The Neural Computation of Scalar Implicature

Joshua K. Hartshorne (jkhartshorne@gmail.com)
Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue
Cambridge, MA 02139 USA

Jesse Snedeker (snedeker@wjh.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St.
Cambridge, MA 02138 USA

Albert Kim (albert.kim@colorado.edu)
Institute of Cognitive Science, University of Colorado-Boulder, 594 UCB
Boulder, CO 80309 USA

Abstract

What psychological and linguistic processes allow one to go beyond the literal meaning of a sentence and infer what was meant but not said ("reading between the lines")? Theorists have differed as to whether these phenomena are driven by complex, online inference processes or by relatively rote rules. The present study uses ERP to explore the cognitive and neural mechanisms involved in scalar implicature (the inference that, e.g., "some" indicates "some but not all"), a test case that has been subject to considerable behavioral research but limited neuropsychological research. Our results challenge both rote-processing and rich-inference accounts. We provide the first ERP results showing that scalar implicature processing depends on context, challenging rote-processing theories of implicature. However, we also fail to find evidence of a processing cost associated with implicature processing, as predicted by many rich-inference accounts. These results point to a novel conceptualization of pragmatic processing in scalar implicature.

Keywords: language; pragmatics; ERP; scalar implicature

Rich vs. Rote Pragmatics

Understanding language appears to involve two broad but distinct kinds of processes: derivation of the semantic meaning (those things entailed to be true) and pragmatic inferences that go beyond this "literal" meaning (Bach, 1999; Grice, 1989; Morris, 1938). 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 about how to distinguish the representations and processes underlying each, as well as their interaction. One particularly contentious point is whether pragmatic inferences result from complex, rich reasoning processes (Grice, 1989; Sperber & Wilson, 1986) or from relatively rote, automatic, almost grammatical rules (Chierchia, Fox, & Spector, 2012; Levinson, 2000).

Perhaps the most-researched test case is scalar implicature, illustrated in (2):

(2) a. John ate some of the cookies.
b. John did not eat *all* of the cookies.

When we hear a sentence like (2a), we typically assume that (2b) is true as well. Although this inference is robust, it can be cancelled (3a), distinguishing it from semantic entailments, which cannot be cancelled (3b) (Hirschberg, 1991; 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.

On the classical view (Horn, 1972), scalar implicature requires rich online counter-factual reasoning: Listeners only infer (2b) from (2a) if they believe i) the speaker knows whether John ate all the cookies, ii) it is relevant whether John ate all the cookies, and iii) assuming (a-b) hold, the speaker would tell them that John ate all the cookies. This view has been questioned, originally by Levinson (2000), who argued that scalar implicatures are triggered automatically, prior even to compositional processing (i.e., processing language at the level of phrases or sentences).

Much of the work addressing this theoretical dispute has been indirect, testing whether scalar implicatures are slow and computationally costly as a proxy for being rich and complex (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009). Results have been inconsistent and controversial. More problematically, the link between "slow and costly" and "rich and complex" can be disputed: Grodner and colleagues (2010) argue that scalar implicature is rich, complex, and *fast*; similarly nothing in principle forbids an automatic process from being slow.

A more direct route is as follows: If scalar implicature is the result of a complex inference process, it should be possible to create contexts in which scalar implicatures are more or less likely to be calculated. If, on the other hand, scalar implicature is an automatic process, it should be relatively impervious to context. A handful of behavioral studies have reported contextual manipulations that affect scalar implicature calculation (Bergen & Grodner, 2012; Bonnefon, Feeney, & Villejoubert, 2009; Hartshorne & Snedeker, submitted; Noveck, Chierchia, Chevaux, Guelminger, & Sylvestre, 2002); we return to these in the Discussion. Nothing is known on a neuropsychological level about scalar implicature's context sensitivity, as no such studies have been reported.[1] Thus, we conducted the present study in order to confirm the behavioral results and extend them to the neuropsychological level.

Grammatical Context

We compare the ERPs elicited by carefully matched sentences that do or do not evoke scalar implicatures. Our method derives from previous findings that scalar implicatures are more likely to be calculated in declarative sentences (4a) than in the antecedent of a condition (4b).

(4) a. Addison ate **some** of the cookies before breakfast this morning, and…
b. If Addison ate **some** of the cookies before breakfast this morning, then…

This has been explained by Chierchia and colleagues (2012) as an effect of entailment context. Scalar implicature usually operates to deny the truth of a logically stronger statement. Since *Addison ate **all** of the cookies* entails that *Addison ate **some** of the cookies*, stating the latter implicates that the former is not true. In contrast, *If Addison ate **all** of the cookies, then Q* does not entail but rather is entailed by *If Addison ate **some** of the cookies, then Q*; thus stating the latter does *not* implicate that the former is not true.

While this entailment manipulation is linguistic in nature, it is nonetheless difficult to account for on a strict automatic processing account like Levinson's (2000), on which scalar implicature is triggered lexically prior to any compositional processes – that is, before sentential context, which is by definition compositional, can play a role.

While intuitions that conditional sentences suppress implicature seem robust, experimental evidence consists of a single published judgment study (Noveck et al., 2002).[2] Thus at best we do not know whether the entailment manipulation in (4) affects scalar implicature online.

The present study addresses this gap as follows:

(5) a. Addison ate some of the cookies before breakfast this morning, and **the rest** are on the counter.
b. If Addison ate some of the cookies before breakfast this morning, then **the rest** are on the counter.

[1] The two previous ERP studies investigated sentences of the form *some elephants have trunks* – literally true but rendered infelicitous by scalar implicature (Nieuwland, Ditman, & Kuperberg, 2010; Noveck & Posada, 2003). The final word evokes an N400 relative to the final word in felicitous sentences (*some dogs have spots*), at least if the sentences are carefully matched.

[2] Panizza, Chierchia, and Clifton Jr. (2009) report an eyetracking-while-reading study with a similar manipulation, but involving number. The relationship between number and scalar implicature is complex, unclear, and controversial.

Note that *the rest* is only felicitous if Addison has not eaten all of the cookies, which is exactly what the scalar implicature implies; thus, by hypothesis *the rest* should be more felicitous in (5a) than (5b). Thus, by testing whether the ERPs to *the rest* are different in (5a) and (5b), we test whether entailment context rapidly modulates scalar implicature, affecting interpretation of content later in the sentence. In addition, by comparing ERPs at *some* – the word that triggers the scalar implicature – we will gain valuable information about the neural processes supporting scalar implicature calculation.

One methodological concern remains: Declarative and conditional sentences differ in numerous ways, not just in how they affect scalar implicature. Thus, any differences observed may be due to implicature-irrelevant processes. Thus, we included a control version of the experiment, where *some* was everywhere replaced with *only some*, a phrase that semantically forces the subset ("not all") reading. Thus, the crucial analyses are interactions – differences seen between the critical declarative and conditional sentences not seen between the control declarative and conditional sentences.

Method

Subjects

49 monolingual native English-speaking right-handed individuals participated. Two were excluded for equipment failure and ten for excessive artifact, leaving 19 in the experimental condition and 16 in the control condition.

Materials and Procedure

Each participant saw 30 critical declarative sentences and 30 critical conditional sentences. Filler sentences consisted of 60 matched in structure – but not content – to the critical sentences but with continuations that did not mention "the rest" and 35 which additionally swapped the word *some* for *all*. These fillers prevented subjects from inferring that all sentences would refer to "the rest" of a previously-mentioned collection. An additional 42 filler sentences involved relative clauses and no quantifiers. Four lists were created by converting the critical declarative sentences into conditional sentences and *vice versa* and by reversing the order of all stimuli (except the first four stimuli, which were always the same fillers). The four experimental and four control lists were identical except that in the latter, the word *some* was always preceded by *only*.

Sentences were presented in eight blocks, with breaks in between. 61 of the sentences were followed by comprehension questions, which were not analyzed. Sentences were presented roughly one word at a time. Wherever two short words appeared consecutively, we presented them together (e.g., *Sally/saw/a cat/on the/table*). This allowed us to present the critical phrase *the rest* as a single unit, rather than in two parts which would potentially add noise to the ERP. *Some* was always presented singly. Stimuli were presented in the center of the screen for 350 ms with a 250 ms blank interval between words. The inter-trial interval ranged from 1600 to 2000 ms, not counting any time spent on questions.

Acquisition and Analysis

Ongoing EEG was recorded from 128 scalp locations using a geodesic sensor net (Electrical Geodesics, Eugene, OR) as subjects read the sentences silently. EEG was recorded relative to a vertex channel and later re-referenced to the average of the mastoid channels. Impedances were maintained below 100 Ω. Signals were recorded at 250 Hz and down-sampled to 200 Hz post-acquisition. A 0.1-30 Hz bandpass filter was applied. Epochs of 1500 ms were selected following the critical phrase (*some* or *the rest*) and were corrected with a 200 ms pre-stimulus baseline. Bad channels were replaced and epochs containing artifact (eye blink, eye movement, etc.) removed, both by computer algorithm. Only participants with at least 19 epochs per cell were included in analyses.

The Bootstrap Cluster Algorithm

The previous literature has focused on the role of the N400 in processing scalar implicature violations. Because no previous study has looked for components indexing scalar implicature *generation*, we needed a mechanism for selecting and analyzing exactly those electrodes in those time periods with the greatest differences between conditions without allowing multiple comparisons to inflate our Type I error rate (cf Vul, Harris, Winkielman, & Pashler, 2009). We adapted the recently developed bootstrap cluster analysis of Maris and Oostenveld (2007).

We calculated the context by condition interaction using a mixed effects model with maximal random effects for each electrode at each time point (to speed processing, we further down-sampled the data to 50 Hz) and recorded the t-value. We then identified all clusters of data points with t-values greater than 1.96 or less than -1.96.[3] Clustering crossed both time (consecutive super-threshold data points on the same electrode were placed in the same cluster) and space (super-threshold data points from the same time point and belonging to neighboring electrodes were placed in the same cluster). Although data points with positive effects (positive t-values) may represent the same underlying dipole as data points with negative effects (negative t-values), we adopted the conservative approach of placing t-values of different polarities in different clusters. Clusters are assigned scores, which are the sum of their t-values; thus, clusters with larger statistical effects and/or which are extended in time and space are assigned larger scores.

Statistical significance was assessed through bootstrapping. The condition labels for the subjects (experimental/control) were shuffled, as were the context codes (declarative/conditional) for each subject's average ERPs. The clustering algorithm was re-run, and the scores for the largest positive and negative clusters were recorded. This process was repeated 100 times. P-values for a given cluster in the actual data are estimated as the number of *larger* clusters from the bootstrapped data (calculated separately for positive and negative clusters).

Results

As can be seen in Figure 1, the interaction between context and condition in the ERPs evoked by *some* were weak, and none of the resulting clusters were significant (*p*s>.2).

In contrast, at *the rest* an interaction was observed, frontally distributed and lasting from approximately 400 to 1300 ms post-stimulus (*p*=.04; see Figures 2 & 3). Inspection of the four waveforms for the four conditions revealed at this interaction was due to a positive deflection for the conditional/experimental sentences relative to the other three conditions. That the conditional/experimental sentences should be the odd one out is expected: only in that condition should *the rest* be difficult to process, and in fact in our norming studies, the conditional/experimental sentences were judged to be less felicitous than the other three types; this effect disappeared if the sentences were truncated prior to *the rest*.

Thus, we interpret the interaction at *the rest* to be due to a positive deflection for the conditional/experimental sentences, reflecting the infelicity of *the rest*, perhaps due to the difficulty assigning its reference.

Discussion

A previous judgment study (Noveck et al., 2002) found that scalar implicatures were more likely in declarative than conditional sentences. If this is the case, and if this contextual manipulation affects processing rapidly, then *the rest* should be more difficult to process in (5b) than (5a). Indeed, we found that the contextual manipulation affected the ERPs to *the rest*. Interestingly, we did not find an effect of the manipulation on the ERPs to the scalar trigger *some*.

We address theoretical implications of these findings momentarily. First, we consider their robustness. Given recent concern about replicability in the cognitive sciences (Hartshorne & Schachner, 2012), we conducted a replication closely matched to the above experiment with the following differences: EEG was recorded using Ag/AgCl electrodes attached to an elastic cap following the extended 10-20 system (Newer et al., 1998), and blink artifact was corrected through linear regression. We coded the stimuli so that the ERPs to *some* in the filler sentences – which up through *some* are indistinguishable from the critical sentences – could be included in analysis, doubling the number of trials for that analysis. Analyses were conducted in identical fashion and with the same result, demonstrating their robustness: no significant clusters were found at *some* (*p*s>.15), but an extended, frontally-distributed cluster was found after *the rest* (*p*<.01).

[3] The choice of threshold (e.g., 1.96) affects the type of clusters found – low thresholds are better at detecting broadly extended but weak effects – but does not affect robustness to multiple comparisons. Other threshold resulted in similar findings.

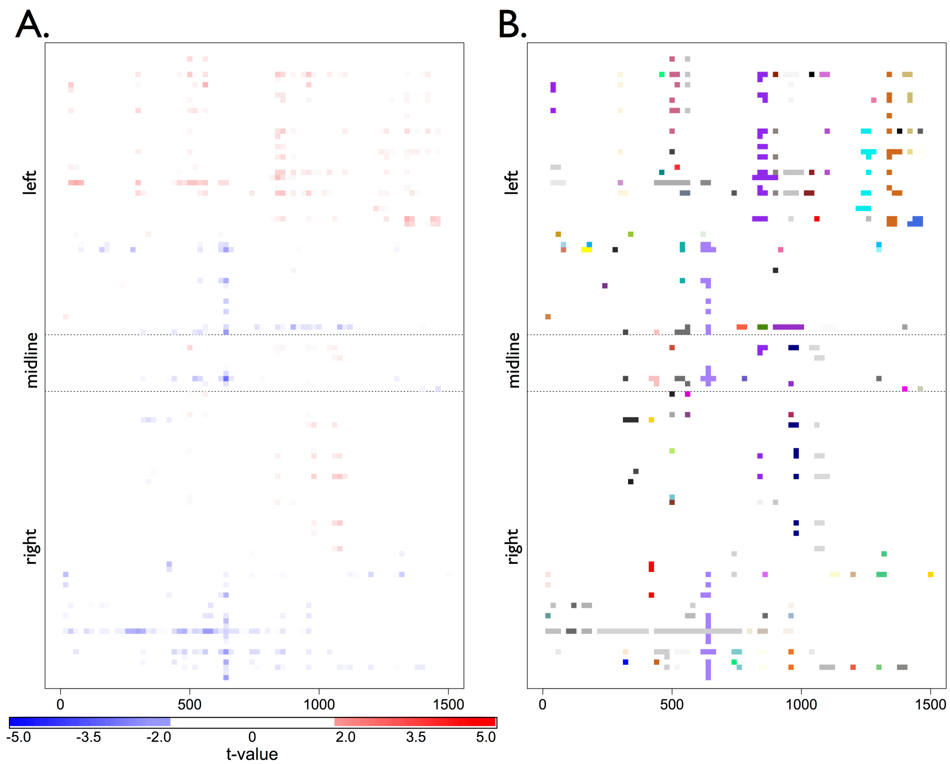

Figure 1: Bootstrap cluster analyses at *some*. In each panel, electrodes grouped into left-hemisphere, midline, and right-hemisphere, with more anterior electrodes placed higher. **Panel A**: *t*-values. **Panel B**: clusters (distinct color for each cluster).

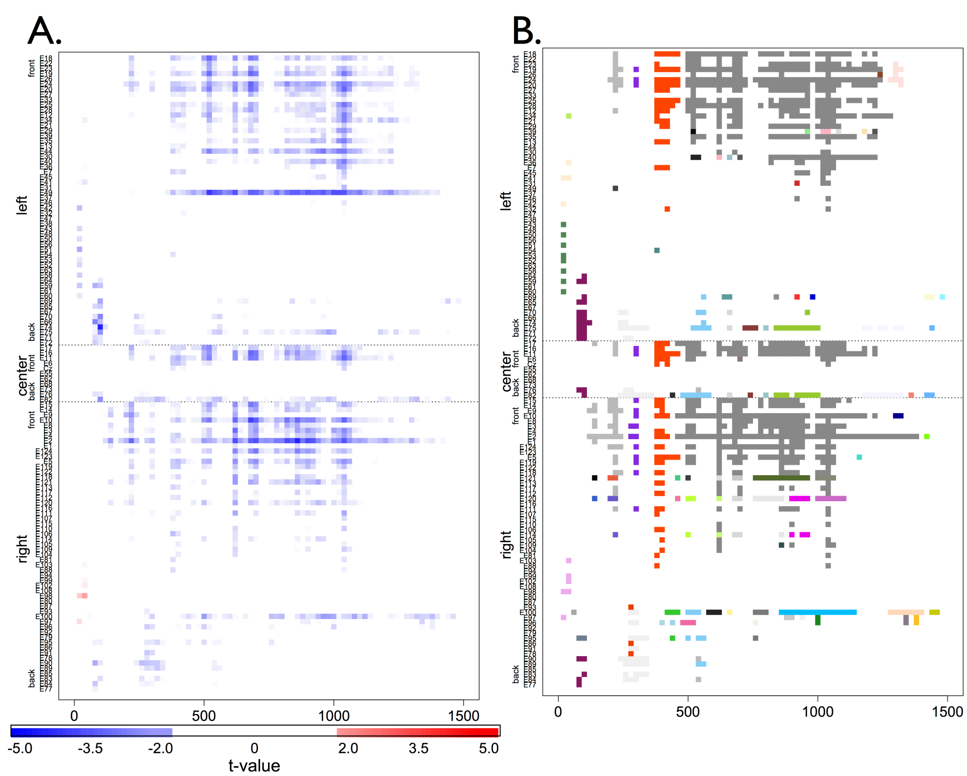

Figure 2: Bootstrap cluster analyses at *the rest*. In each panel, electrodes grouped into left-hemisphere, midline, and right-hemisphere, with more anterior electrodes placed higher. **Panel A**: *t*-values. **Panel B**: clusters (distinct color for each cluster).

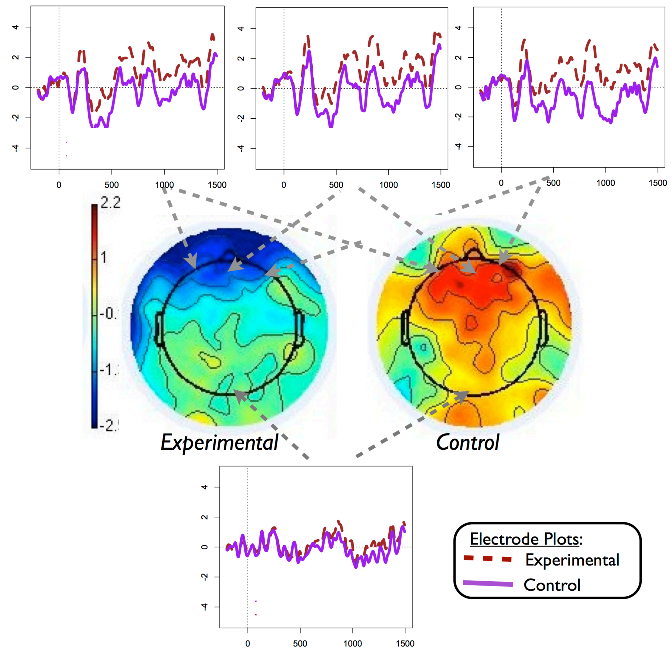

Figure 3: Difference waves (declarative – conditional) at *the rest*. Topographical plots are shown at 600 ms post-stimulus. Four representative electrodes are depicted for the entire epoch. The relative negativity for the difference waves in the experimental sentences is driven by a *positive* deflection for the conditional sentences (see main text).

We consider first the positive results. As predicted, scalar implicatures are less likely to be calculated in the antecedents of conditionals (as evidenced by results at *the rest*), confirming a strong prediction of Chierchia and colleagues' Grammatical Theory. Moreover, context's effect was sufficiently rapid to affect processing of content (*the rest*) later in the sentence. This is difficult to reconcile with a strong lexicalist position like Levinson's (2000), on which scalar implicatures are always triggered by words like *some*, though they may be explicitly cancelled as in (3a). Note that not only was the implicature not cancelled in our conditional sentences, not calculating the implicature renders the sentences infelicitous.

Perhaps the most intriguing finding was the lack of any effect at the scalar implicature trigger *some*. This finding matches those of five self-paced reading experiments involving similar stimuli, for which Hartshorne and Snedeker (submitted) similarly report no effect: *some* was read no faster or slower whether a scalar implicature was calculated or not. These finding are in apparent conflict with a single experiment by Bergen and Grodner (2012), which showed slower reading times for *some* in implicature-promoting conditions, which they interpreted as indexing the computational cost of scalar implicature calculation. However, Bergen and Grodner used a different manipulation, an issue we return to shortly.[4]

There are at least three logically possible explanations of the result. The first is that the scalar implicature processing's effect on ERP (and self-paced reading) is quite small and thus we had insufficient statistical power to detect it. This would raise an interesting question: Why is the effect so small relative to typical language ERP effects (such as the effect we observed on *the rest*) which *are* observable with a study this size?

A second possibility is that scalar implicature is always triggered by *some*, and thus the ERPs were identical across conditions (Levinson, 2000). As already noted, this runs afoul of our results at *the rest*; we would have to stipulate that the entailment context acts to cancel the implicature in the conditional sentences. What the mechanism would be is unclear. Moreover, the effect of the cancellation should be measurable, and though we explored ERPs to several words subsequent to *some*, we saw no evidence of it.

A third possibility is that *not* computing the scalar implicature is also a complex and lengthy inference, sufficiently similar to actually computing the scalar implicature that the two could not be distinguished in our study. On the Grammatical Theory, the parser attempts to insert scalar implicatures at any appropriate insertion site, and they are retained if certain criteria are met, such as its resulting in a more informative (i.e., logically stronger) interpretation of the utterance. Presumably, the only way the grammar can know that these criteria have been met is to actually carry out the operations; thus, similar work is done whether the scalar implicature is ultimately endorsed or not. Similarly, on the Gricean account, scalar implicatures are calculated only when certain conditions are met (e.g., the speaker would make the stronger statement if it were true and the speaker knows whether the stronger statement is true). Whether these conditions are met affects whether the implicature is endorsed, not whether the complex set of conditions must be *checked.* In short, even if calculating a scalar implicature is costly, that does not necessarily mean that manipulations which affect whether the implicature is ultimately endorsed affect the computational cost.

Why then did Bergen and Grodner find an effect on *some*? While we manipulated whether the scalar implicature was appropriate, they manipulated the salience of the stronger alternative (e.g., *all*). Since scalar implicature processing requires a stronger alternative to get off the ground, their manipulation may have more directly affected whether processing happened at all.

Conclusion

We find the grammatical entailment context modulates scalar implicature processing rapidly enough to affect

[4] Breheny, Katsos, and Williams (2006) report longer reading times for scalar triggers in contexts expected to promote scalar implicature calculation. However, the contextual manipulations are uncontrolled, making its results difficult to interpret. In the case of Exp. 3, the manipulation is fully confounded with a repeated name penalty, sufficient to explain their results.

processing of subsequent words in the sentence. At the same time, this manipulation did not affect the EEG evoked by the scalar implicature trigger (*some*). These findings present a first step in uncovering the neural processes underlying the factors driving scalar implicature and also present challenges to existing theories.

Acknowledgments

The authors thank Miki Uruwashi and especially Stephanie Liem for help with data acquisition, and attendees at AMLaP 2012 for comments and suggestions. This work was supported by NDSEG and NSF GRFP fellowships to JKH and NSF-BCS 0921012 to JS.

References

Bach, Kent. (1999). The semantics-pragmatics distinction: What it is and why it matters. In K. Turner (Ed.), *The Semantics-Pragmatics Interface from Different Points of View* (pp. 65-84). Oxford, UK: Elsevier.

Bergen, Leon, & Grodner, Daniel J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1450-1460.

Bonnefon, Jean-Francois, Feeney, Aidan, & Villejoubert, Gaelle. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition, 112*, 249-258.

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

Breheny, Richard, Katsos, Napoleon, & Williams, John. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*, 434-463. doi: 10.1016/j.cognition.2005.07.003

Chierchia, Gennaro, Fox, Danny, & Spector, Benjamin. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, K. v. Heusinger & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 2, pp. 2297–2332). Boston, MA: De Gruyter Mouton.

De Neys, Wim, & Schaeken, Walter. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology, 54*, 128-133.

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

Grodner, Daniel J., Klein, Natalie M., Carbary, Kathleen M., & Tanenhaus, Michael K. (2010). "Some," and possible all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116*, 42-55.

Hartshorne, Joshua K., & Schachner, Adena. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience, 6*(8).

Hartshorne, Joshua K., & Snedeker, Jesse. (submitted). The speed of inference: Evidence against rapid use of context in calculation of scalar implicature.

Hirschberg, Julia. (1991). A Theory of Scalar Implicature. *book*, 1-220.

Horn, Laurence R. (1972). On the semantic properties of logical operators in English. *book*, 1-309.

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

Levinson, Stephen C. (2000). Presumptive Meanings: The Theory of Generalized Conversational Implicature, Speech, and Communication. *book*, 1-480.

Maris, Eric, & Oostenveld, Robert. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177-190. doi: 10.1016/j.jneumeth.2007.03.024

Morris, Charles. (1938). Foundations of the theory of signs. *International Encyclopdia of Unified Science, 1*(2).

Newer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., Hinrichs, H., . . . Rappelsburger, Peter. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology, 106*, 259-261.

Nieuwland, Mante S., Ditman, Tali, & Kuperberg, Gina R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Langauge, 63*, 324-346.

Noveck, Ira A, Chierchia, Gennaro, Chevaux, Florelle, Guelminger, Raphaele, & Sylvestre, Emmanuel. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning, 8*, 297-326.

Noveck, Ira A, & Posada, Andres. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and language, 85*, 203-210. doi: 10.1016/S0093-934X(03)00053-1

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

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

Vul, Edward, Harris, Christine, Winkielman, Piotr, & Pashler, Harold. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion , Personality , and Social Cognition. *Perspectives on Psychological Science, 4*, 274-290.