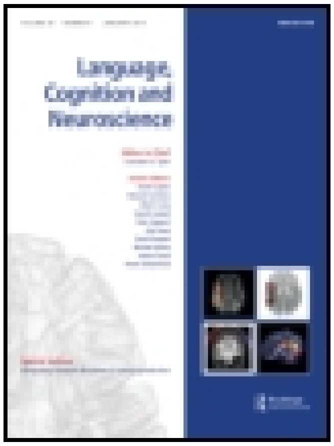

Language, Cognition and Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp21

The neural computation of scalar implicature

Joshua K. Hartshorne[ab], Jesse Snedeker[b], Stephanie Yen-Mun Liem Azar[b] & Albert E. Kim[c]
[a] Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[b] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA
[c] Institute of Cognitive Science and the Department of Psychology & Neuroscience, University of Colorado, Boulder, CO 80390, USA
Published online: 24 Dec 2014.

To cite this article: Joshua K. Hartshorne, Jesse Snedeker, Stephanie Yen-Mun Liem Azar & Albert E. Kim (2014): The neural computation of scalar implicature, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2014.981195

To link to this article: http://dx.doi.org/10.1080/23273798.2014.981195

Language, Cognition and Neuroscience, 2014
http://dx.doi.org/10.1080/23273798.2014.981195

The neural computation of scalar implicature

Joshua K. Hartshorne[a,b]*, Jesse Snedeker[b], Stephanie Yen-Mun Liem Azar[b] and Albert E. Kim[c]

[a]*Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA;* [b]*Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA;* [c]*Institute of Cognitive Science and the Department of Psychology & Neuroscience, University of Colorado, Boulder, CO 80390, USA*

(*Received 19 April 2014; accepted 22 August 2014*)

Language comprehension involves not only constructing the literal meaning of a sentence but also going beyond the literal meaning to infer what was meant but not said. One widely studied test case is scalar implicature: The inference that, e.g., *Sally ate some of the cookies* implies she did not eat all of them. Research is mixed on whether this is due to a rote, grammaticalised procedure or instead a complex, contextualised inference. We find that in sentences like *If Sally ate some of the cookies, then the rest are on the counter*, that *the rest* triggers a late, sustained positivity relative to *Sally ate some of the cookies, and the rest are on the counter.* This is consistent with behavioural results and linguistic theory suggesting that the former sentence does not trigger a scalar implicature. This motivates a view on which scalar implicature is contextualised but dependent on grammatical structure.

Keywords: scalar implicature; pragmatics; ERP

Understanding language is often divided into two types of processes: the derivation of the semantic meaning (those things entailed by the statement) and the calculation of pragmatic inferences that go beyond this literal meaning (Bach, 1999; Grice, 1989; Morris, 1938). For example, given sentence (1), the claim that Gabe is the agent of the drinking event is based on semantic decoding, while the inference that he is an inconsiderate lout who has annoyed the speaker is a pragmatic inference.

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

Pragmatics may seem to be a peripheral phenomenon – the occasional minor inference that should not distract from the meat of language (syntax and semantics). In fact, pragmatic inferences are pervasive, affecting many if not all communicative acts, and often comprising core aspects of the linguistic message (Horn & Wald, 2004). For example, consider the inference in (1) that Gabe put the carton back in the fridge *after* drinking all the milk. Without a full account of pragmatic inference, our psychological and neuroscientific understanding of language will be severely limited.

In recent years, there has been an explosion of experimental and theoretical work in pragmatic inference (for review, see Noveck & Reboul, 2008). Scalar implicature – the focus of our investigation below – has emerged as a particularly important test case. In scalar implicatures, we infer from one statement that a stronger claim is false (Hirschberg, 1991; Horn, 1972). Consider:

(2) John ate some of the cookies.

The literal meaning of this statement is that John ate a quantity of cookies that is greater than zero, which leaves open the possibility that he ate *all* of the cookies. Nevertheless, most listeners infer from (2) that John did *not* eat all of the cookies.

This inference (John did not eat all of the cookies) has a different status from the literal meaning (John ate a non-zero quantity of cookies). If someone stated (2), but it turned out that John had in fact eaten all of the cookies, the speaker could be accused of being misleading or imprecise, but not of lying. However, if it turned out that John had not eaten any cookies, the speaker was lying.

What accounts for scalar implicature? Beginning with Grice (1989), most theories incorporate the intuition that if John had eaten all of the cookies and the speaker knew it, then s/he would have said so:

(3) John ate all of the cookies.

One can infer from the fact that the speaker did not say (3) that it is not true.[1] Part of the explanation seems to lie in informativity: If John has eaten all of the cookies, (2) is true but less informative than (3). *Informativity* can be formalised in terms of asymmetric entailment: Whenever

*Corresponding author. Email: jkhartshorne@gmail.com

Table 1. Informativity (upward-entailing context).

Some P	All P
T	F
T	T
F	F

Note: Truth tables for (2) and (3).

(3) is true, (2) must also be true, whereas the reverse is not the case (Table 1).

Informativity: lexical or higher level?

An important open question is whether, for the purposes of scalar implicature, informativity is calculated at the lexical level (*some*) or at a higher level of representation, such as the entire utterance (*John ate some of the cookies*). A number of factors suggest a lexical account. One is that informativity accounts, without further constraints, can overgenerate. For instance, we do not normally infer from (2) that John does not like scuba diving, even though (4) is more informative than (2) (cf. Horn, 1972).

(4) John ate some of the cookies and likes scuba diving.

For this reason, many researchers, beginning with Horn (1972), have proposed that scalar implicatures derive from specific lexical scales, such as (*some*, *all*): Barring specific contextual support,[2] the only alternative utterances considered are those involving other members of the entailment scale (cf. Chierchia, Fox, & Spector, 2012; Gazdar, 1979; Hirschberg, 1991; Levinson, 2000; Sauerland, 2004). In addition to (*some*, *all*), a number of other scales have been proposed, such as (*or*, *and*) and (*warm*, *hot*).

The introduction of lexical scales raises the possibility that scalar implicatures are calculated at the lexical level: On encountering the weak scalar term *some,* listeners could retrieve the more informative alternative *all*, negate it and replace the original meaning with the enriched one (some-but-not-all). Below, we refer to this as the *lexical informativity account*, a position typically associated with Levinson (2000; see discussion in Breheny, Katsos, & Williams, 2006; Degen, 2013; Huang & Snedeker, 2009a, 2009b).[3] This proposal gains some support from the relative robustness of scalar implicatures, which, unlike many other linguistic inferences, are typically derived without specific contextual support.[4]

However, other researchers have presented theoretical arguments that scalar implicature should not be calculated at the lexical level. Importantly, the relative informativity of items on these scales (*some*, *all*) interacts with the broader propositional content of the sentence. Consider sentences like the following:

Table 2. Informativity (downward-entailing context).

Some P	All P	Q	If some P, Q	If all P, Q
T	F	T	T	T
T	F	F	F	T
T	T	T	T	T
T	T	F	F	F
F	F	T	T	T
F	F	F	T	T

Note: Truth tables for (6) and (7).

(6) If John ate some of the cookies, then he cannot have dessert.

(7) If John ate all of the cookies, then he cannot have dessert.

In this case, the sentence with *some* (6) is more informative than the sentence with *all* (7): Whenever (6) is true, (7) is true, but there are cases where (7) is true but (6) is not (Table 2). Thus, the relative informativity of the utterances is reversed relative to the relative informativity of the lexical items. Grammatical contexts like this are called "downward entailing" contexts, in order to distinguish them from the typical "upward-entailing" contexts, in which the relative informativity of the utterances and lexical items are matched.

If scalar implicature is calculated lexically, entailment context makes no difference and readers should get the *some-but-not-all* reading for (6). However, if scalar implicatures are calculated over higher level structure, then no implicature for (6) is expected, since there is no more informative utterance to negate. Indeed, this is what some theorists predict (Chierchia, 2006; Chierchia et al., 2012; Geurts, 2009; Geurts & Pouscoulous, 2009; Katsos, 2008; Sauerland, 2004).[5] Note that some of these accounts (e.g., Chierchia et al., 2012) have alternative mechanisms allowing for scalar implicatures within downward-entailing contexts. However, for all these *higher level informativity accounts*, scalar implicatures in downward-entailing contexts should be rare at best.

Although lexical-level processing is often assumed to be fast and automatic, in principle this distinction crosscuts the long-standing debate about whether scalar implicature is calculated by default or not (Bott & Noveck, 2004; Feeney, Scrafton, Duckworth, & Handley, 2004; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009a, 2009b, 2011; Levinson, 2000). There is now considerable evidence that scalar implicature machinery is deployed flexibly: Listeners are less likely to make scalar implicatures when they believe that the speaker is not motivated to be informative (Bonnefon, Feeney, & Villejoubert, 2009) or is unlikely to know whether the more informative statement is true (Bergen & Grodner, 2012). However, the machinery may be deployed flexibly but still operate over a lexical representation rather than over the proposition as a whole. In fact, the psycholinguistic studies suggesting rapid use of context are amendable to a purely lexical explanation. Specifically, in two reading studies that show immediate context sensitivity, the contexts that result in an implicature often (Bergen & Grodner, 2012) or always (Breheny et al., 2006) include *all* or a near synonym (e.g., *each*) earlier in the discourse, while the contexts that do not support implicature never do. Thus, the effects of context in these studies could result from the priming of the lexical scale.

Lexical or higher level informativity: experimental evidence

There is currently relatively little evidence to tell between the lexical and higher level accounts, and none that speaks to the role of grammatical context in the neural processing of implicature. To date, evidence has primarily come from studies in which participants make explicit judgements about the interpretation of scalar terms. In the first of these studies (Noveck, Chierchia, Chevaux, Guelminger, & Sylvestre, 2002), the critical utterances were part of an exercise in logical reasoning (11).

(11) If there is a P or a Q then there is an R.

There is a P.
There is a Q.
Is there an R?

If scalar implicatures are calculated over lexical scales ignoring higher level structure, then given the lexical scale <or, and>, *or* should be interpreted exclusively (*not both*), and the answer should be "no". On the higher level informativity account, we should expect participants to answer "yes". This is because "or" in the first sentence of (11) is in a downward entailing context, thus the implicature should be suspended, and *or* should be interpreted inclusively. The actual participants overwhelmingly answered "yes", supporting the higher level informativity account. Some subsequent judgement studies have confirmed this pattern (Chemla & Spector, 2011; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001), though not all have (Geurts & Pouscoulous, 2009).[6]

However, because these studies use explicit judgement tasks, in which participants consider the acceptability of the utterance or engage in explicit logical reasoning, they do not tell us how scalar inferences are made during every day, unreflective language comprehension. The present study explores the role of grammatical context during ordinary language comprehension, using measures that allow us to watch the process of implicature unfold in the brain over time.

Overview of the experiments

In the present work, we compare processing of the scalar implicature trigger *some* in declarative (upward-entailing) and conditional (downward entailing) sentences, using event-related potentials (ERPs). As we noted above, *higher level informativity accounts* predict that this manipulation should affect processing, whereas *lexical informativity accounts* do not.

This work expands and improves on previous research in several ways. First, as noted above, the experimental evidence thus far is limited to offline judgements, which can only tell us about final stage results. Moreover, most of these tasks have used highly repetitive abstract stimuli, which may have encouraged participants to treat the task as one of logical reasoning rather than linguistic communication, which could lead them to focus on logical entailment rather than lexical knowledge or communicative habits. Using an ERP paradigm allows us to assess interpretation without requesting explicit judgements about meaning. In addition, using a method that provides rich temporal information allows us to measure early interpretations that are less likely to be contaminated by explicit metalinguistic reasoning.

Second, despite an explosion of behavioural experiments on scalar implicature (Barner, Brooks, & Bale, 2010; Bonnefon et al., 2009; Bott & Noveck, 2004; Breheny, Ferguson, & Katsos, 2012; Chemla & Spector, 2011; Chevallier et al., 2008; Chevallier, Wilson, Happé, & Noveck, 2010; De Neys & Schaeken, 2007; Feeney et al., 2004; Goodman & Stuhlmüller, 2013; Foppolo, Guasti, & Chierchia, 2012; Grodner et al., 2010; Huang & Snedeker, 2009a, 2009b, 2011; Marty, Chemla, & Spector, 2013; Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Papafragou, 2006; Pouscoulous, Noveck, Politzer, & Bastide, 2007), relatively little is known about the neural computation of scalar implicature. Two studies have investigated the brain response produced when a scalar implicature conflicts with world knowledge. Nieuwland, Ditman, and Kuperberg (2010) presented participants with sentences made felicitous or infelicitous by a prior scalar implicature (*Some people have pets/? lungs…*), finding a larger N400 to the infelicitous word (*lungs*).[7] Politzer-Ahles, Fiorentino, Jiang, and Zhou (2013) manipulated the felicity of the word *some* itself: participants were presented with sentences such as *Some of the girls are sitting on blankets sun-tanning*, after they had seen either a picture in which some-but-not-all of the girls are sitting on blankets sun-tanning (making the scalar implicature-enriched interpretation of the sentence felicitous) or a picture in which all of the girls are sitting on blankets sun-tanning (making the scalar implicature-enriched interpretation of the sentence infelicitous). This resulted in a sustained negativity to *some* in the infelicitous condition. Thus, these studies measure processing when a scalar implicature-enriched interpretation mismatches prior knowledge, rather than manipulating whether the scalar implicature itself was calculated, which is the focus of the present study.

Experiment 1

The present study contrasts lexical informativity and higher level informativity accounts of scalar implicature by contrasting declarative sentences (12a) and conditional sentences (12b). As discussed above, if scalar implicatures are calculated based on higher level informativity, then the some-but-not-all interpretation should be available in (12a) but not (12b).

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

Note that in both cases, the most likely referent for *the rest* is the remaining cookies that Addison did not eat. However, according to the higher level informativity account, listeners do not normally calculate scalar implicatures in the antecedents of conditionals. As such, interpreting *the rest* in (12b) – but not in (12a) – requires retroactive calculation of the scalar implicature. This method of probing interpretation was borrowed from Breheny et al. (2006). Thus, interpreting *the rest* should be more difficult in (12b) than (12a). Associated ERPs may reflect the more involved search for the referent, the retroactive calculation of the scalar implicature or both. In addition, by comparing ERPs at *some*, the word that triggers the scalar implicature, we may gain valuable information about the neural processes supporting scalar implicature calculation.

In contrast, the lexical informativity account predicts that the some-but-not-all interpretation is calculated regardless of entailment context, and thus *the rest* should be equally easy to interpret in both cases.

One methodological concern remains: Declarative and conditional sentences differ in numerous ways, not just in how they affect scalar implicature. The critical phrase *the rest* is preceded by different connectives – *and* in (12a) and *then* in (12b) – and thus differences seen in processing of *the rest* may merely reflect late components of the ERP to the different connectives. The two clauses in (12b) depend upon one another in a way that the two clauses in (12a) do not. Determining the truth of a declarative depends on states of the world, whereas determining the truth of a conditional depends on possible states: If Addison has not eaten any cookies, (12a) cannot be true but (12b) might be. Differences in the ERPs between our declarative and conditional sentences may reflect these or other implicature-irrelevant factors.

To address this issue, we conducted a control version of the experiment, where *some* was replaced everywhere by *only some*, a phrase that semantically forces the subset ("not all") reading:

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

This allows us to control for differences between declarative and conditional sentences. Any differences between declarative and conditional sentences should affect both the experimental sentences (12) and the control sentences (13), whereas only the experimental sentences (12) manipulate scalar implicature processing. Thus, the crucial analyses are interactions: differences seen between the experimental declarative and conditional sentences (12) that are not seen between the control declarative and conditional sentences (13).

Method

Subjects

Forty-nine monolingual native English-speaking right-handed adults 18–38 years old participated. Two were excluded for equipment failure and 10 for excessive artefact, leaving 35 participants (17 female, 18–38 years old, $M = 21$ years old, SD = 4): 19 in the experimental condition and 16 in the control condition. Participants were recruited from the broader Harvard University community and were compensated with either partial course credit or a small payment.

Materials and procedure

We created 60 sentences, each of which could occur in four forms, as depicted in (12–13). Each participant saw 30 critical declarative sentences and 30 critical conditional sentences in *either* the experimental or the control conditions. Filler sentences consisted of 60 sentences matched in structure – but not content – to the critical sentences but with continuations that did not mention "the rest" and 35 that 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 for the experimental condition. Within each list, 30 sentences appeared in the declarative version and the other 30 in the conditional form. By counterbalancing the order of the stimuli (except the first four stimuli, which were always the same fillers) and counterbalancing which form the sentence appeared in (declarative or condition), we obtained the four lists. Four lists for the control condition were created as well, which were identical to the experimental lists, except that the word *some* was always preceded by *only* – both in the critical sentences and the filler sentences.

Sentences were presented in eight blocks, with breaks in between. Sixty-one of the sentences were followed by comprehension questions, which were not analysed. Sentences were presented in displays of 1–2 words 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 centre of the screen for 350 ms with a 250-ms blank interval between words. Immediately prior to the beginning of the next trial, subjects saw a blank inter-trial screen whose duration varied randomly in a range of 1600–2000 ms. The aim of the variable ISI was to reduce the degree to which oscillatory activity in the encephalogram (EEG) would entrain to the regular frequency of the stimuli presented in the sentences.

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 75 Ω. Signals were recorded at 250 Hz and down-sampled to 200 Hz post-acquisition using the signal processing toolbox's resample() function (Mathworks, 2012). We down-sampled to 200 Hz to facilitate comparison with other datasets that were already sampled at 200 Hz. 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 artefact (eye blink, eye movement, etc.) removed, both by computer algorithm. Bad channels were identified by hand, with an average of 1 (range: 0–5) replaced per participant. An additional automated screen of channels with spectral power more than 12.5 times the standard deviation did not identify any additional bad channels. Bad epochs were those with a maximal amplitude of ±125 μV from baseline on any channel. Only participants with at least 19 epochs per cell were included in analyses.

The permutation 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 associated with scalar implicature *generation*, we needed a mechanism for selecting and analysing 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 permutation cluster analysis of Maris and Oostenveld (2007).

We calculated the context (declarative/conditional) by condition (experimental/control) interaction using a mixed effects model with maximal random effects for each electrode at each time point and recorded the t value (to speed processing, we further down-sampled the data to 50 Hz, resulting in 20-ms windows). We then identified all clusters of data points with t values greater than 1.96 or less than −1.96.[8] 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 neighbouring electrodes were placed in the same cluster). 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 permutation analysis. 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 500 times. P values for a given cluster in the actual data are estimated as the number of clusters of equal or greater size from the permuted data (calculated separately for positive and negative clusters).

Results and discussion

At *the rest*, an interaction of condition and context was observed, frontally distributed and lasting from approximately 400 to 1300 ms post-stimulus (p = .048; see Figures 1 and 2). This is inconsistent with the lexical informativity account, on which no interaction was expected. The higher level informativity account did predict an interaction, driven extra processing of *the rest* in the conditional sentences in the experimental condition (12b). This result is consistent with our norming studies (Hartshorne & Snedeker, submitted), in which the experimental conditional sentences (12b) were judged to be less felicitous than the other three types (an effect which 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 experimental conditional sentences (12b), reflecting the added difficulty of interpreting *the rest*.

In contrast, the interaction at *some* was weak, and none of the resulting clusters were significant (ps > .47; Figure 3). This finding is somewhat surprising: The results

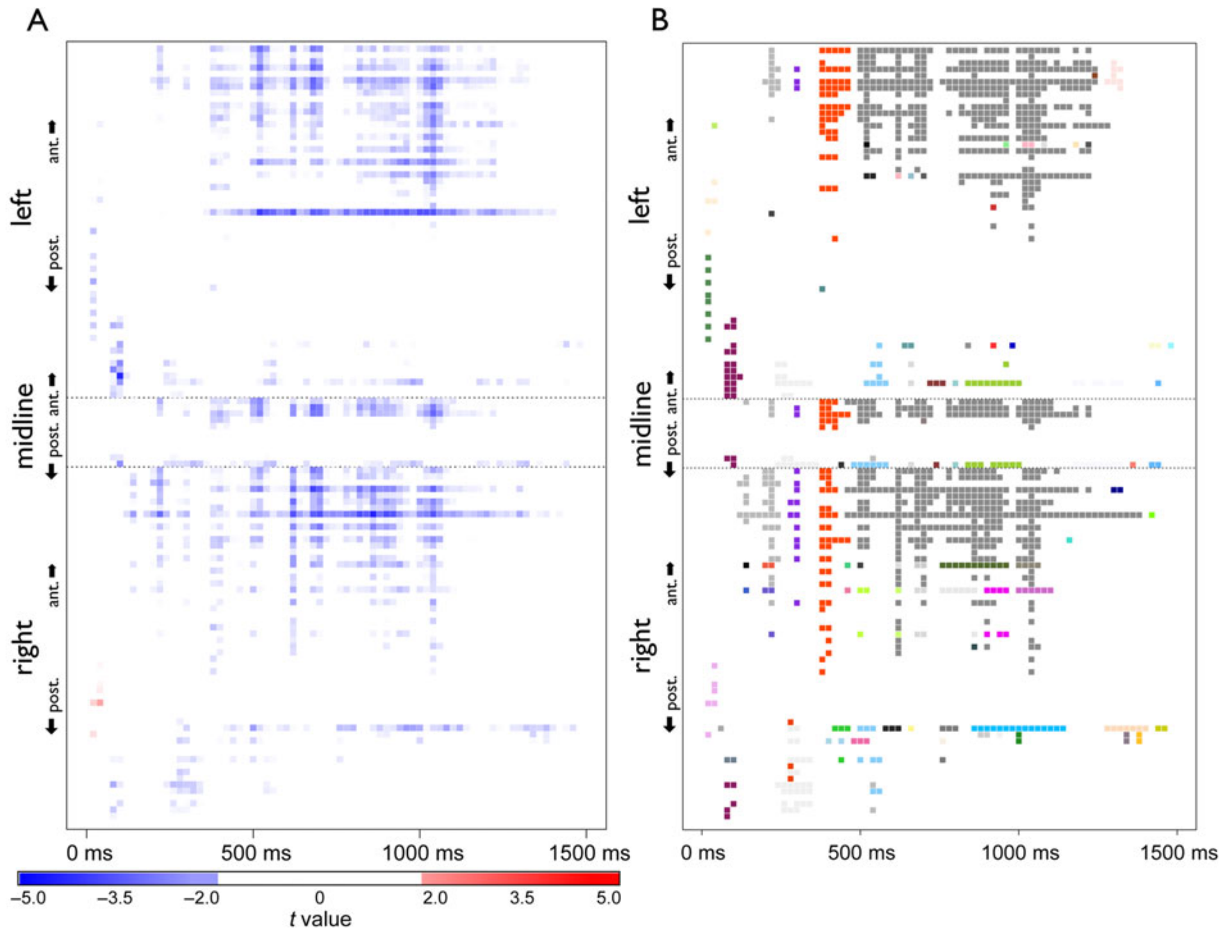

Figure 1. Permutation cluster analyses at the rest in Experiment 1. In each panel, electrodes are grouped into left hemisphere, midline and right hemisphere, with more anterior electrodes placed higher. Panel A: *t* values. Panel B: clusters (distinct colour for each cluster). The significant cluster is in grey.

at *the rest* indicate that that the context manipulation did affect scalar implicature calculation, thus one might have expected to see a neural signature of that calculation at *some* in the experimental declarative sentences (12a) as compared to the other sentences (where the implicature is not calculated). There are several possible interpretations of this null result, which we return to in the General Discussion. First, in Exp. 2, we confirm these findings through replication.

Experiment 2

In Experiment 1, we found evidence that entailment context affected the ultimate interpretation of *some* (based on the results at *the rest*). However, we found no signature of the differential processing in the ERP signature at *some* itself. Given this, and given recent concern about replicability in the cognitive sciences (Open Science Collaboration, 2012; Hartshorne & Schachner, 2012; Pashler & Wagenmakers, 2012), we conducted a replication.

Method

Thirty-two monolingual native English-speaking right-handed adults participated, 16 in each condition. Data for one participant in the control condition were corrupted and so excluded. Materials and ERP data analysis were identical to that of Experiment 1. EEG was recorded from 64 sintered Ag/AgCl electrodes embedded in an elastic cap (Neuroscan QuikCaps) following the extended 10–20 system (Nuwer et al., 1998). EEG data were sampled at 1000 Hz and later down-sampled to 200 Hz. Blink artefact was corrected through a subject-specific regression-based algorithm (Semlitsch, Anderer, Schuster, & Presslich, 1986). Reference location, filtering and ERP epoching were as in Experiment 1. We also 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.

Results and discussion

As in Experiment 1, permutation cluster analyses on the ERPs evoked by *the rest* detected a significant interaction

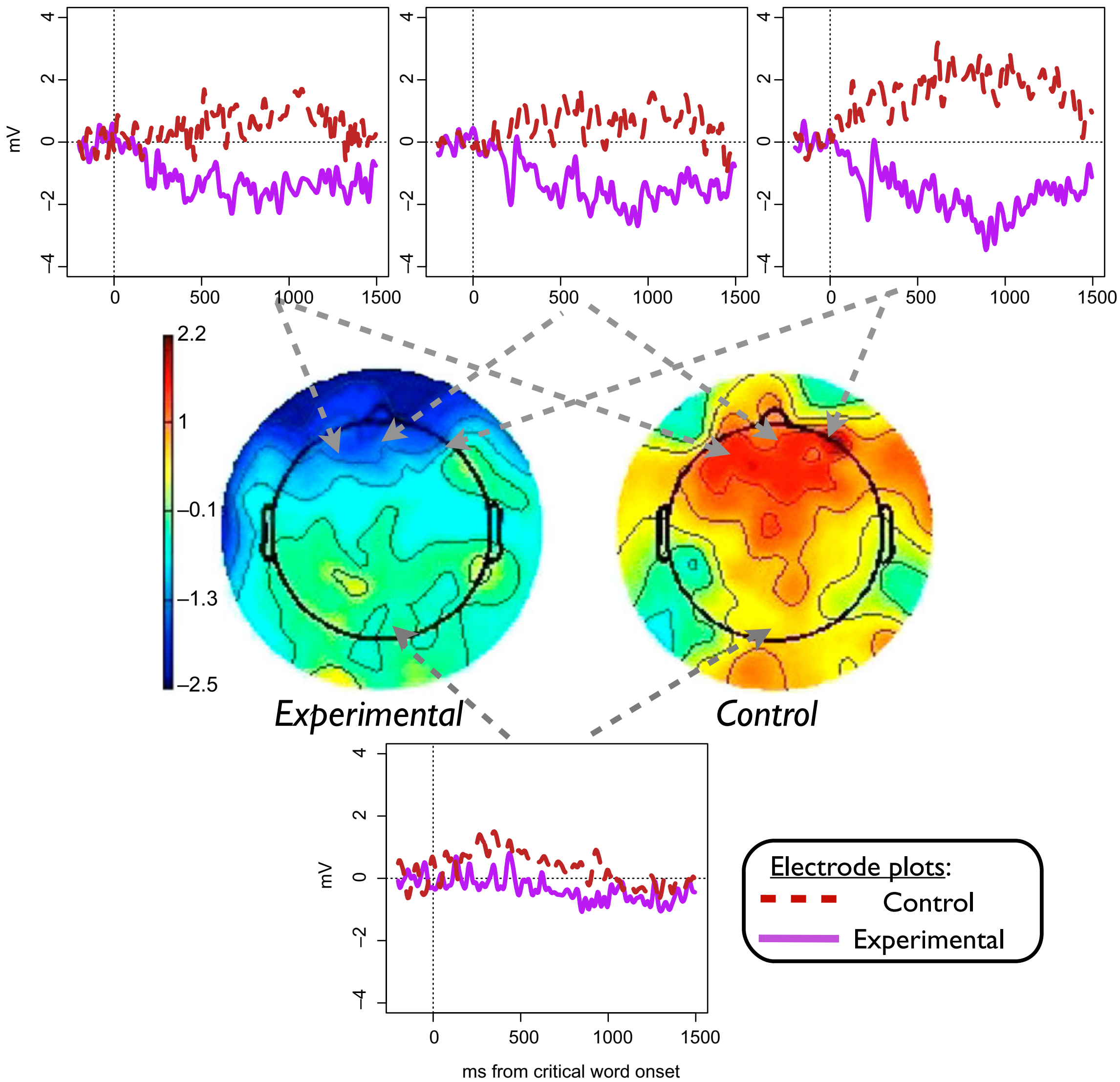

Figure 2. Difference waves (declarative–conditional) at the rest in Experiment 1. Topographical plots are shown at 600 ms post-stimulus. Four representative electrodes are depicted for the entire epoch. The relative negativity for the difference waves in the experimental sentences is driven by a positive deflection for the conditional sentences (see main text).

(Figures 4 & 5). The single significant cluster (p = .008) begins somewhat later than in Experiment 1, starting at around 850 ms post-stimulus (as in Experiment 1, it extends to the end of the analysed epoch). However, this cluster appears to be the extension of an effect beginning at around 600 ms (see Figure 4A). The distribution of the effect still appears skewed anterior but is more broadly distributed than in Experiment 1. As in Experiment 1, there were no significant effects at *some* (ps > .4; Figure 6).

Thus, the results of Experiment 1 and 2 are quite similar, despite the differences in recording equipment and subject population. In both cases, we observed a widely distributed, sustained positivity beginning around half a second after onset of *the rest*. In neither case do the effects resemble any of the three other frequently observed ERP effects in language processing studies: N400, Left Anterior Negativity or central-parietal P600 effects. In addition to resembling each other, both effect patterns bear some resemblance to recently reported anterior positivities, which have been elicited by plausible but unexpected words (e.g., Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007), as discussed below. No significant effects were seen after *some* for either experiment.

General discussion

A core issue in scalar implicature research is whether scalar implicatures are lexical inferences (e.g., *some* → *not all*) or are driven by broader informativity considerations. We found a robust difference in the ERPs triggered by *the rest* in declarative and conditional sentences, suggesting that

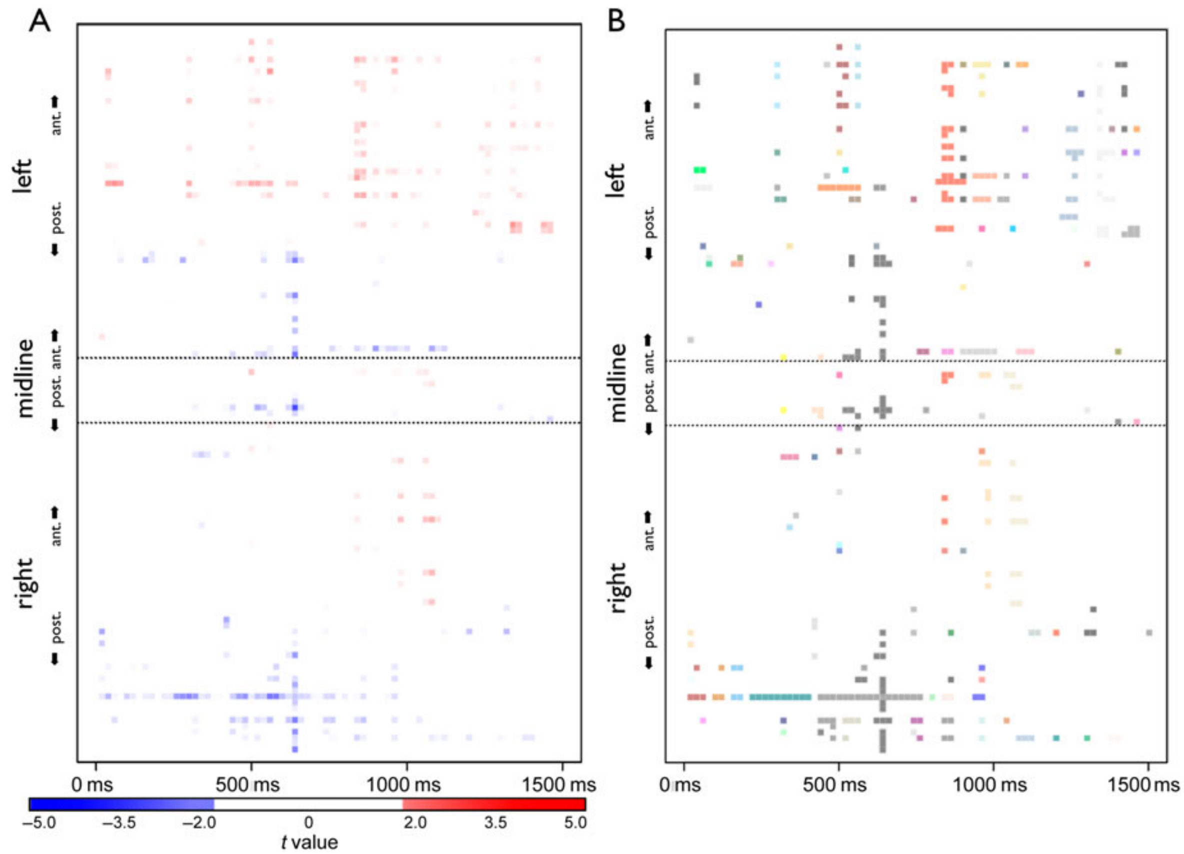

Figure 3. Permutation cluster analyses at some in Experiment 1. In each panel, electrodes are grouped into left hemisphere, midline and right hemisphere, with more anterior electrodes placed higher. Panel A: *t* values. Panel B: clusters (distinct colour for each cluster).

scalar implicature calculation is modulated by the informational content of the sentence as a whole, rather than by lexical scales alone. Crucially, we controlled for implicature-irrelevant differences by replacing *some* with *only some*, forcing the subset interpretation of the quantifier. These results are consistent with several previous explicit judgement studies that had queried people's intuitions about logical reasoning, similarly finding that participants were

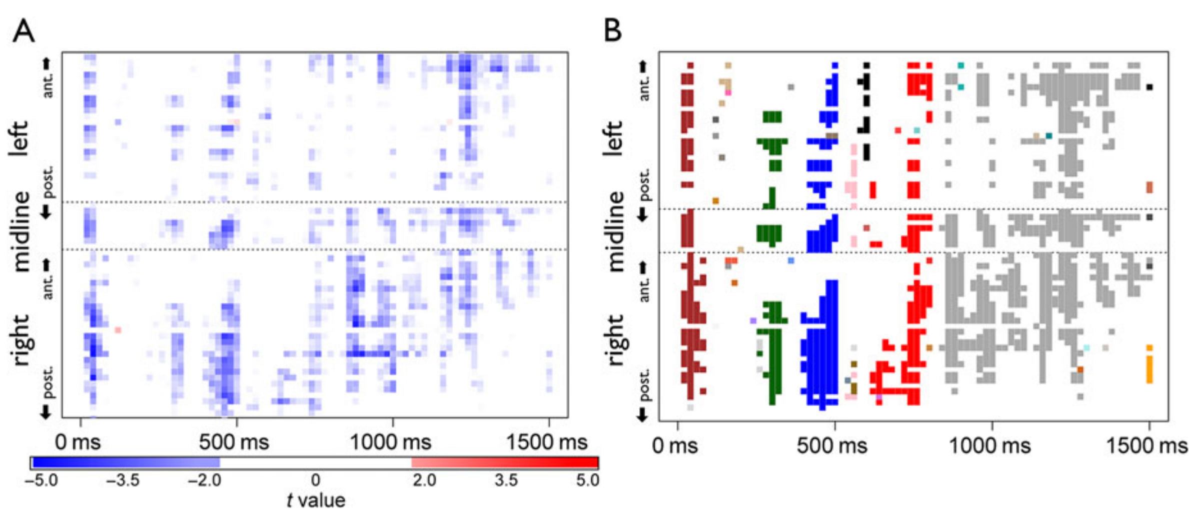

Figure 4. Permutation cluster analyses at the rest in Experiment 2. In each panel, electrodes are grouped into left hemisphere, midline and right hemisphere, with more anterior electrodes placed higher. Panel A: *t* values. Panel B: clusters (distinct colour for each cluster). The significant cluster is in grey.

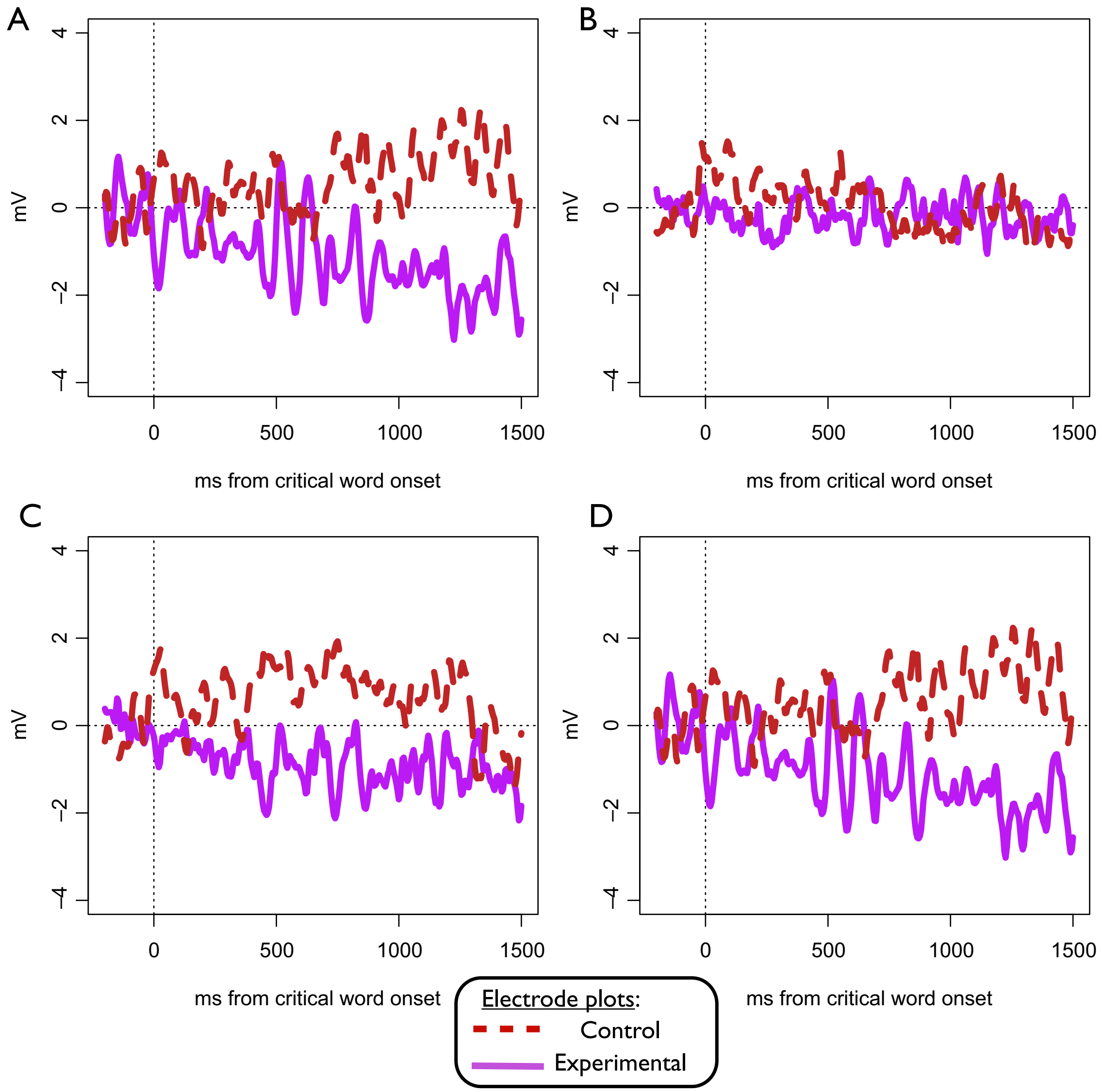

Figure 5. Difference waves (declarative-conditional) at the rest for individual electrodes in Exp. Panel A: F7. Panel B: F8. Panel C. Fz. Panel D. POz.

less likely to compute scalar implicatures within the antecedents of conditionals (see Introduction).

The present results demonstrate that scalar implicatures are calculated over propositional content rather than lexical items per se. Taken alone, however, this observation fails to explain why the terms on lexicalised scales seem to produce robust implicatures while other quantity implicatures require more support (see Introduction and Horn, 1989). As noted above, many theorists still make use of sets of lexical alternatives, positing that they are used to determine which sentences are considered as relevant alternatives. On some accounts, these sets of alternatives are lexicalised and included in the grammar (e.g., Chierchia et al., 2012), whereas on others they are a side effect of those terms frequently being used explicitly as alternatives to one another (e.g., Barner & Bachrach, 2010; Barner et al., 2010). How this would play out during online processing is not yet well understood. Interestingly, the studies to date do not find any clear indication that there is a qualitative difference between the online processing of scalar implicatures that are based on sets of lexical alternatives and those based on contextual scales (see Breheny, Ferguson, & Katsos, 2013; Zevakhina & Geurts, 2014).

There is an additional reason to doubt the viability of an account on which scalar implicatures are always calculated based on lexical alternatives, independent of proposition-level informativity. Considerable evidence has accumulated indicating that interpreting a scalar term pragmatically (e.g., *some-but-not-all*) takes longer and requires more cognitive resources than interpreting it semantically (e.g., *at least some*), which many researchers

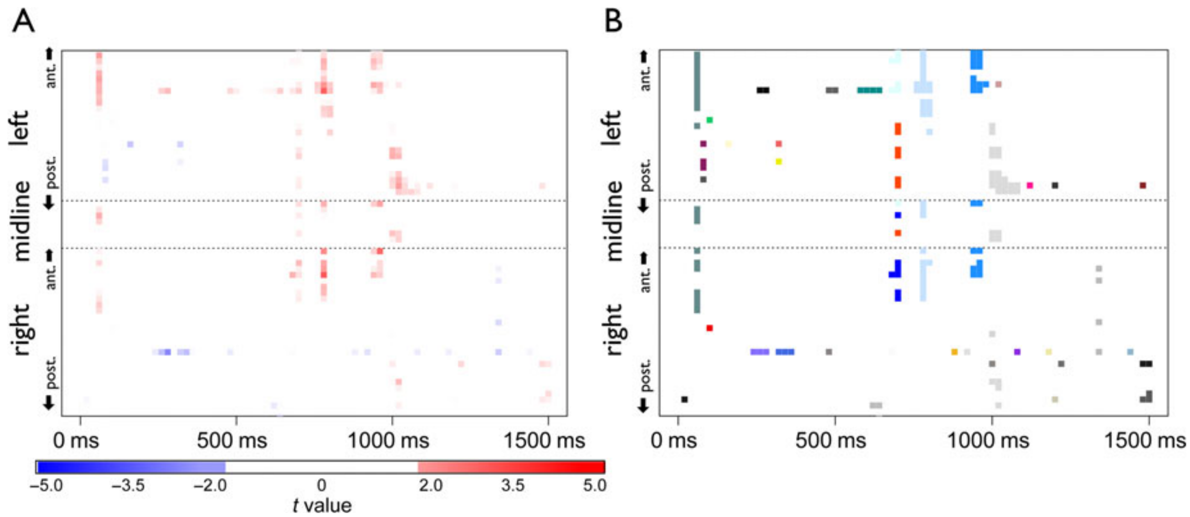

Figure 6. Permutation cluster analyses at some in Experiment 2. In each panel, electrodes are grouped into left hemisphere, midline and right hemisphere, with more anterior electrodes placed higher. Panel A: *t* values. Panel B: clusters (distinct colour for each cluster).

have argued is the opposite of what one would expect if implicatures are always calculated lexically and then cancelled as needed (Bott, Bailey, & Grodner, 2012; Bott & Noveck, 2004; De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Feeney et al., 2004; Huang & Snedeker, 2009a, 2011; Marty et al., 2013; Noveck & Posada, 2003; but see Grodner et al., 2010). Whether this is a fatal argument depends on exactly what the underlying processing model looks like and how the underlying processing gets transformed into the manifestations that we can measure (e.g., reaction time).

In the remainder of this section, we consider the implications of the ERP results in the context of our current understanding of ERPs and the neural bases of sentence comprehension.

Processing *the rest*

Two previous scalar implicature EEG studies observed negativities in relation to scalar implicature processing (Nieuwland et al., 2010; Politzer-Ahles et al., 2013). Though the distribution and onset of the negativities were different across the two studies, that could be due to the differences in methodology: Politzer-Ahles and colleagues (2013) involved sentence–picture matching, whereas Nieuwland and colleagues (2010) used reading-only paradigm. The authors of both studies interpreted their effects as reflecting pragmatic activity related to scalar implicature processing, in which case one might have expected a similar effect in our study. Instead, we observed a late, sustained positivity. Thus, across three studies, three different ERP effects have been observed.

One likely important difference between our study and the previous ones is that while the infelicitous sentences in the previous studies were infelicitous because the scalar implicature had been calculated (e.g., 14), our infelicitous sentences were infelicitous because the scalar implicature had not been calculated (12b, repeated below):

(14) Some people have lungs…
(12) If Addison ate some of the cookies before breakfast this morning, then the rest are on the counter.

It may be that the effects observed across the studies reflect integrating words that are made more or less felicitous/expected by the preceding context, rather than implicature processing per se. Nieuwland and colleagues (2010) observed an enhanced N400 on the critical infelicitous word (*lungs*), which typically indexes the overall degree of match between semantic features of an observed word and those that are predicted based on prior context (cf. Kutas & Federmeier, 2011). The negativity observed by Politzer-Ahles and colleagues may be related. The difference between our findings and theirs could reflect a difference in the degree of infelicity of the continuation in the critical condition. Nieuwland's infelicitous stimuli were quite infelicitous (14), and Politzer-Ahles's involved reading sentences that did not match a previously viewed picture. In contrast, our manipulation was more subtle: *The rest* was difficult to interpret in the conditional sentences in the experimental condition, but the sentences are nonetheless relatively felicitous.

To verify this difference in felicity, we asked 40 native English speakers (23 female, 15 male, 2 no response; 23–67 years old, $M = 42$, SD = 13) to rate the felicity (9-point Likert scale) of 40 randomly chosen stimuli from our study and 40 from Nieuwland's. Critically, we turned those sentences into sentence fragments by truncating the sentence either before or after the critical word (e.g. *lungs* or *the rest*; Table 3). We counterbalanced felicity and where the truncation occurred across the participants using a Latin Squares design. The three-way interaction of experiment (Nieuwland vs. present) × truncation (pre-/post-critical word) × felicity (felicitous vs. infelicitous) was significant (t = 4.4, p < .0001).[9] The Nieuwland stimuli showed a large difference in acceptability post-critical word (felicitous: M = 8.2, SE = 0.1; infelicitous: M = 6.3, SE = 0.3) that was not present when the sentence was truncated prior to the critical word (felicitous: M = 8.0, SE = 0.3; infelicitous: M = 8.1, SE = 0.3), resulting in a significant interaction (t = 11.4, p < .0001). For our stimuli, the analogous interaction was much smaller and not significant (pre-critical word, felicitous M = 7.9, SE = 0.2; infelicitous: M = 7.3, SE = 0.2; post-critical word, felicitous: M = 8.1, SE = 0.2; infelicitous: M = 7.7, SE = 0.2; t = 1.5, p = .13). Note that while the conditional sentences appear to be slightly less acceptable than the declarative sentences – perhaps because of their greater complexity – any such issues are controlled for in the main experiments by the use of the "only some" control conditions.

Thus, the semantic felicity differences in our stimuli – in contrast to those of Nieuwland and colleagues – may have been too small to affect the N400 component. What can account for the sustained positivity we observed? One possibility is that it reflects retroactive calculation of the scalar implicature. If so, one might have expected to see a similar effect at *some* in the declarative sentences, which was not the case. However, as discussed in the next section, many researchers have argued that scalar implicature processing is slow and probabilistic. If so, ERPs related to scalar implicature processing at *some* may be spread out over time; after averaging across trials, they would be difficult to detect. It may be that by forcing retroactive calculation of the scalar implicature at *the rest*, we concentrated the ERP, making it more detectable.

Another possibility is suggested by several studies that found sustained frontal positivities not unlike our own, which typically appeared in cases where the critical word is unexpected but plausible (DeLong, Urbach, Groppe, & Kutas, 2011; Federmeier et al., 2007; Moreno, Federmeier, & Kutas, 2002). For instance, Federmeier and colleagues (2007) reported a sustained positivity starting at about 500 ms, triggered where a plausible word (*look*) appeared in a context where another word (*play*) was considerably more likely (*The children went outside to*…).

Intuitively, *the rest* is plausible but unexpected in our conditional sentences in the experimental condition (12b), and thus a similar explanation might apply to our results. It should be noted, however, that the analogy is not perfect. We modified the judgement study above to ask participants how likely the critical word (*the rest*) was in our upward- and downward-entailing contexts; while it was judged to be significantly more likely in the upward-entailing context, in neither case was it judged highly likely (upward-entailing: M = 4.9, SE = 0.4; downward-entailing: M = 4.4, SE = 0.3; difference: t = 5.7, p < .001; judgements were made on a 7-point Likert scale). This could simply reflect the fact that all long sentences are unlikely, but if taken at face value, this contrasts with Federmeier and colleagues' (2007) finding that their late positivity was specific to strongly constraining semantic contexts and did not appear for more weakly constraining contexts (*Joy was too frightened to move/look*). However, the comparison is complicated by the fact that we manipulated context whereas they manipulated the critical word. Further work will be needed to better characterise these sustained, late positivities and determine if they are indeed a single component.

In summary, it may be that the ERPs to infelicitous critical words in our study, the Nieuwland study and the Politzer-Ahles study reflect general revision processes, rather than scalar implicature processes per se. While this is unfortunate news for those trying to uncover pragmatics-specific neural processes, it points in a profitable direction for uncovering the mechanisms by which pragmatic and semantic information is integrated online.

Table 3. Stimuli from the sensibility judgment study.

Study	Truncation	Felicity	Sentence
Nieuwland et al.	Before	Felicitous	Some people have …
Nieuwland et al.	Before	Infelicitous	Some people have …
Present Study	Before	Felicitous	Addison ate some of the cookies before breakfast this morning, and …
Present Study	Before	Infelicitous	If Addison ate some of the cookies before breakfast this morning, then …
Nieuwland et al.	After	Felicitous	Some people have pets, …
Nieuwland et al.	After	Infelicitous	Some people have lungs, …
Present Study	After	Felicitous	Addison ate some of the cookies before breakfast this morning, and the rest …
Present Study	After	Infelicitous	If Addison ate some of the cookies before breakfast this morning, then the rest …

Processing *some*

Perhaps the most intriguing finding was the absence of an effect at the scalar implicature trigger *some*. We found parallel results in five self-paced reading experiments involving similar stimuli (Hartshorne & Snedeker, 2014): *some* was read no slower (or faster) in contexts where the scalar implicature was calculated. Nonetheless, analyses at *the rest* – both here and in the self-paced reading experiments – show that our manipulation affected scalar implicature calculation.

The lack of an effect at *some* is made additionally surprising by a recent study by Bergen and Grodner (2012), in which they manipulated the speaker's knowledge state: In typical scalar implicature theory, listeners should not infer from the utterance of a weaker statement (*Addison ate some of the cookies*) that a stronger one (*Addison ate all of the cookies*) is false if the speaker does not herself know whether the stronger statement is true or false. Essentially, this lack of knowledge provides an explanation for the fact that the speaker did not use the stronger statement, and no implicature is necessary. Using such a manipulation, Bergen and Grodner (2012) reported that reading time was longer for *some* in conditions where participants infer implicatures (speaker full-knowledge condition) relative to conditions where participants do not (speaker partial-knowledge condition) (for converging offline results, see Goodman & Stuhlmüller, 2013).

One possible, if unlikely, explanation is that the scalar implicature is always triggered by *some*, and thus the ERPs were identical across conditions. On this account the implicature must then be cancelled in the downward entailing conditions (in order to be consistent with our results at *the rest*). We have already discussed some of the difficulties with this prediction. Additionally, when we used longer analysis windows or analysis windows beginning with words subsequent to *some*, we saw no evidence of an ERP effect that could reflect this cancellation process. Finally, this theory fails to explain other recent findings (e.g., Bergen & Grodner, 2012).

A second possibility, which we also deem unlikely, is that scalar implicature is so fast and effortless that its computation is not detectable using EEG or self-paced reading. Not only does that suggestion hard to reconcile with Bergen and Grodner (2012), who found slower reading of *some* when scalar implicatures were calculated, it would require re-consideration of a large body of literature suggesting that scalar implicature calculation is slow (Bott et al., 2012; Bott & Noveck, 2004; Feeney et al., 2004; Huang & Snedeker, 2009a, 2011; Noveck & Posada, 2003; but see Grodner et al., 2010) and requires considerable computational resources (De Neys & Schaeken, 2007; Dieussaert et al., 2011; Marty et al., 2013).

A third, more likely reason why our EEG (and self-paced reading) experiments may be insensitive to scalar implicature calculation is that scalar implicature calculation is so spread out in time that any effect largely washes out after averaging across trials. Numerous studies have shown that scalar implicature calculation is relatively slow, taking approximately one second from the onset of *some* (Bott et al., 2012; Bott & Noveck, 2004; Feeney et al., 2004; Huang & Snedeker, 2009a, 2011; Noveck & Posada, 2003; but see Grodner et al., 2010). If the cost of scalar implicature calculation is spread out across our analysis time window, that would diminish the effect size at any given point, rendering it difficult to detect without current analysis techniques and feasible sample sizes. One reason that Bergen and Grodner (2012) might nonetheless have detected this effect is that many of their implicature inducing contexts (speaker full-knowledge condition) – and none of their implicature-suppressing contexts (speaker partial-knowledge condition) – used the word *all*, potentially priming the scale and thus scalar implicature itself, speeding up the process and condensing it into a small enough time window to be detectable.

A final possibility is that while the *outcome* of scalar implicature processing in upward and downward entailing contexts is different, the *processing* is largely the same. On the Gricean account, scalar implicatures are calculated only when listener decides that some alternative would have been more informative. Presumably the listener must check the relative informativity of the alternatives before this decision can be made, and thus the overall processing may not be much different regardless of whether the ultimate interpretation does or does not include an implicature. Similarly, on Chierchia and colleagues' (2012) Grammatical Theory, the parser entertains every possible insertion site for the exhaustivity operator, though whether the operator is retained in that position depends on several factors including its resulting in a more informative interpretation of the utterance (the authors consider two different algorithms which make different predictions; see Chierchia et al., 2012, Section 4.6). Presumably, the only way the grammar can know that these criteria have been met is to actually carry out the operations. Thus, if EEG and self-paced reading are sensitive primarily to processing and not the outcome of that processing, it may be difficult to distinguish our two conditions. This same reasoning need not apply to Bergen and Grodner's (2012) study, because in their partial-knowledge condition, the speaker is assumed not to know whether *all* applies, and thus determining whether the same statement with *all* substituted for *some* is more informative is moot. As such, if the processor is reasonably efficient, implicature processing would only occur in their full-knowledge condition, resulting in an observable difference in processing cost across conditions.

Some headway might be made on determining exactly why Bergen and Grodner (2012) find effects of context at the word *some*, whereas we and Hartshorne and Snedeker

(2014) do not, by investigating online processing of other contextual effects. For instance, listeners reportedly do not make scalar implicatures if they believe the speaker would be unlikely to use the more informative statement, even if it were true (Bonnefon et al., 2009).

Conclusion

We find that grammatical entailment context modulates online scalar implicature processing. In particular, encountering *some of the X* sets up a possible antecedent for a later anaphor, but only in contexts that support scalar implicature. If the scalar implicature was not calculated, resolution of the anaphor is more difficult, resulting in a sustained positivity. This finding strongly challenges lexical-based accounts.

Acknowledgement

The authors thank Miki Uruwashi and for help with data acquisition and attendees at AMLaP 2012, CogSci 2013 for comments and suggestions.

Funding

This work was supported by NDSEG, NSF GRFP and NIH NRSA HD072748 fellowships to JKH, NICHD R03HD071094-01A1 to AK, and NSF-BCS 0921012 to JS.

Notes

1. This inference does not necessarily go through if the speaker does not know whether John ate every single one (maybe he was still eating when she left the room). While there is evidence that listeners are sensitive to the speaker's knowledge state when calculating implicatures (Bergen & Grodner, 2012; Bonnefon et al., 2009; Breheny et al., 2013; Goodman & Stuhlmüller, 2013), it is not clear whether implicature processing involves explicit mental state reasoning or simply approximates it. That is, like gazelle stotting or communication between bees, the fact that the problem solved by scalar implicature is inherently social does not mean that the underlying mechanisms explicitly invoke mental state reasoning.
2. If a specific alternative statement has been made contextually relevant, other implicatures may apply: (5) Alfred: Did John eat some of the cookies, and does he like scuba diving?
   Beatrice: John ate some of the cookies.
   Note that regardless of what Alfred said, Beatrice's statement implies that John did not eat all of the cookies. Additional inferences – e.g., about whether John likes scuba diving – may depend on what exactly Alfred asked.
3. Like the lexical alternatives account, the grammatical theory of implicature (Chierchia et al., 2012) operates over sub-propositional units and invokes lexical scales. However, on this proposal, the occurrence of a scalar implicature is sensitive to the grammatical context in which the scalar term appears.
4. See Note 3 above.
5. Note that these accounts differ along a number of other dimensions, such as whether scalar implicature involves grammatical processing. The last two decade has witnessed an explosion of work on scalar implicature within theoretical linguistics, resulting in a rich literature and detailed theories. Many of these distinctions are beyond the scope of the present work. We refer the interested reader to Chierchia et al. (2012), Sauerland (2012), and Geurts (2010).
6. Panizza, Chierchia, and Clifton (2009) report an eyetracking-while-reading study that manipulates entailment context. But critically this work focuses on the interpretation of number words ("two" means two and not three). Whether number interpretation involves scalar implicature is controversial (Breheny, 2008), and a variety of behavioural paradigms have found categorical differences in how numbers and scalar quantifiers are interpreted (Huang & Snedeker, 2009a, 2009b, 2011; Huang, Spelke, & Snedeker, 2013; Marty et al., 2013), suggesting that we cannot generalise from one to the other.
7. Note that this requires that the semantic relatedness to prior context be controlled (cf. Noveck & Posada, 2003; see also Kounios & Holcomb, 1992).
8. 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 it does not affect robustness to multiple comparisons. Other thresholds resulted in similar findings.
9. These and other analyses in this section utilised mixed effects models with subjects and items as random effects and with maximal random slopes design. *P* values are derived from model comparison.

References

Bach, K. (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: Elsevier.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*, 40–62. doi:10.1016/j.cogpsych.2009.06.002

Barner, D., Brooks, N., & Bale, A. (2010). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 87–96. doi:10.1016/j.cognition.2010.10.010

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

Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*, 249–258. doi:10.1016/j.cognition.2009.05.005

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

Bott, L., & Noveck, I. A. (2004). Some utterances are under-informative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*, 437–457. doi:10.1016/j.jml.2004.05.006

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, *25*, 93–139. doi:10.1093/jos/ffm016

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*, 443–467. doi:10.1080/01690965.2011.649040

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*, 423–440. doi:10.1016/j.cognition.2012.11.012

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

Chemla, E., & Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of Semantics*, *28*, 359–400. doi:10.1093/jos/ffq023

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. doi:10.1080/17470210701712960

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. A. (2010). Scalar inferences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*, 1104–1117. doi:10.1007/s10803-010-0960-8

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*, 535–590. doi:10.1023/B:LING.0000023378.71748.db

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University conference on language development* (pp. 157–168). Somerville, MA: Cascadilla Press.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In C. Maienborn, K. von Reusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning III* (pp. 2297–2331). Boston, MA: Walter de Gruyter.

Degen, J. (2013). *Alternatives in pragmatic reasoning*. (Thesis). University of Rochester, Rochester.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*, 1203–1207. doi:10.1111/j.1469-8986.2011.01199.x

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

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology 64*, 2352–2367.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. doi:10.1016/j.brainres.2006.06.101

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/Revue canadienne de psychologie expérimentale*, *58*, 121–132. doi:10.1037/h0085792

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*, 365–394. doi:10.1080/15475441.2011.626386

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York, NY: Academic Press.

Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind & Language*, *24*, 51–79. doi:10.1111/j.1468-0017.2008.01353.x

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

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

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

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

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some", and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55. doi:10.1016/j.cognition.2010.03.014

Hartshorne, J. K., & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience*, *6*. doi:10.3389/fncom.2012.00008

Hartshorne, J. K., & Snedeker, J. (2014). *The speed of inference: Evidence against rapid use of context in calculation of scalar implicature*. Manuscript submitted for publication.

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

Horn, L. R. (1972, March). *On the semantic properties of logical operators in English* (PhD Thesis). University of California-Los Angeles, Los Angeles.

Horn, L. R. (1989). *Natural history of negation*. Chicago, IL: University of Chicago Press.

Horn, L. R., & Wald, G. (2004). *The handbook of pragmatics*. Malden, MA: Blackwell.

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

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*, 1723–1739. doi:10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*, 1161–1172. doi:10.1080/01690965.2010.508641

Huang, Y., Spelke, E., & Snedeker, J. (2013). What exactly do number words mean? *Language Learning & Development*, *9*, 105–129.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: The case of scalar implicature. *Synthese*, *165*, 385–401. doi:10.1007/s11229-007-9187-4

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, *121*, 459–479. doi:10.1037/0096-3445.121.4.459

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi:10.1146/annurev.psych.093008.131123

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

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, *133*, 152–163. doi:10.1016/j.lingua.2013.03.006

Mathworks. (2012, January 25). *Signal processing toolbox: User's guide* (*2012a*). Retrieved from www.mathwords.com/help/signal

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*, 188–207. doi:10.1006/brln.2001.2588
Morris, C. (1938). Foundations of the theory of signs. *International Encyclopedia of Unified Science*, *1*.
Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324–346. doi:10.1016/j.jml.2010.06.005
Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165–188. doi:10.1016/S0010-0277(00)00114-1
Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning*, *8*, 297–326. doi:10.1080/13546780244000079
Noveck, I. A., & Posada, A. (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
Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*, 425–431. doi:10.1016/j.tics.2008.07.009
Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-frederiksen, A., Guérit, J.-M., Hinrichs, H., … Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology*, *106*, 259–261. doi:10.1016/S0013-4694(97)00106-5
Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657–660. doi:10.1177/1745691612462588
Panizza, D., Chierchia, G., & Clifton, C. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*, *61*, 503–518. doi:10.1016/j.jml.2009.07.005
Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language*, *33*, 721–757. doi:10.1017/S0305000906007550
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*, 253–282. doi:10.1016/S0010-0277(02)00179-8
Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*, 71–82. doi:10.1207/s15327817la1201_3
Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530. doi:10.1177/1745691612465253
Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research*, *1490*, 134–52. doi:10.1016/j.brainres.2012.10.042
Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*, 347–375. doi:10.1080/10489220701600457
Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics*, *23*, 361–382. doi:10.1093/jos/ffl008
Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367–391. doi:10.1023/B:LING.0000023378.71748.db
Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass*, *6*, 36–49. doi:10.1002/lnc3.321
Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*, 695–703. doi:10.1111/j.1469-8986.1986.tb00696.x
Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290.
Zevakhina, N., & Geurts, B. (2014). Scalar diversity. Manuscript submitted for publication.