*all of his relatives*), whereas in the scalar implicature-nonsupporting context (9b) the entities in question were never mentioned earlier in the paragraph. Thus, the slower reading times in the scalar implicature-supporting context could reflect the infelicity of using a full noun phrase to refer to a set that is salient in the discourse and might be better described with a pronoun (e.g., *some of them*). Processing costs of this kind have been observed in prior studies (the repeated name penalty: Gordon & Chan, 1995; Gordon et al., 1993).

As a first-pass test of the plausibility of this hypothesis, we conducted a sentence completion task with 10 of the stimulus pairs used in BKW's Experiment 3. Each story was truncated just prior to the word *some*. 24 participants each completed 5 scalar implicature-supporting or 5 non-scalar implicature-supporting sentences, with a total of 10 sentences of each type tested (due to a typo on one of the forms, one sentence type was excluded from analyses, excluding this item did not change the pattern of results). In the scalar implicature-supporting context, 55% (20/37) of the continuations that referred to the target set (*books, relatives, etc.*) used a pronoun instead of a noun, whereas in the scalar implicature-nonsupporting context 0% of these continuations did (0/18). These results indicate that participants' expectations for referential forms were very different in these two contexts. They support our hypothesis that the quantifier segment was infelicitous in the scalar implicature-supporting context (9a), and thus the longer reading times should be interpreted as a repeated named penalty (Gordon & Chan, 1995; Gordon et al., 1993) rather than as evidence for scalar implicature.

BKW did try to address potential confounds in the context that might affect reading of *the rest*, by including a control version of (9a) in which *some* was replaced with *only some*. In this condition, *the rest* was read just as quickly as in (9a). Unfortunately, they did not include a control version of the scalar implicature nonsupporting context (9b) (as they had in Experiment 2). Thus it is impossible to know whether the delay in (9b) is attributable to the failure to make the implicature as opposed to different expectations about how the discourse will continue given the prior context.

In sum, we find that the three experiments reported by BKW lack the controls necessary to pin their results on scalar implicature processing rather than on some other confounded aspect of the manipulations. Their conclusions may be correct, but the evidence itself is ambiguous. Thus new, unambiguous data are needed.

The Present Study

In the present study, we investigate online contextual sensitivity in scalar implicature processing, borrowing many elements from BKW's design but using a more constrained contextual manipulation. This manipulation derives from linguistic analyses proposing that scalar implicatures are dispreferred in contexts, such as the antecedent of a conditional, in which direction of semantic entailments are reversed (downward entailing contexts, Chierchia, 2006; Chierchia et al., *in press*; Geis & Zwicky, 1971). For example, (10a) does not give rise to the interpretation in (10b).[3]

[3] It is sometimes incorrectly suggested that Presumptive Meanings also predicts no implicatures in downward-entailing contexts such as conditionals. This appears to be due to Levinson's (2000, pp. 254-55) claim that positive lexical scales -- <some, many, most, all> -- have negative

(10) a. If you eat some of your vegetables, you can have dessert.

b. If you eat some *but not all* of your vegetables, you can have dessert.

While there is not yet any published experimental data confirming this pattern for *some*, conditionals and other downward-entailing contexts have been found to block scalar implicatures for *or* in offline judgments (Chierchia et al., 2001; Noveck et al., 2002, see Panizza et al., 2009 and Panizza, Huang, Hartshorne, Snedeker & Chierchia, *in prep*, for related data on numbers). By contrasting declarative and conditional sentences we were able to reduce our contextual manipulation to two words (see (11), below). We controlled for any remaining scalar-implicature-irrelevant differences between declarative and conditional sentences by creating a parallel sentences of *both* types in which *some* was replaced with *only some*.

It should be stressed that the present study does not seek to replicate BKW. Our manipulation of grammatical context is different in kind from the contextual manipulations they used. Rather, the present study seeks to test some of the same questions with a method inspired by BKW but freed of that study's confounds. We also added one additional methodological innovation, not present in BKW: Between Experiments 1 and 2, we varied the number of words between the scalar implicature trigger (*some of her homework*) and *the rest*, thus varying how much time participants had to integrate the contextual information.

Experiment 1

Participants read matched declarative (11a) and conditional (11b) sentences segment-by-

---

counterparts -- <not all, not most, not many = few, not some = none> -- where the ordering of terms indicates that each term implicates the negation of the term on the right. Thus *some* implicates *not all*, but *not some* (*none*) does not implicate *not not all* (*all*). Note that on Presumptive Meanings, scalar implicatures operate on individual words in a context-independent manner. Levinson (2000) posits that *not all* must in some way be treated as an individual word. Leaving aside how this could happen pre-compositionally, it is clear that *if…some* and *if….all* cannot be so treated, as an arbitrary amount of lexical material may separate *if* from the quantifier.

segment in a self-paced reading task. The differences between our scalar implicature-supporting (11a) and scalar implicature nonsupporting (11b) contexts were limited to two words: the addition of *if* in (11b) and the replacement of *and* with *then*. If context has an early effect on scalar implicature processing – and if scalar implicature calculation is computational costly and leads to longer reading times (Bott & Noveck, 2004; Breheny et al., 2006; De Neys & Schaeken, 2007) – then we would expect longer RTs for the quantifier region (*some of her homework*) in (11a) relative to (11b). As in BKW, we can probe the effects of context on the outcome of scalar implicature processing – the interpretation of *some* – by checking RTs at *the rest.* If a scalar implicature has been calculated (*some but not all of her* homework), it should be easier to determine reference for *the rest* (*the rest of her homework*) than if no scalar implicature has been calculated. Thus, we expect longer reading times at *the rest* in (11b) relative to (11a).

While the simplicity of this context manipulation removes many potential confounds, any difference in reading time between (11a) and (11b) could still reflect differences in the semantic or syntactic processing of declarative and conditional sentences, rather than the likelihood of scalar implicature computation per se. Thus we ran a separate control condition comparing declarative and conditional sentences in the absence of implicature. This was done by changing *some* in both sentences to *only some* (12). The presence of *only* semantically forces the reader to interpret *some* as *some but not all* regardless of context, eliminating any role for scalar implicature.

(11) a. SI-supporting: Mary did/some of her homework/this morning/before breakfast,/and the rest/must be done later today.

b. SI-nonsupporting: If Mary did/some of her homework/this morning/before breakfast,/then the rest/must be done later today.

(12) a. Mary did/only some of her homework/this morning/before breakfast,/and the rest/must be done later today.

b. If Mary did/only some of her homework/this morning/before breakfast,/then the rest/must be done later today.

Method

Participants. Twenty-eight native English speakers participated in Experiment 1 (21 female; 18-37 y.o., M=24, SE=1):14 in the experimental condition (9) and 14 in the control condition (10). The experimental/control manipulation was tested between subjects both to shorten the experiment and to avoid excessively priming the subset interpretation of *some* in the experimental condition. One additional participant failed to complete the experiment and was replaced.

Materials. Eighty sentence templates were created, with nine versions of each template (Table 1). These included two experimental (*some*) sentences, as in (9), and two control (*only some*) sentences, as in (10). For each of these four sentences, a filler sentence was created that was identical through the first clause, but which concluded without referring to the set mentioned in the first clause, and thus without *the rest*. This ensured that the occurrence of this phrase could not be predicted solely on the basis of seeing *some*. As results from these filler items were not to be analyzed, effort was made to make the continuations maximally natural rather than uniform (see Table 1). Finally, a declarative filler sentence was created by substituting *all* for *some* in the first clause and adding a natural continuation that did not re-mention the set. Computer error led to sentences from 2 of the 80 templates to be displayed incorrectly. These errors affected only some of the participants, but for consistency items derived from these templates were removed for all subjects and from all experiments. However, leaving these items in does not affect the qualitative pattern of results.

Additional fillers were as follows: 40 from an experiment on direct object sentence complement ambiguities (e.g., *The local veterinarian/advised (that) the horse/should be prevented from running*) and 40 from a pronoun reference experiment (e.g., *John feared Sally because (s)he was from Macedonia*). None of these fillers contained partitive expressions or *the rest*.

Procedure. Each participant was randomly assigned to the experimental condition (*some*) or the control condition (*only some*). Each participant was presented with 20 critical sentences (10 in declarative contexts and 10 in conditional context), 20 matched fillers (10 in declarative contexts and 10 in conditional contexts) and 40 *all* fillers. These items were chosen randomly without replacement from the 80 templates, separately for each participant. Thus, counterbalancing was standard (each item/template appearing only once and the same sentence template being used in every condition), except that there were no item groups (groups of items that rotate through the conditions together) and hence no lists. In addition to those 80 sentences, each participant also viewed the 80 additional filler sentences described above. The order of these 160 sentences was randomized for each participant.

Sentences were presented one segment at a time in the center of the screen. Participants pressed the spacebar to advance to the next segment. Participants answered comprehension questions after 34% of trials – slightly more than the 25% figure used by BKW. Participants were additionally told to expect a further comprehension test after the experiment, which did not take place. Thus any differences between the studies are unlikely to be due to our participants paying *less* attention.

Experiment 1: Stimuli

Experimental (*some*)

Declarative: *Addison ate/some of the cookies/before breakfast/this morning,/and the rest/are on the counter.*

Conditional: *If Addison ate/some of the cookies/before breakfast/this morning,/then the rest/are on the counter.*

Control (*only some*)

Declarative: *Addison ate/only some of the cookies/before breakfast/this morning,/and the rest/are on the counter.*

Conditional: *If Addison ate/only some of the cookies/before breakfast/this morning,/then the rest/are on the counter.*

Filler (*some*)

Declarative: *Addison ate/some of the cookies/before breakfast/this morning,/and then/she ate some cake.*

Conditional: *If Addison ate/some of the cookies/before breakfast/this morning,/then that/is OK.*

Filler (*only some*)

Declarative: *Addison ate/only some of the cookies/before breakfast/this morning,/and that/is OK.*

Conditional: *If Addison ate/only some of the cookies/before breakfast/this morning,/then that/is OK.*

Filler (*all*)

Declarative: *Addison ate/all of the cookies/before breakfast/this morning,/and now/she feels sick.*

Table 1. Example of all nine versions of one template used in Experiment 1.

Results & Discussion

Results (Table 2) were analyzed using mixed-effects models with subjects and sentence templates as random effects in R using the lme4 package, and p-values were estimated using the function pvals.fnc, which implements Markov chain Monte Carlo sampling with 10,000 samples (*see* Baayen, 2008). Accuracy on the comprehension questions was moderately high, given that they were designed to be difficult and thus promote attention (M=88%, SE=1%). RTs shorter than 200ms and longer than 7sec were removed (2% of data). The interaction between context (declarative vs. conditional) and condition (control vs. experimental) was significant only for the final segment (Table 3). As expected, this interaction was driven by significantly shorter RTs in the declarative context relative to the conditional context in the experimental condition. We interpret this as a spill-over effect reflecting differences in the interpretation of the *the rest*. This greater difficulty in interpreting *the rest* is consistent with making a scalar implicature in the declarative contexts but not the conditional contexts: without a prior scalar implicature, there is no accessible referent for *the rest* in participants' discourse model. This interpretation is supported by the absence of any effect of context in the control condition, where the subset reading was semantically forced, indicating that the differences in the experimental condition were due to the interpretation of *some*, not to scalar implicature-irrelevant differences between processing declarative and conditional sentences. Because the scalar implicatures were not explicitly cancelled, this finding is incompatible with Levinson's (2000) Presumptive Meanings account which, strictly speaking, provides no other mechanism by which context can influence scalar implicature.

We did not, however, find any effect of the scalar trigger itself, suggesting again that BKW's effects on the scalar trigger were driven by confounding variables, such as the repeated name penalty. Thus, in this experiment, we find no evidence of a very early modulation of scalar implicature processing by context, an issue that we return to the General Discussion.

Experiment 1: Results

	(If) Mary did	(only) some of her homework	this morning	before breakfast	and/then the rest	must be done later today
			Experimental sentences (some)			
Declarative:	902 (86)	846 (68)	767 (47)	858 (74)	694 (44)	982 (61)
Conditional:	1012 (108)	935 (96)	790 (47)	837 (72)	675 (47)	1244 (114)
Difference:	*-110*	*-89*	*-23*	*21*	*19*	*-262*
Cohen's d:	*0.2*	*0.2*	*0.1*	*0.1*	*0.1*	*0.6*
t-stat/p-value	1.81 (p=.07)	1.31 (p=.19)	0.72 (p=.47)	0.33 (p=.74)	0.37 (p=.71)	3.58 (p=.00*)
			Control sentences (only some)			
Declarative:	772 (72)	868 (100)	642 (62)	736 (89)	568 (49)	873 (121)
Conditional:	790 (77)	874 (87)	660 (54)	723 (77)	601 (53)	920 (120)
Difference:	*-18*	*-6*	*-18*	*13*	*-33*	*-47*
Cohen's d:	*0.1*	*0*	*0.1*	*0*	*0.1*	*0.1*
t-stat/p-value	0.40 (p=.69)	0.04 (p=.97)	0.61 (p=.54)	0.05 (p=.96)	0.96 (p=.34)	0.99 (p=.32)

Table 2. Segment-by-segment RTs (SEs) in milliseconds for Experiment 1, with t-statistics, p-values, and Cohen's d.

Experiment 1: Statistical Comparisons

	(If) Mary did	(only) some of her homework	this morning	before breakfast	and/then the rest	must be done later today
context:	1.96 (p=.05)	1.48 (p=.14)	0.73 (p=.47)	0.40 (p=.69)	0.42 (p=.67)	4.35 (p=.00*)
condition:	1.09 (p=.28)	0.15 (p=.88)	1.62 (p=.11)	1.13 (p=.26)	1.76 (p=.08)	0.43 (p=.67)
Interaction:	1.12 (p=.26)	0.95 (p=.34)	0.20 (p=.84)	0.08 (p=.93)	1.03 (p=.31)	2.28 (p=.02*)

Table 3. Segment-by-segment t-statistics and p-values for effects of context (declarative vs. conditional), condition (experimental vs. control), and for the interaction between context and condition. Reference condition is the declarative experimental sentences.

Experiment 2

Experiment 1 confirmed theoretical assertions that scalar implicatures are not calculated by default in all contexts, contradicting predictions by Levinson (2000). The results moreover support the claim that scalar implicatures are less likely to be computed in the antecedent of conditional sentences (Chierchia et al., *in press*).

However, Experiment 1 provides relatively little information about when and how context impacts scalar implicature calculation. We know only that grammatical context affects scalar implicature calculation within approximately 2.5 seconds of seeing the scalar trigger (this is the average time elapsed between the onset of the trigger *some* segment and the probe *the rest* segment). These results are consistent with the possibility that scalar implicatures are immediately calculated by default in both conditional and declarative sentences and then cancelled in the downward entailing (conditional) context prior to encountering our probe (*the rest*). Thus, our psycholinguistic theory has been only minimally constrained.

In Experiment 2, we address these issues by removing the two segments between the quantifier phrase *some* and *the rest*, thus restricting the time available to approximately 900 ms (the length of the quantifier segment). If this is not enough time to integrate contextual information, there should no longer be any difference in RTs between the declarative and conditional conditions at *the rest*.

If indeed no difference between declarative and conditional sentences is found, we can determine whether this is due to (a) a lack of scalar implicature computation in the declarative sentences, as would be consistent with context-dependent calculation, or (b) a lack of scalar implicature cancellation in the conditional sentences, as would be consistent with default computation, in the following way: (a) predicts that slower reading times subsequent to *the rest* for declarative sentences in Experiment 2 relative to Experiment 1, as the lack of a scalar implicature will make integrating *the rest* more difficult in Experiment 2, whereas (b) predicts faster reading times subsequent to *the rest* for conditional sentences in Experiment 2 relative to Experiment 1, as failure to cancel the scalar implicature will make integrating *the rest* easier in Experiment 2.

Method

Participants. Forty-six native English speakers participated in Experiment 2 (36 female, 9 male, one "NA"; 16-38 y.o., M=23, SE=1), 22 in the experimental (*some*) condition, and 24 in the control (*only some*) condition. A larger number of participants were tested than in Experiment 1 to help ensure that any null result was not due to insufficient power.

Materials and Procedure. The stimuli for Experiment 2 were created from those of Experiment 1 by removing the middle two segments in the experimental and control sentences. Filler items were similar to those in Experiment 1 except that the pronoun sentences involved present-tense verbs instead of past-tense verbs.

Results and Discussion

Accuracy to comprehension questions was acceptable (M=91%, SE=1%). Removing one participant with lower accuracy (70%) does not change the pattern of results. Response times were analyzed in the same way as before (Table 4). In contrast to Experiment 1, the sentence context by condition interaction was not significant at any segment (*p*s>.1; Table 5) despite the fact that there were nearly twice as many subjects in Experiment 2 as in Experiment 1. Thus, the lack of a contextual effect at the trigger segment (*some*) was replicated. However, we no longer found a difference at or subsequent to *the rest*, indicating that there was now no difference in the interpretation of *some* in the declarative sentences relative to the conditional sentences. To compare the two experiments, we entered reading times for the final segment in both experiments into a single analysis. The three-way interaction of experiment by sentence context by condition was significant at the final segment ($t=2.60$, $p<.01$).

Was this difference between experiments due to scalar implicatures being in force in both declarative and conditional sentences in Experiment 2 – as predicted by the default accounts – or due to scalar implicatures not being in force in either sentence type in Experiment 2 – as predicted by the context-dependent accounts? Paired comparisons support the later conclusion: Specifically, the reading times for the final segment in the declarative *some* sentences in Experiment 1 (982 ms) were shorter than those in Experiment 2 (1141 ms) ($t=2.15$, $p=.03$, $d=.34$), suggesting that a scalar implicature had been calculated in Experiment 1 (aiding interpretation of *the rest*) but not in Experiment 2. In contrast, reading times for the conditional sentences did not differ significantly across experiments ($t=0.48$, $p=.63$, $d=.20$). Similarly, in the control *only some* condition, neither the declarative sentences nor conditional sentences were significantly affected by the length manipulation (ts<1). This observation is further bolstered by a marginally significant two-way interaction between *some* vs. *only some* for the declarative sentences ($t=1.70$, $p=.09$), whereas the two-way interaction for conditional sentences was not significant ($t=0.46$, $p=.64$). Thus, the primary effect of the extra time afforded by Experiment 1 was to make *the rest* easier to integrate in the declarative *some* sentences, consistent with the onset of a scalar implicature.

Experiment 2: Results

	(If) Mary did	(only) some of her homework	and/then the rest	must be done later today
		Experimental sentences (some)		
Declarative:	927 (74)	848 (64)	690 (42)	1141 (83)
Conditional:	1011 (66)	962 (83)	696 (38)	1135 (80)
Difference:	*-84*	*-114*	*-6*	*6*
Cohen's d:	*.2*	*.2*	*0*	*0*
t-stat/p-value	*1.32 (p=.19)*	*1.93 (p=.05*)*	*0.13 (p=.90)*	*0.09 (p=.93)*
		Control sentences (only some)		
Declarative:	859 (67)	987 (60)	674 (31)	951 (59)
Conditional:	952 (77)	1009 (58)	687 (34)	1052 (85)
Difference:	*-93*	*-22*	*-13*	*-101*
Difference:	*.2*	*.1*	*.1*	*.2*
t-stat/p-value	*1.85 (.06)*	*0.44 (.66)*	*0.47 (.64)*	*1.88 (.06)*

Table 4. Segment-by-segment RTs (SEs) in milliseconds for Experiment 2, with t-statistics, p-values, and Cohen's d.

Experiment 2: Statistical Comparisons

	(If) Mary did	(only) some of her homework	and/then the rest	must be done later today
context:	1.49 (p=.14)	2.13 (p=.03*)	0.20 (p=.84)	0.21 (p=.83)
condition:	0.58 (p=.56)	1.48 (p=.14)	0.29 (p=.77)	1.72 (p=.09)
Interaction:	0.02 (p=.99)	1.28 (.20)	0.07 (.94)	1.37 (.17)

Table 5. Segment-by-segment t-statistics and p-values for effects of context (declarative vs.

conditional), condition (experimental vs. control), and for the interaction between context and condition. Reference condition is the declarative experimental sentences.

Experiment 3

Experiment 1 confirmed that scalar implicature is affected by (grammatical) context. The comparison of Experiment 1 and 2 indicates that this contextual information first affects the interpretation of *some* somewhere between 1 second and 2.5 seconds after encountering *some*.

Having established that the effect of context on *the outcome* of scalar implicature calculation is observed only relatively late, in this final experiment, we address a related issue: Is scalar implicature *processing* gated by context? BKW reported longer reading times on scalar triggers (*some*, *or*) in scalar implicature-supporting conditions, which they interpret as evidence that the implicature processing begins immediately and is immediately gated by context. However, in the Introduction, we noted that the manipulations used in the BKW experiments were confounded in ways that prevent us from drawing such strong conclusions from this data pattern.

In neither of our experiments did we observe a similar slow-down in the scalar implicature-supporting (declarative) contexts, which could be taken as evidence that context does not gate processing, but only the outcome of processing. In fact, in both Experiments 1 and 2, reading times were non-significantly *longer* in the scalar implicature-nonsupporting (conditional) context –counter to BKW's hypothesis. To examine this more closely, we conducted an additional analysis in which we pooled together the critical items from Experiments 1 and 2 as well as the matched filler items in each experiment which had quantifier phrases but did not conclude with reference to *the rest* (see Table 1), thus doubling the number of observations per participant. Despite this substantial increase in power, the critical 2x2 interaction between context and condition did not emerge (t=1.31, p=.19) and the trend in the means was in the wrong direction, with reading times numerically longer in the conditional sentences -- this effect being numerically larger for the *some* sentences than for the *only some* sentences (dec. some: M=856, SE=48; con. some: M=918, SE=54; dec. only some: M=957, SE=56; con. only some: M=968, SE=57). The results for the subsequent segment were similar (t=0.31, p=.75).

We also conducted an exact replication of Experiment 1 (N=38) and two additional versions of Experiment 2 (Ns=30 & 43). The first replication of Experiment 2 presented words individually and presented both experimental (*some*) and control (*only some*) stimuli within subjects. The second presented content words individually or with accompanying function words (*Addison/ate/some of/the cookies/…*) and presented experimental and control stimuli between subjects. The replication of Experiment 1 replicated the effect at *the rest* -- with longer RTs in the conditional vs. declarative sentences in the experimental condition only -- and the replications of Experiment 2 similarly failed to show significant effects at *the rest*. In none of these experiments did we find an interaction between context (declarative/conditional) and condition (*some*/*only some*) at the quantifier segment or at the segments immediately following (ps<.05). Thus, these results seem very robust.

Nevertheless, we considered one final reason why this effect may have failed to appear in the present studies. In a downward entailing context, the concept expressed by *some* can also be expressed by the negative polarity item *any* (see Chierchia, 2006, for a relevant discussion). The availability of this alternative could potentially make *some* less felicitous in such a context.[4]

---

[4] We thank Napoleon Katsos and one anonymous reviewer for raising this possibility.

This infelicity could lead to longer reading times for *some* in the conditionals, which could cancel out any hypothesized difference in reading times caused by scalar implicature processing in the declarative sentences. In Experiment 3, we tested this possibility by eliciting naturalness judgments for the stimuli used in Experiment 1, manipulating whether the quantifier in the conditional sentences was *some* or the negative polarity item *any*.

Method

Participants. Two groups of thirty English-speaking participants, drawn from the same population, were tested, half in the *some* condition (ages 19-59, M=37, SD=12; 17 female; 28 native English speakers) and half in the *any* (ages 20-64, M=37, SD=13, 1 no report; 19 female, 1 no report; 27 native English speakers, 2 no report). Four additional participants were excluded for leaving more than two items blank.

Materials and Procedure. The critical sentences from Experiment 1 were modified by terminating them before the comma with an ellipsis:

(13) a. Mary did some of her homework…

b. If Mary did some of her homework…

A random ordering of the 80 sentence templates was chosen, and half the templates were put in declarative form (13a) and half in conditional form (13b) such that there were never more than three consecutive sentences of the same type. Four lists were created by counter-balancing condition (declarative/condition) and sentence order (forwards/backwards) across lists. Another set of four lists were created by replacing *some* with *any* in all the conditional sentences (the *any* condition). Participants were asked to rate each item in terms of naturalness on a 9-point Likert scale.

Results and Discussion

Overall, sentences were rated as quite natural (M=7.5, SE=0.2). Results (Figure 1) were analyzed using mixed effects modeling with subjects and sentence templates as random factors. There was a significant main effect sentence context (t=7.04, p<.001), reflecting lower naturalness ratings for conditional sentences across the board, which was also noted spontaneously by participants ("I didn't feel any of these sentences should have begun with the word 'if.'"). However, this was not a function of the polarity of the quantifier, as in fact the difference between sentence contexts was slightly *larger* for the negative polarity item *any*, reflected in a significant interaction between sentence context and quantifier polarity (*some* vs. *any*; b=-0.18, t=2.07, p=.04).

Thus, there is no evidence that *some* was infelicitous in the conditional sentences. These results, taken together with the results of Experiments 1 and 2, suggest that processing effort – as measured by reading time – was roughly equivalent whether or not an implicature was produced, though of course we cannot rule out the possibility that self-paced reading time is simply not sensitive to the relevant processing effort. We interpret these results in the General Discussion.

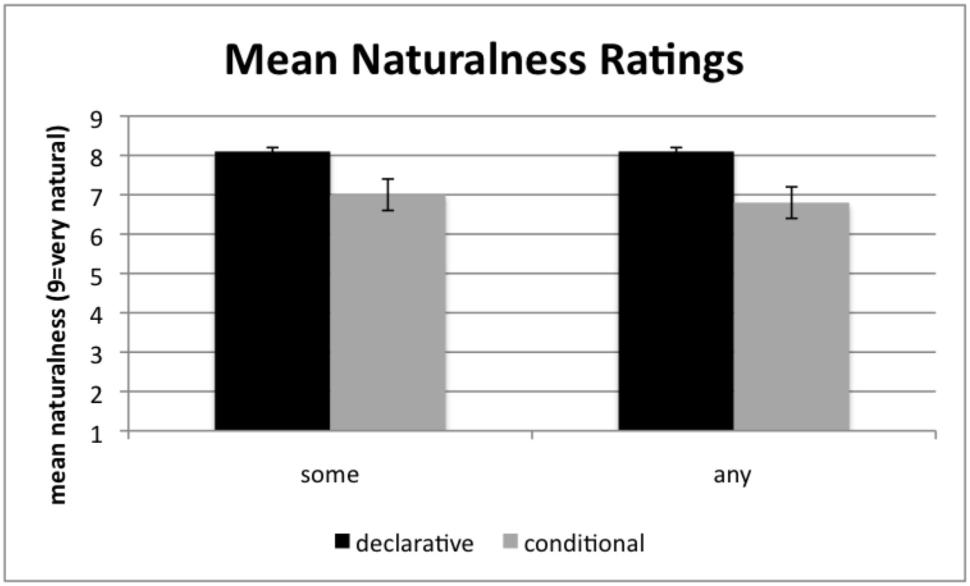

Figure 1. Rated naturalness of the first clause in the sentences used in Experiment 1.

General Discussion

Experiment 1 confirmed that scalar implicature is affected by context: specifically, whether a scalar term in an upward or downward entailing environment. The comparison of Experiment 1 and 2 put a rough bound on when that contextual information affects the interpretation of *some*: somewhere between 1 second and 2.5 seconds after encountering this scalar trigger.

These combined results suggest that scalar implicatures were only calculated in licensed contexts, rather than being calculated by default and then revised. Thus our findings put strong constraints on the space of possible psycholinguistic theories. They are particularly problematic for Levinson's Presumptive Meanings account (2000), on which scalar implicatures are always calculated and may only be explicitly overridden. Even if the theory were revised to allow downward-entailing contexts to implicitly cancel the scalar implicature, these data would be a challenge. Such a theory would have to posit: a) that automatic default processes that produce implicatures are so slow and effortful that they have no effect on interpretation of the rest for about 1000 ms but b) the rich contextual processes that override implicatures are so rapid that they are completed 1500ms later. Thus this data pattern undermines any motivation for a two-stage processing account.

Interestingly, at no time point were reading times significantly longer in scalar implicature-supporting (declarative) contexts, a robust result that cannot be explained by the felicity of *some* in the downward entailing contexts (Experiment 3). This finding suggests that the effects that BKW observed on the scalar trigger were due to confounds in their manipulation of context. Although this is a null result and thus difficult to interpret, this pattern seems to suggest that the processing costs related to the scalar trigger are the same regardless of whether the implicature is calculated or not (see below).

In the remainder of this section, we discuss these findings in terms of both theory and previous experimental results.

The speed of scalar implicature calculation

Experiment 2 suggests that approximately 900 ms after encountering the scalar term *some*, participants still have not finished computing the scalar implicature. These results may help inform an ongoing debate on the speed of scalar implicature computation.

A number of studies report that it takes as much as 1-2 seconds longer to judge sentences like *some elephants have trunks* as false – consistent with the scalar implicature – as it does to judge them true – consistent with literal meaning and no implicature (Bott & Noveck, 2004; Feeney et al., 2004; Noveck & Posada, 2003). While these results are roughly consistent with the present findings, it has been suggested (e.g., Grodner et al., 2010) that these judgment studies may merely reflect that establishing the falsity of the *some but not all* interpretation requires seeking and failing to find a subset of elephants who do not have trunks, while agreeing with the *at least some* interpretation requires only conjuring up a few previously encountered elephants who had trunks. That is, the former is a more complex process and may be expected to take more time.

Three visual world studies have attempted to study the interpretation of *some* more directly, reporting divergent results (Grodner et al., 2010; Huang and Snedeker, 2009a, *in press*). In Huang and Snedeker (2009, *in press*) participants were asked to select "the girl that has some of the soccer balls" from a scene including: a girl with 2 soccer balls, a boy with 2 soccer balls, a girl with three socks, and a boy with nothing.[5] If participants calculate the scalar implicature, then they should infer that the sentence refers to the girl with the subset of soccer balls after the onset of *some*. Instead, Huang and Snedeker (2009, *in press*) found that participants considered both girls as possible referents – consistent with the literal *at least some* interpretation – until roughly one second after the onset of *some*. In control trials, in which numerical quantifiers were used (*the girl that has two of the soccer balls*), participants rapidly accessed the upper bound and ruled out the competitor (the girl with three socks). In contrast, Grodner et al. (2010) recently reported a very similar experiment which showed no evidence for a delay in calculating scalar implicature for *some*: a preference for the subset match began emerging about 200-300 ms after the onset of *some*.

[5] Some researchers (e.g., Breheny, 2009) have suggested that the processes that give the *some but not all* interpretation to "some" in these definite descriptions is different from the processes that underlie other scalar implicatures. As of yet, none of the experimental data suggest any distinction between the phenomena, so here we treat the phenomenon inclusively. We recognize, however, that what counts as an implicature will depend on your theory of implicature, and that these inferences would not be implicatures on theories that posit that implicatures **necessarily** result in globally more informative utterances.

Both groups of researchers have concluded that differences result from differences in the design of the visual stimuli and the types of filler trials that were employed, but there is no agreement as to which set of findings present the more accurate view of scalar implicature (Degen & Tanenhaus, 2010; Grodner et al. 2010; Hartshorne, Khan & Snedeker, 2010; Huang & Snedeker, 2009b). Grodner and colleagues have suggested that the presence of trials with numbers in Huang & Snedeker's experiments makes *some* a less natural description of a small set than it would otherwise be, which makes it more difficult for participants to interpret *some* as referring to either of the possible sets in the display, resulting in slower processing. In contrast, Huang and Snedeker have suggested that in the absence of numbers participants verbally encode each set as it is presented, since they quickly learn that subsets will generally be described with *some* and total sets with *all* (see also Hartshorne et al., 2010). Thus processing of the scalar term can sometimes precede its realization in the input, such that at the point when the participants hear the critical scalar term (*some*) there is an existing connection between that term and particular items in the display (e.g., the girl with the subset of soccer balls) and thus reference restriction is more rapid. In the absence of an explicit proposal of the mechanism by which naturalness influences processing, it is unclear whether these stories are truly distinct.

In this context, it is relevant that we find slow emergence of the scalar implicature, consistent with Huang and Snedeker (2009, *in press*) and the previous judgment studies (Bott & Noveck, 2004; Feeney et al., 2004; Noveck & Posada, 2003). Moreover, this pattern is robust. As noted above, we conducted an exact replication of Experiment 1, which similarly showed that *the rest* was more difficult to integrate in the conditional sentences at 2.5 seconds after *some*. To ensure that our failure to find a similar effect in Experiment 2 was due to the shorter distance between *some* and *the rest* (approximately 900 ms) and not due to irrelevant design issues, we conducted two modified versions of Experiment 2, adjusting design parameters in several ways (see above). Neither showed any evidence for scalar implicature computation prior to the probe, consistent with the results of Experiment 2.

It is not clear how the present findings would be explained on Grodner and colleagues hypothesis. There are no number trials, and our norming study indicating that some was perceived as natural in both the downward and upward entailing contexts. However, the present findings are clearly compatible with Huang's hypothesis that rapid scalar implicature requires verbal pre-encoding. In the present study, there were no pictures that could set in motion the process of verbal encoding prior to sentence onset. The use of filler trials ensured that presence of the scalar trigger, the probe phrase, and the need for implicature could not be predicted.

Mechanisms like verbal encoding are often construed of as strategies or confounds, which should be rooted out of experiments. In this case, we think that this construal is misleading. Any strategy that can be built up this quickly must have a powerful basis in the existing cognitive system and thus has the potential to be used in ordinary pragmatic processing. Thus we see this mechanism as concrete hypotheses about how context and ongoing language processing may be integrated. The degree of situational constraint varies widely across different instances of language use. In some situations, we can predict what the speaker is likely to say next, or how they might refer to a particular entity. In such contexts, predictions of this kind may often allow us to get a head start on processes which would otherwise be quite slow.

The discussion in the previous section suggests an interpretation of our finding that the effect of context on scalar implicature calculation is relatively delayed: the effects of a context that blocks scalar implicatures cannot emerge any faster than the scalar implicature otherwise would, and current evidence suggests scalar implicature processing (at least in non-predictable contexts) is slow.

Is scalar implicature processing gated by context?

We found no difference in reading times at the scalar quantifier between our scalar implicature-supporting and scalar implicature-nonsupporting contexts.[6] There are several logical explanations: a) the contextual manipulation we employed may be too weak to show an effect on *some*, b) reading time may not be sensitive to the cost of scalar implicature calculation even if these costs exist, or c) the processing costs in both contexts may be similar.

There are reasons to doubt the possibility that our manipulation was too weak to elicit a detectable effect. On many theories, our manipulation of entailment context would be expected to produce more rapid and robust effects than more idiosyncratic contextual constraints, such as those employed by BKW. This is trivially true for the Grammatical Theory, which gives precedence to this systematic semantic variable (Chierchia, 2006; Chierchia et al., *in press*). But it would also be true on any theory in which the frequency of a contextual cue influences processing speed (since all utterances, except questions and commands, are either downward or upward entailing). The antecedent of a conditional generally expresses a state of affairs which is not currently known to be true or false (otherwise why condition on it). Thus this contextual manipulation would be expected to have a powerful effect in any theory in which implicature generation depends upon beliefs about the speakers knowledge of the event. Moreover, our entailment manipulation did lead to a robust effect detectable by self-paced reading: the effect at

[6] Note that Panizza et al.'s (2010) eyetracking-while-reading study of number interpretation found that participants fixated numbers in declarative sentences 14 ms longer than in conditional sentences, an effect significant for conditional regression path but not regression path, gaze duration or first fixation duration. We note again that the relationship between studies of numbers and scalar quantifiers is unclear (Breheny, 2008; Horn, 2003). Moreover, they did not include the equivalent of our *only some* controls, so any differences may be due to differences between declarative and conditional sentences *per se*.

the rest. Thus one must explain why the manipulation was sufficient to produce this downstream effect but not any effect at *some*. In this context, it is worth noting that self-paced reading is known to be sensitive to a wide array of processing considerations (*inter alia:* Gordon et al., 1993; McNamara & Healy, 1988; Trueswell, Tanenhaus & Kello, 1993; Zola, 1984), many of which are quite subtle (e.g., transitional probabilities), and for this reason self-paced reading remains a key, commonly-utilized methodology in psycholinguistics (e.g., Amato & MacDonald, 2010; Carreiras, Dunabeitia, Vergara, de la Cruz-Pavia & Laka, 2010; Goldwater, Markman & Stilwell, 2011; Rohde, Levy & Kehler, 2011). Thus, there is no reason believe that any other contextual manipulation of scalar implicature processing would be stronger, though the possibility cannot be logically excluded.

The second possibility discussed above is that (b) the underlying mechanisms probed by self-paced reading (e.g., the factors that lead to the decision to press a button and receive the next chunk of text) are different from and unaffected by those which govern scalar implicature processing *per se*. That is, the mechanisms involved in strengthening the reading of *some* are in fact quite costly, but their cost is born by an independent portion of the mind/brain which does not directly interact with self-paced reading. To make this proposal more plausible, one would want a well-specified model of both processes and explain why they do not interact, while nonetheless allowing the *output* of scalar implicature processing (the strengthened reading of *some*) to eventually affect self-paced reading downstream at *the rest*. This proposal also motivates searching for convergent evidence from different paradigms, such as EEG.

Finally, it is (c) possible that failing to calculate the scalar implicature (strengthen *some*) may be just as costly as calculating it. Such a finding is consistent with a number of theoretical perspectives. Chierchia et al. (*in press*) suggest that the grammar always attempts to insert the grammatical *only* operator – the operator which drives scalar implicatures on their Grammatical Theory – but this operator is (typically) only retained in the grammatical structure when it leads to logically strengthened meanings, which is not the case in downward-entailing contexts such as conditionals. On Relevance Theory (Noveck & Sperber, 2007; Sperber & Wilson, 1989), processing ceases when a threshold of "sufficient" relevance has been achieved. Whether more processing is required to achieve sufficient relevance in our scalar impliccature supporting contexts than the scalar implicature nonsupporting contexts is an open *theoretical* question for Relevance Theory, and one could certainly imagine a resolution consistent with the present data.

Conclusion

In the above paper, we investigate online contextual dependence of scalar implicature processing using a carefully-controlled, non-confounded contextual manipulation. We find that scalar implicature computation is in fact modulated by context online, presenting a strong challenge to the framework proposed by Levinson (2000). However, our findings suggest that this contextual dependence -- and, indeed scalar implicature computation in general -- is sluggish, consistent with a broad range of studies reporting slow, effortful scalar implicature calculation (Bott & Noveck, 2004; Chevallier et al., 2008; Chierchia et al., 2001; De Neys & Schaeken, 2007; Feeney et al., 2004; Huang & Snedeker, 2009a, 2009b, 2009c; Noveck & Posada, 2003; Pouscoulous, Noveck, Politzer & Bastide, 2007). One important remaining question is how contextual information beyond entailment context (e.g., Bonnefon et al., 2009) is integrated into scalar implicature calculation online.

References

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, 116, 143-148.

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge, UK: Cambridge University Press.

Benz, A. & van Rooij, R. (2007). Optimal assertions, and what they implicate. A unified approach. *Topoi*, 26, 63-78.

Bezuidenhout, A., & Morris, R. (2004). Implicature, relevance and default inferences. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 257-283). Basingstoke: Palgraev Macmillan.

Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, 34, 433-456.

Bonnefon, J.-F., Feeney, A., Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112, 249-258.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: the onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437-457.

Breheny, R. (2009). How pragmatic reasoning affects incremental utterance interpretation. Talk given at Experimental Pragmatics 2009, Lyon, France.

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), 434-63.

Brown, P. & Levinson, S. (1987). *Politeness: Some Universals in Language Use*. Camridge, England: Cambridge University Press.

Carreiras, M., Dunabeitia, J. A., Vergara, M., de la Cruz-Pavia, I. & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, 115, 79-92.

Chemla, E. (2009). Presuppoisitions of quantified sentences: experimental data. *Natural Language Semantics*, 17, 299-340.

Chevallier, C., Wilson, D., Happe, F., & Noveck, I. (2010). Scalar inferences in Autism Spectrum Disordes. *Journal of Autism and Developmental Disorders*, 40, 1104-1117.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V. & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, 61, 1741-1760.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the 'logicality' of language. *Linguistic Inquiry*, 37(4), 535-590.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Luisa, M. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. *Proceedings of the 25th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press.

Chierchia, G., Fox, B., & Spector, D. (*in press*). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn & K. von Heusinger, (eds.), *Handbook of Semantics*. Mouton de Gruyter

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology*, 11, 430-477.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology*, 54(2), 128-33.

Degen, J. & Tanenhaus, M. K. (2010). Naturalness of lexical alternatives influences interpretation of ‘some’. Poster presented and CUNY Conference on Human Sentence Processing, New York, USA.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, 58(2), 121-32.

Frazier, L., Clifton Jr., C., & Stolterfoht, B. (2008). *Cognition*, 106, 299-324.

Frisson, S., Rayner, K. & Pickering, M. J. (2005) Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 862-877.

Gazdar, G. (1977). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, 2(4), 561-566.

Geurts, B. & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics & Pragmatics*, 2, 1-34.

Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, 118, 359-376.

Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse context. *Journal of Memory and Language*, 34(2), 216-231.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 331-347.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grodner, D., Klein, N., Carbary, K., & Tanenhaus, M. K. (2010). “Some” and possible all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116, 42-55.

Hartshorne, J. K., Khan, M., & Snedeker, J. (2010). Predictability in the visual world. Paper presented at Architectures and Mechanisms in Language Processing (AMLaP) 2010, York, U.K.

Hirschberg, J. (1985). *A Theory of Scalar Implicature*. Uniersity of Pennsylvania.

Holtgraves, T. & Yang, J.-N. (1990). Politeness as universal: cross-cultural perceptions af request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, 59, 719-729.

Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. University of California at Los Angeles: PhD thesis.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58(3), 376-415.

Huang, Y. T., & Snedeker, J. (2009b). Some questions are still unresolved: prosody, predictability, and speed of scalar implicatures. Presented at the 2009 Experimental Pragmatics Conference, Lyon, France.

Huang, Y. T., & Snedeker, J. (2009c). Semantic meaning and pragmatic interpretation in five-year olds: Evidence from real time spoken language comprehension. *Developmental Psychology*, 45(6), 1723-1739.

Huang, Y. T. & Snedeker, J. (*in press*). *Logic and Conversation* revisited: Evidence for a division between semantic and pragmatic content in real time language comprehension. *Language and Cognitive Processes*.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: The case of scalar implicature. *Synthese*, 165, 385-401.

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Coversational*

Implicature. Cambridge, MA: MIT Press.

McNamara, T. P., & Healy, A. F. (1988). Semantic, phological, and mediated priming in reading and lexical decisions. *Journal of Experimental Psychology: Leaning, Memory, and Cognition*, 14(3), 398-409.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Jouranl of Memory and Language*, 63, 324-346.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, 78(2), 165-88.

Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning*, 8(4), 297-326.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language*, 85(2), 203-10.

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12(11), 425-431.

Noveck, I. A. & Sperber, D. (2007). The why and how of experimental pragmatics: The case of ‘scalar inferences.’ In N. Burton-Roberts, N. (Ed.), *Pragmatics*. Basingstoke, UK: Palgrave Macmillan, pp. 184-212.

Panizza, D., Chierchia, G., & Clifton Jr., C. (2009) On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*, 61, 503-518.

Panizza, D., Huang, Y. T., Hartshorne, J. K., Snedeker, J., & Chierchia, C. (*in prep*). *Some* times words have *two* meanings. A study on the on line interpretation and processing of number words and scalar quantifiers.

Pinker, S. Nowak, M. A. & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences USA*, 105, 833-838.

Pouscoulous, N., Noveck, I. A., Plitzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in scalar implicature calculation. *Language Acquisition*, 14, 347-375.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, 118, 339-358.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, 2, 447-457.

Strawson, P. F. (1952). *Introduction to Logical Theory*. Oxford, England: Wiley.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528-553.

Winterstein, G. (2009). Argumentative properties of pragmatic inferences. *New Frontiers in Artificial Intelligence*, 5447, 161-176.

Zola, D. (1984). Redundancy and word perception during reading. Perception & Psychophysics, 36(3), 277-284.