Journal of Neurolinguistics 46 (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Processing scalar implicatures in conversational contexts: An ERP study

Thomas Holtgraves*, Brian Kraus

Ball State University, United States

ARTICLE INFO

Keywords:
Scalar implicatures
P300
Face-threat
Pragmatics
Comprehension
Conversation

ABSTRACT

Scalar expressions are words that have both a semantic meaning (e.g., the semantic meaning of "some" is "more than one") and a pragmatic meaning (e.g., the pragmatic meaning of "some" is "some but not all"). The majority of experimental research on scalar terms has focused on the quantity expression "some" and its use in nonconversational contexts. In contrast, in this research we examined five different scalar expressions that were embedded in a conversational context with varying degrees of face-threat. Participants read scenarios followed by a target utterance containing a scalar expression in the first half of the utterance (e.g., some), with a second half continuation of the utterance containing either the pragmatic meaning (e.g., not all) or the semantic meaning (e.g., all). ERPs in response to the scalar term and subsequent meaning were examined. Neural responses to the scalar term did not vary as a function of face-threat. However, the semantic meaning resulted in a larger P300 than did the pragmatic meaning, a difference that was greater when the situation was face-threatening than when it was not face-threatening. This pattern did not vary over the five different scalar expressions and suggests that in conversational contexts, it is the pragmatic meaning that is expected.

1. Introduction

Scalar expressions are words that can be ordered on a scale with respect to their strength (e.g., warm/hot, good/excellent, some/all) (Levinson, 1983) and have both a semantic meaning (e.g., the semantic meaning of "some" is "more than one") and a pragmatic meaning (e.g., the pragmatic meaning of "some" is "not all"). The pragmatic meaning of a scalar expression is referred to as a scalar implicature and by definition this meaning is not mandatory. This semantic/pragmatic distinction can be verified by considering whether the meaning can be cancelled; the pragmatic meaning can be cancelled (e.g., Some of your cookies were left, in fact, all of your cookies were left) but the semantic meaning cannot (e.g., Some of your cookies were left, in fact, none of your cookies were left).

Why do scalar implicatures occur? One possible reason is based on the Gricean notion of a cooperative speaker (Grice, 1968). If it is the case that the stronger term holds, it is assumed that the speaker would have used that term. Given that the stronger term was not used, the inference is that it does not hold. Hence, the use of "some" typically involves an inference that the speaker means some ***and not all***. In general, it is assumed that this principle can be extended to any set of terms which can be placed on an ordinal scale and which differ in their strength (Horn, 1976, 1989; Levinson, 2000). Empirical research on processing scalar expressions has been restricted primarily to investigations of quantity expressions (some) and in nonconversational contexts. The purpose of the present research was to examine the comprehension of several different types of scalar terms when those terms were embedded within conversational contexts.

* Corresponding author. Dept. of Psychological Science, Ball State University, Muncie, IN 47306, United States
E-mail address: 00t0holtgrav@bsu.edu (T. Holtgraves).

https://doi.org/10.1016/j.jneuroling.2017.12.008
Received 27 March 2017; Received in revised form 21 December 2017; Accepted 21 December 2017
0911-6044/ © 2017 Elsevier Ltd. All rights reserved.

1.1. Processing scalar implicatures

In general, there are two, over-arching approaches to the comprehension of scalar terms. One approach, attributed primarily to Levinson (2000), suggests that scalar implicatures (e.g., some but not all) are part of the meaning of the expression. That is, they are assumed to be lexicalized and hence represent the default interpretation. In this view, then, there is no processing cost associated with scalar implicatures unless they are overridden by the context. The alternative approach, the prototype of which is relevance theory (Sperber & Wilson, 1995), suggests that whether or not a scalar implicature occurs is entirely context dependent. A number of studies have been designed to investigate these two approaches.

One empirical approach has been to examine the processing of underinformative sentences containing the scalar expression "some". Bott and Noveck (2004) argued that the truth value of underinformative sentences (e.g., some elephants are mammals) depends on whether or not a scalar implicature is generated; they are false if the implicature is generated and true if not. They trained participants in two experiments to either generate the implicature or not; participants were slower and less accurate in the former relative to the latter. In a third experiment, participants were allowed to interpret the targets as they saw fit, and mean reaction times were longer when participants generated the implicature than when they did not generate the implicature. In a fourth experiment, time constraints were manipulated and when a short period of time was available, participants were more likely to respond "True" to the underinformative sentences, suggesting that they were less likely to derive the inference when they were under time pressure. These results, then, demonstrate that in these contexts there is a processing cost associated with generating the implicature, a finding that supports the contextualist view.

Using similar materials, Noveck and Posada (2003) examined electrophysiological responses to sentence-final words in underinformative sentences containing scalar terms (e.g., Some elephants have trunks), patently false sentences (e.g., Some crows have radios) and patently true sentences (e.g., Some houses have bricks). The patently true and false sentences elicited larger N400s than did the underinformative sentences, and this occurred regardless of whether participants judged the uninformative sentence to be true. However, judging the underinformative sentence to be false (indicating the generation of the implicature) took significantly longer than judging it to be true, a pattern consistent with the contextualist view that the generation of the implicature is an effortful process.

More recently, Nieuwland, Ditman, and Kuperberg (2010) conducted two experiments to examine the neural response to informative (e.g., Some people have pets) versus underinformative (e.g., Some people have lungs) sentences. If scalar implicatures are generated incrementally (i.e., when the scalar term is encountered), then an N400 should be larger for the critical word (e.g., lungs) in the underinformative sentences relative to the critical word in the informative sentences (e.g., pets). In the first experiment, there was not an overall difference between the informative and underinformative conditions. However, for more pragmatically skilled participants, the underinformative condition resulted in a significantly larger N400 than the informative condition. At least for the pragmatically skilled individuals, the pragmatic meaning was being used incrementally. In a second experiment, the sentences were made locally underinformative by deleting the comma and adding a relative clause. Hence, "Some people have lungs," was replaced with "Some people have lungs that are diseased by viruses". Critical words elicited a larger N400 in the informative condition than in the (locally) underinformative condition, and this difference was not moderated by individual differences in pragmatic ability. The authors interpret this finding as indicating that semantic processing costs are sometimes more influenced by lexical-semantic relationships rather than pragmatic informativeness. In these sentences, the under-informativeness was simply not noted due to the scope of the scalar quantifier. More recently, Bott, Bailey, and Grodner (2012) replicated these results but with a response-deadline procedure in order to investigate implicature speed and accuracy separately. Their results demonstrated that implicature generation for these items does entail a processing cost, and that it is a cost that is not attributable to syntactic complexity or retrieval probabilities.

Other researchers, using slightly different materials and procedures, have also reported scalar implicatures to be context dependent and to involve a processing cost (Breheny, Katsos, & Williams, 2006), although Politzer-Ahles and Fiorentino (2013) reported context effects (i.e., implicatures were more likely when supported by the context) but no evidence for a processing cost when comprehending the scalar term itself (i.e., reaction times for the scalar term did not vary as a function of the context). Also, there have also been demonstrations of various constraints on the processing of scalar implicatures. For example, Degen and Tanenhaus (2011) demonstrated that syntactic features (e.g., Some vs. Some of the) can influence interpretation of these terms (but not response times). Similarly, Hartshorne et al. (2015) examined the neural responses to critical words in sentences containing the scalar term "some" that were either declarative (e.g., Addison ate some of the cookies before breakfast, and the rest are on the counter) or conditional (e.g., If Addison ate some of the cookies before breakfast, then the rest are on the counter). There was a more extreme response (an enhanced late positivity) to the critical phrase (e.g., the rest) in the conditional version than the declarative condition. This suggests that the scalar implicature was not generated initially in the conditional version (thereby demonstrating the importance of the sentence context for the generation of scalar implicatures) and required some type of revision process (as indicated by the late positivity) when the critical phrase was encountered. Interestingly, and consistent with Politzer-Ahles and Fiorentino (2013), there were no differences in the neural response to the scalar term itself (although Politzer-Ahles and Fiorentino (2013) report evidence for earlier inferential activity than Hartshorne et al., 2015). However, in research where individuals read a sentence immediately before viewing a picture which was either semantically or pragmatically consistent with the previous sentence, a sustained posterior negativity was observed indicating that processing pictures within a context may elicit different neural activity versus sentence processing (Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013). Also, Bergen and Grodner (2012) demonstrated that implicature generation is sensitive to assumptions regarding the speaker's knowledge state (i.e., whether the speaker is likely to know if the stronger statement would be true).

More recent research has focused more specifically on the sequential nature of the comprehension of scalar expressions. For example, Tomlinson, Bailey, and Bott (2013) conceptualized the processing issue in terms of one-step vs. two-step models, with the latter involving multiple, sequential, processing stages (i.e., one meaning is first activated and then subsequently replaced by a different meaning). In a series of experiments using a mouse tracking procedure, these researchers demonstrated that scalar implicatures for "some" are generated via a two-step process, with the initial activated interpretation (some and possibly all) being subsequently replaced by the implicature (some but not all). Politzer-Ahles and Gwilliams (2015) used magnetoencephalography to examine the neural correlates of scalar inferences. Participants listened to descriptions of contexts that either strongly or weakly supported the implicature for "some". Increased activation in the middle portion of the lateral prefrontal cortex was observed when the context was minimally supportive of the implicature, relative to the strongly supportive context. These results are consistent with a gradient account (rather than all or nothing) in which contextual cues moderate the ease and effort required for the implicature.

1.2. Politeness and scalar implicatures

An alternative approach to scalar expressions, one that emphasizes their interpersonal properties, has been articulated by Bonnefon and colleagues (Bonnefon, Feeney, & Villejoubert, 2009; Bonnefon, De Neys, & Feeney, 2011). They argue that scalar terms can serve a politeness function, and that this function influences how they are processed. In several studies, they had participants read a scenario and utterance containing a scalar term (e.g., some), and then provide a judgment regarding the possibility that the semantic (i.e., unenriched) interpretation was true. When the information was face-threatening (e.g., Some people hated your x), participants were more likely to endorse the possibility of the semantic interpretation than when it was face-boosting (e.g., Some people loved your x). Bonnefon and colleagues argue that this is because in the face-threatening (but not face-boosting) context, participants recognize that "some" may be serving a politeness function and that this serves to block the pragmatic (i.e., enriched) interpretation. Moreover, in one study in which reaction times were collected (Bonnefon et al., 2011), true judgments of the semantic interpretation took significantly longer than false judgments when the information was face-threatening, indicating that in these contexts it is the semantic interpretation rather than the pragmatic interpretation that requires effort and is a time-consuming process. Overall, then, these studies suggest that face-threatening contexts may block the scalar implicature due to the hearer's recognition of a possible politeness motivation, and that the endorsement of the semantic interpretation involves an effortful process.

Recently, Mazzarella (2015) has raised questions regarding the precise role played by face-threat in the comprehension of scalar terms. She argues that the empirical demonstrations of Bonnefon and colleagues do not distinguish between comprehension and acceptance. Instead, she argues that when the scalar expression is first encountered there may be some activation of the pragmatic interpretation (i.e., the implicature), and that it is only later, when queried about the speaker's meaning, that reasoning about a possible politeness motive has its effect and results in endorsement of the semantic interpretation. This is certainly a plausible account, yet to our knowledge there have been no attempts to address this issue empirically.

1.3. Present research

Our goal was to extend research on the comprehension of scalar terms in several ways. First, most empirical research on scalar terms has focused on quantity expressions, primarily the quantifier "some". In contrast, we examined the comprehension of five different scalar expressions: quantity (some), frequency (sometimes), preference (like), evaluation (good) and probability (possible). These expressions (with the exception of the probability term) were all included in Levinson's (1983) treatment of scalar terms. Based on his work, our working assumption was that they will all operate in the same way (i.e., the use of the less extreme term implicates that the more extreme does not hold). For example, "sometimes" pragmatically implies "not always", "like" pragmatically implies "don't love", and so on.

Second, we were interested in examining the processing of these terms when they were used in conversational contexts in which interpersonal considerations (i.e., politeness) might play a role. To do this we combined an electrophysiological procedure with a self-paced reading task to examine the neural responses to both the scalar term itself, and subsequent (either pragmatic or semantic) interpretations of that term occurring in the second half of the utterance. Although a self-paced reading task requires a motor response which may create EEG artifacts, Dittman, Holcomb, and Kuperberg (2007) have demonstrated that no discernible ERP artifacts are present when a self-paced reading procedure is used. Moreover, because this procedure is consistent over conditions, any artifacts associated with a self-paced reading task will occur for all conditions. Finally, because this procedure involves participants responding to the stimuli (rather than passively reading or listening) our primary interest was in variability in the amplitude of the P300 (although other neural responses were examined as well).

Participants read scenarios describing different interpersonal situations in which one interactant replies to a question from another interactant. The first half of this reply contained a scalar expression (e.g., some) and the second half contained a phrase that constituted either the pragmatic meaning of the term (e.g., not all) or the semantic meaning of the term (e.g., all). An example follows:

John couldn't make it to Susan's party. To make up for it, he made her some cookies and brought them over for the party. After the party was over, John asked Susan if any of his cookies were left over. Susan says:

*There were <u>some</u> left over, specifically, they were **not all** left over. (Pragmatic)*

*There were <u>some</u> left over, specifically, they were **all** left over. (Semantic)*

Our interest was in the neural response to the scalar terms, and to the words constituting the pragmatic (e.g., not all) versus semantic (e.g., all) interpretations. Hence, we examined neural responses time-locked to the scalar term and neural responses time-locked to the subsequent interpretation. For the scalar term, our primary interest was in the neural responses to the term as a function of face-threat. Following Holtgraves (2014), we manipulated face-threat by varying the referent of the target utterance; if it applied to one of the interactants it was assumed to be more face-threatening than if it pertained to a non-present person. If variability in face-threat influences the comprehension of the term when it is encountered, then we would expect to see variability in the neural response as a function of face-threat. On the other hand, if fixing the meaning of the scalar term is delayed, then no variability as a function of face-threat should occur. Prior research is not clear in this regard because no one has yet examined the impact of face-threat on the comprehension of a scalar term, and research examining the effects of other context variables on processing the scalar term itself have been mixed. Hence, we did not have a directional hypothesis for neural response to the scalar term.

Regardless of whether the interpretation is immediate or delayed, any difference in the neural response to the subsequent pragmatic vs. semantic interpretation should be informative regarding participants' expectations about the meaning of the scalar term in these conversational contexts. One of the prominent theories about the function of the P300 is the context updating hypothesis, which states that the P300 amplitude is a measure of context updating of a mental representation (Donchin & Coles, 1988). Thus, the added contextual information of the conversation may elicit a P300 in response to a perceived mismatch between the previous context of the conversation and its final interpretation. Prior research examining scalar terms in nonconversational contexts has generally supported initial activation of the semantic interpretation, due to the contextual sensitivity and time-consuming process for the pragmatic interpretation. However, no one has yet examined neural responses to scalar terms in interpersonal conversational contexts. In addition, research conducted by Shetreet, Chierchia, and Gaab (2014) suggests that the effortful process associated with the pragmatic meaning of underinformative sentences is at least partially due to a mismatch between the utterance and the context. That is, because people know that all people have lungs, the sentence “Some people have lungs” is anomalous and hence relatively difficult to comprehend. Hence, we did not have an a priori directional hypothesis for the interpretation.

Finally, if there are no differences in neural responses to the scalar term as a function of face-threat, then, as suggested by Mazzarella (2015), face-threat may play a later role in the processing of a scalar expression, specifically, when participants encounter the (pragmatic or semantic) interpretation. When the context is more face-threatening, expectations of politeness are increased and thus context updating may need to occur to evaluate new information versus this existing expectation. Because the semantic interpretation violates that expectation, we expected an enhanced P300 to the semantic interpretation relative to the pragmatic interpretation when the context was face-threatening.

2. Method

2.1. Participants

Forty participants were recruited to participate in this study. Of these, 25 were recruited using the Ball State University Psychological Science subject pool and were compensated with course credit for participating. The other 15 subjects were recruited via a campus wide e-mail and paid $20 for their participation.

2.2. Experimental materials

Stimulus materials consisted of a set of scenarios and corresponding utterances containing a scalar term. Examples are presented in Table 1; all materials are presented in the Appendix. There were 12 scenarios for each of the following five scalar terms: quantity (some), frequency (sometimes), preference (like), evaluation (good), and probability (possible). Each critical utterance contained one

Table 1
Sample experimental materials.

Frequency: John and Jake (and Kevin) decided to move into an apartment together off campus. After living together for about a month, *Jake* (Kevin) asked John if he thought *he* (Jake) was lazy. John responds:
You are (Jake is)/sometimes lazy,/specifically,/*you're* (he's)/(not) always lazy.
Quantity: Nina was a graduate student and taught her first class. After the class was over, *Nina* (one of the students) asked one of the students named James how ma[illegible]dents were using their phone during class. James replied:
James: [illegible] were/some students using their phone,/specifically,/they were/(not) all using their phone.
Probability: Mrs. Taylor was cleaning the room of her two sons, Jon and Joe. While doing so, she discovered a pack of cigarettes. While Joe was at soccer practice, she asked Jon whether the cigarettes were *his (Joe's)*. Jon replied:
It's possible/that the cigarettes are *mine* (Joe's),/specifically,/it's (not) certain/that the cigarettes are mine (Joe's).
Preference: Ben spent a good part of the afternoon making his favorite meal for his wife, Julie. After the meal, *he* (Julie's friend, Libby), asked if she liked the meal. [illegible] said:
Juli[illegible]ked the meal,/specifically,/I/(didn't) love(d) it.
Ev[illegible]on: Grace made her roommate, Lucy, a pumpkin cheesecake for her birthday. Later on, *she* (their neighbor Dan) asked Lucy what she thought of it.
Lu[illegible]: It was/good,/specifically,/it was/(not) excellent.

Note: The non-face-threatening version was created by substituting the material in parentheses for the italicized material. For the pragmatic interpretation, the negation in parentheses (not or didn't) was presented. The words between each / denote what text appeared on each of the 5 slides presented for each slide of each utterance (see Fig. 1). The scalar expression and subsequent semantic/pragmatic interpretation whose onset was time locked for ERP analyses are underlined.

Fig. 1. Trial procedure for all experimental trials. Utterances during which the scalar terms and interpretation terms occurred in each trial were used for analysis.

of the scalar terms in the first half of the reply, with the second half of the reply containing either a semantic or pragmatic interpretation. For example, a sample target utterance for the preference term was: "I like it, specifically, I love it" (semantic), or "I like it, specifically, I don't love it" (pragmatic).

There were two versions of each scenario, a face-threatening and non-face-threatening version. In the face-threatening version the utterance pertained to the recipient of the remark; in the non-face-threatening version the utterance pertained to someone who was not present. This manipulation was identical to that used in Holtgraves (2014), and all of the items used in that experiment (and hence pretested for the effectiveness of the face-threat manipulation) were used in the current experiment. Additional items using the same face-threat manipulation were created for this experiment in order to increase the number of trials necessary for an ERP study (see Table 1 for examples).

Four versions of each scenario/utterance were created by crossing face-threat (threat vs. nonthreat) with interpretation type (pragmatic vs. semantic). Four different stimulus sets were then compiled so that each participant saw an equal number of face-threat by interpretation type versions of each of the five scalar types, and across the experiment an equal number of participants saw each version of each scenario; each participant saw a total of 60 critical trials. There were also 52 filler trials that were identical to the critical trials except that the second half of the critical phrase did not contain a pragmatic or semantic interpretation. Instead, it was a plausible completion (e.g., I liked it, specifically, I didn't hate it) or logical violation of the scalar term (e.g., There were some left, specifically, there were none left). In order to keep participants focused and attending closely to the experimental materials, they responded to a question following each critical utterance. To ensure that participants attended to both the scenario and the utterance, these questions dealt with material from either the scenario or in the utterance; half the time the correct response was "Yes" and half the time the correct answer was "No".

2.3. Procedure

After giving informed consent, participants were seated approximately 54 cm in front of a 24" Dell LCD monitor with the refresh rate set to 60 Hz and the resolution held at 800 × 600 pixels. Participants then both read and verbally received detailed instructions regarding the use of scalar implicatures in the task, subsequently performed six practice trials, and received feedback from the experimenter regarding their performance. E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2002) was used for stimulus presentation. Participants responded by using a response box to advance the screen at their own pace.

Each trial proceeded as follows (see Fig. 1). First, a short scenario describing a conversation between two or more people was presented on the screen. Participants indicated comprehension of the scenario by pushing the center key on the response box. In contrast to prior psychophysiological research on scalar expressions, then, our procedure involved a response component. Prior research has demonstrated the feasibility of this approach, with prior research showing that self-paced reading paradigms do not introduce any discernible artifacts into the EEG recordings versus RSVP reading tasks (Dittman et al., 2007). An utterance said by one of the scenario interactants was then displayed in groups of between one and ten words (see Table 1). Participants used the center key on the response box to advance through the utterance. A 500 ms ISI occurred after the participant advanced through each part of the sentence. After a participant indicated via the response box that they understood the fifth utterance after each scenario, a statement referencing either the scenario (e.g., Jack asked Mark what he thought of his new car) or the utterance (Mark said he loved Jack's new car) was presented. Participants were instructed to indicate whether or not this sentence was accurate or inaccurate by using the left and right keys on the response box respectively.

2.4. EEG data analysis

EEG recordings were taken from 64 locations (see Fig. 2) on the scalp arranged as a dense array 10–20 electrode system using a BioSemi ActiveTwo EEG amplifier system with two 32 electrode strands mounted in an electrocap using BioSemi recording software

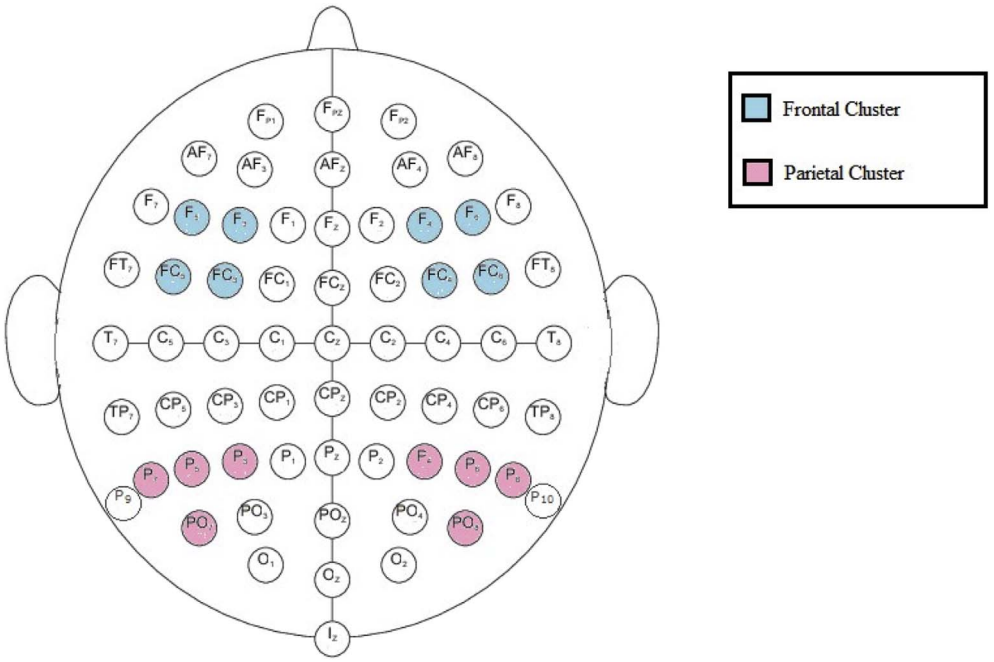

Fig. 2. Electrode clusters used for analysis. The frontal clusters were used for analyses of the scalar term and the parietal clusters for the analyses of the pragmatic/semantic meaning terms.

(BioSemi, Amsterdam, The Netherlands). Two EoG electrodes were placed adjacent to the right eye to capture vertical and horizontal eye movements. Electrodes were placed on each mastoid and were averaged offline as a reference during data processing. Data was recorded continuously throughout the entire task. Recordings varied in length from 30 to 67 min as the task was self-paced. ERPs were recorded from the onset of both the utterance slide in each sentence that contained the scalar term and the onset of the utterance slide that contained the semantic/pragmatic interpretation term.

The data was analyzed using the EEGLAB plugin for MATLAB (Delorme & Makeig, 2004). A high pass FIR filter was applied at 0.5 Hz to reduce spatial drift and help satisfy the assumption of spatially stationary sources for Independent Component Analysis (ICA) (Onton, Westerfield, Townsend, & Makeig, 2006). A low pass FIR filter was also applied at 55 Hz. The data were visually inspected for artifacts and all muscle artifacts and external interference were visually rejected. Bad channels were then removed from the file for interpolation after the data had undergone ICA artifact pruning. No more than 6 electrodes ($< 10\%$ of the 64 scalp electrodes) were interpolated for any one subject. No more than two adjacent electrodes were removed for interpolation or that subject was excluded from further analysis. Spherical interpolations were performed using the *interp* function in EEGLAB. By making sure that adequate usable data channels are present in proximity to a bad channel, its activity can be adequately estimated with a high density electrode system (Greischar et al., 2004). The data was then processed into epochs scaling -100 ms to 900 ms in relation to the onset of each section of the utterance and each response.

ICA was used to remove any remaining artifacts in the data. Decomposition of the independent components was performed in EEGLAB (version 13.2.2) using the 'runica' INFOMAX algorithm (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). Each subject in this analysis had $>$ 500 1s long epochs of data from which ICA components were determined for up to 64 scalp electrodes and 2 EoG electrodes. If any scalp electrodes were deemed unsuitable for analysis, they were removed for interpolation before performing ICA. The initial learning rate was 0.001 and the ICA converged when weight change was smaller than $1E^{-7}$. The components were visually inspected and artefactual components were rejected. Rejected components were primarily related to eye movements captured by the EoG electrodes and muscle artifacts (McMenamin et al., 2010).

Each epoch was then baseline corrected and any previously removed channels were interpolated using spherical interpolation. The epochs then underwent automatic artifact rejection and epochs with any channel that exceeded $+/-$ 75 absolute threshold at any point during the trial were excluded from further analysis. All of the participants were then compiled into a grand average and waveforms of different conditions were visually inspected. Electrodes for analysis were chosen based on visual inspection of the grand average. The mean ERP amplitude (μV) between two latencies was baseline corrected and then extracted for each participant for each condition. Parametric statistical tests were then run on the extracted means for each participant.

3. Results

Each participant completed 60 critical trials. Trials were rejected from further analysis during data processing based on the visual rejection of artifacts in the continuous data. Participants were excluded from the final analysis if their processed data did not contain

more than 70% of both the scalar and implicature utterances. Of the 40 original subjects, 26 met this exclusionary criterion and only their data was used for further analysis. The median number of scalar term trials analyzed for the remaining subjects was 54 out of 60 with a range of 44–60. The median number of interpretation term trials analyzed for each subject was 55 with a range of 47–60.

3.1. Behavioral analyses

Participants' responses to the question following the utterance during each trial were assessed for accuracy. On average, participants selected the correct answer 88% of the time for the critical trials with a range of 65–98%. Hence, these results suggest that participants were attending closely to the experimental materials.

In addition to accuracy, reaction time (RT) data (reported in ms) were also examined. The overall mean reaction time was 582.79 (SD = 337.14). For the scalar term the mean was 597.01 (SD = 460.37) and for the interpretation the mean was 764.07, SD = 678.03.

Scalar term and interpretation reaction times were each analyzed with a 5 × 2 × 2 Scalar type (quantity, frequency, preference, evaluation, probability) X Threat (threat, no threat) X Hemisphere (left, right) ANOVA. For the scalar term there was a significant main effect of Type, $F(1,25) = 63.898$, ($p < .001$), an effect that is understandable given the variability in the size of the phrases for the different types of scalar expressions. No other effects were significant ($p > .05$) for the scalar term. For the interpretation phrase, there was again a significant main effect of Type, $F(1,25) = 63.176$, ($p < .001$). In addition, there were significant main effects for threat, $F(1,25) = 5.987$, ($p < .022$) with threatening trials ($M = 791.35$, $SD = 716.98$) having significantly longer RTs than non-threatening trials ($M = 736.79$, $SD = 635.99$), and interpretation, $F(1,25) = 56.367$, ($p < .001$), with pragmatic interpretations ($M = 885.08$, $SD = 742.74$) having significantly longer RTs than semantic interpretations ($M = 643.06$, $SD = 582.28$). No other significant effects were present ($Fs < 1.1$). The same results occurred when the analyses were restricted to trials for which participants correctly answered the comprehension check item.

3.2. ERP analyses

Our primary interest was the neural response to the scalar terms and two subsequent interpretations (pragmatic or semantic) as a function of face-threat. Additionally, we were interested in whether the response would be similar for all scalar types, and the possible role of the RH in comprehending scalar terms. We first present our initial analyses of the neural responses to the scalar term, followed by responses to the interpretation. Due to the large number of variables we examined, and the necessity of a large number of statistical tests in order to evaluate those variables, our analyses proceeded in stages. We first analyzed a specific neural response in which we included scalar type as an independent variable. If scalar type did not interact with any other independent variables, we collapsed over scalar type in subsequent analyses.

3.3. Analyses of the scalar term

Visual examination of responses to the scalar term suggested the occurrence of a P200 in the bilateral fronto-cental areas (F3, F4, F5, F6, FC3, FC4, FC5, FC6)[1]. These electrodes were clustered within their respective hemispheres for further analysis (see Fig. 2). We first conducted a 5 × 2 X 2 Scalar type (quantity, frequency, preference, evaluation, probability) X Threat (threat, no threat) X Hemisphere (left, right) repeated measures ANOVA for a P200 (150–250 ms) in these areas in response to the presentation of the scalar term. The face-threat manipulation was not significant, $F(1,25) < 1$; hence, neural responses to the scalar term did not vary as a function of face-threat. However, there was a significant Threat × Hemisphere interaction, $F(1,25) = 8.535$, ($p = .007$) (See Fig. 3). Post-hoc tests indicated that when a trial was threatening, the P200 was significantly larger in the left hemisphere ($p < .05$) than the right hemisphere, a difference that did not occur when the situation was not threatening. The effect of face-threat was not significant ($p > .10$) in either hemisphere.

There was a significant main effect for scalar type (See Fig. 4), $F(4,100) = 4.348$, $p = .003$), and post-hoc tests indicated that the P200 for preference (like) was larger than all other scalar types except for evaluation (good) ($p < .05$).[2]

3.4. Analyses of the interpretation

Visual examination of responses to the interpretations suggested differences in a P300. Accordingly, we first conducted a 5 × 2 X 2 Scalar type (quantity, frequency, preference, –evaluation, probability) X Interpretation (semantic, pragmatic) X Hemisphere (left,

[1] Visual examination of responses to the scalar term also suggested differences in a P300 in the same bilateral fronto-central areas (see Fig. 1) and the results generally paralleled the pattern observed for the P200. The face-threat manipulation was not significant, $F(1,25) < 1$ and there was a significant main effect for scalar type, $F(4,100) = 5.328$, $p = .001$. In addition, there was a significant main effect for hemisphere, $F(1,25) = 6.031$, $p = .021$, and a significant Scalar Type × Hemisphere interaction, $F(4,100) = 6.093$, $p < .001$.

[2] There was also a significant Scalar Type × Hemisphere interaction, $F(4,100) = 7.463$, $p < .001$. Post-hoc tests indicated that in the left hemisphere, the P200 for probability (possible) was smaller than all other types except for quantity (some) ($p < .05$). However, in the right hemisphere the P200 for preference (like) was significantly larger than all other types except for evaluation (good) and probability (possible) while the P200 for quantity (some) was significantly smaller than all other types except for frequency (sometimes) ($p < .05$). Additionally, the P200s for both quantity and frequency were significantly larger in the left hemisphere than the right hemisphere ($p < .05$).

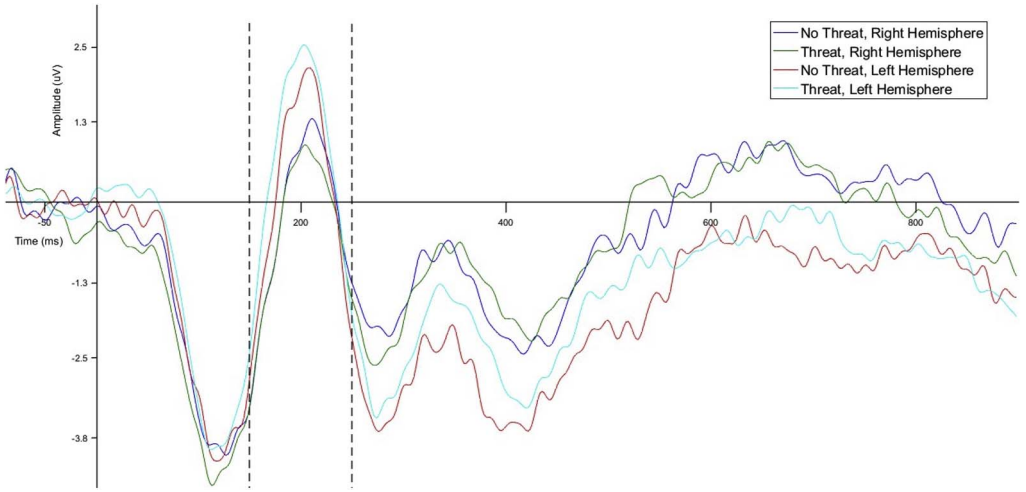

Fig. 3. P200 for Scalar Term as a Function of Threat and Hemisphere. The left and right hemisphere electrodes plotted are averages of (F5, F3, FC5, FC3) and (F6, F4, FC6, FC4) respectively.

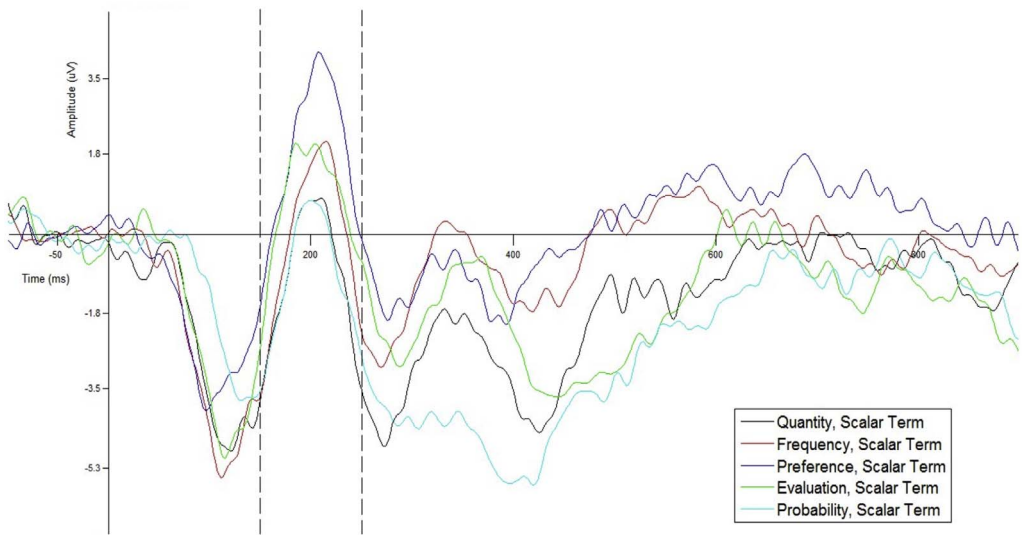

Fig. 4. Bilateral P200 for Scalar Term as a Function of Scalar Type at electrodes (F5, F3, FC5, FC3, F6, F4, FC6, FC4).

right) repeated measures ANOVA for a P300 (300–500ms) in the bilateral parietal areas (P3, P4, P5, P6, P7, P8, PO7, PO8) in response to the presentation of the interpretation. These electrodes were clustered within their respective hemispheres for further analysis (see Fig. 2). A significant main effect for interpretation was found $F(1,25) = 4.964$, $p = .035$. As can be seen in Fig. 5, the P300 was significantly smaller for the pragmatic interpretation than for the semantic interpretation. The Type × Interpretation interaction was not significant $F(4,100) = 1.024$, ($p = .398$), indicating that this response did not vary over scalar type. There was, however, a significant main effect for scalar type, $F(4,100) = 4.922$, $p = .001$. Follow up simple effects tests indicated that the P300 for quantity (some) was significantly smaller than all other scalar types except probability (possible) ($p < .05$). In addition, probability (possible) trials had a significantly smaller P300 amplitude than preference (like) and evaluation (good) trials ($p < .05$). No other significant effects were found.

We then collapsed over scalar type and conducted a 2 X 2 X 2 Threat (threat, no threat) X Interpretation (semantic, pragmatic) X Hemisphere (left, right) repeated measures ANOVA for a P300 (300–500ms) in the bilateral parietal areas (P3, P4, P5, P6, P7, P8, PO7, PO8) in response to the presentation of the interpretation. As in the prior analysis there was a significant main effect for interpretation, $F(1,25) = 5.898$, ($p = .023$), demonstrating that the P300 was significantly larger for the semantic interpretation (See Fig. 6). However, there was also a marginally significant Threat × Interpretation interaction, $F(1,25) = 3.340$, ($p = .08$). As can be seen in Fig. 6, the P300 difference between the pragmatic and semantic interpretations was larger and significant ($p < .05$) in the face-threatening context relative to the non-face-threatening context where the difference was not significant.[3]

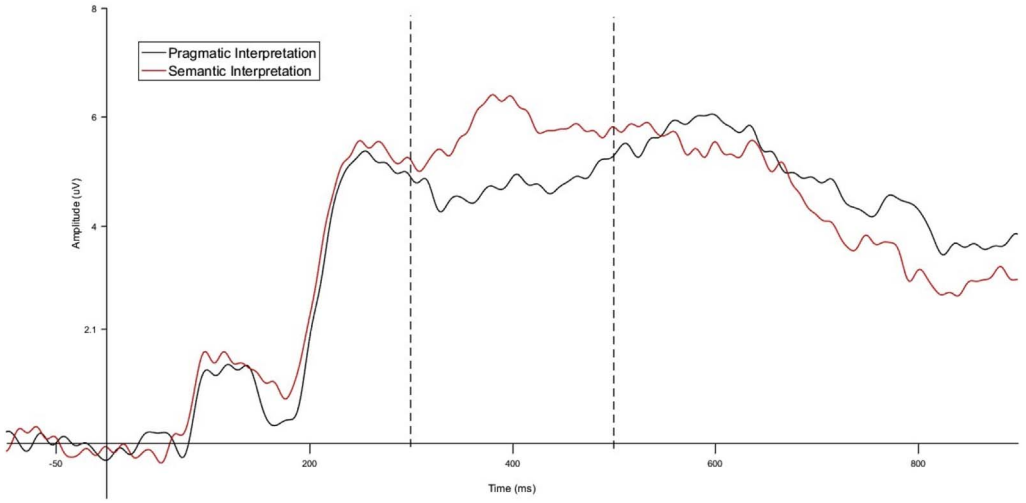

Fig. 5. P300 for Interpretation as a Function of Interpretation Type at electrodes (P3, P5, P7, PO7, P4, P6, P8, PO8).

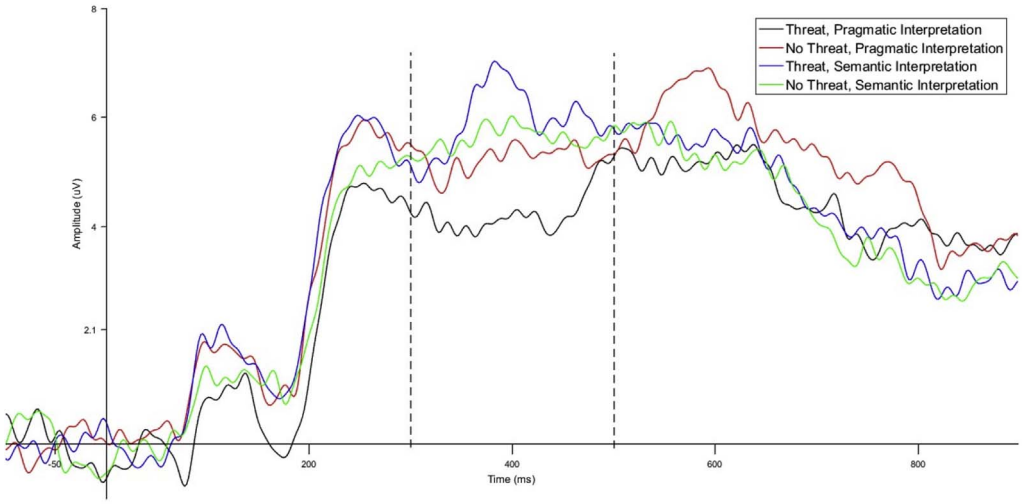

Fig. 6. P300 (dashed line) for Interpretation as a Function of Threat and Interpretation Type at electrodes (P3, P5, P7, PO7, P4, P6, P8, PO8).

4. Discussion

The purpose of this experiment was to examine the processing of several different types of scalar expressions in conversational contexts. Prior research in this area has focused primarily on comprehension of the quantity expression “some” and in non-conversational contexts. In general, the majority of this research has provided support for a contextualist view, that is the pragmatic meaning (i.e., implicature) is more likely to be generated if there is contextual support, and the generation of the implicature is a time-consuming, effortful process. The situation changes somewhat when interpersonal considerations come into play. Bonnefon and colleagues have argued that the processing of scalar expressions is influenced by politeness considerations such that the recognition of a possible politeness motive blocks the pragmatic interpretation (Bonnefon et al., 2009; 2011). It is not clear, however, when and how this process occurs. Mazzarella (2015) has argued that the impact of politeness on the comprehension process occurs later, a view that is consistent with studies demonstrating no effect of context manipulations on processing of the scalar term itself (Hartshorne et al., 2015; Politzer-Ahles & Gwilliams, 2015).

In the present research we examined the processing of scalar terms in a conversational and, more importantly, interpersonal context. We approached this issue by examining neural responses to both the scalar term itself, and to continuations of the utterance

[3] To examine later neural responses, we examined a P600 (525–625ms) in the same bilateral parietal areas (P3, P4, P5, P6, P7, P8, PO7, PO8) in response to the presentation of the interpretation. There was a marginally significant Threat × Interpretation interaction, $F(1,25) = 3.324$, ($p = .08$). Post-hoc simple effects were marginally significant and showed that the P600 for the pragmatic interpretation was larger in nonthreatening contexts than in threatening contexts ($p = .054$). No other effects were significant.

that constituted either a semantic (i.e., unenriched) or pragmatic (i.e., enriched) interpretation of the scalar term. We observed no difference in neural responses to the scalar term itself as a function of face-threat. This finding is consistent with recent studies demonstrating a lack of context effects in the processing of the scalar term itself (Hartshorne et al., 2015; Politzer-Ahles & Fiorentino, 2013; Politzer-Ahles & Gwilliams, 2015), as well as with Mazzarella's (2015) argument that politeness exerts a downstream effect on interpretation. Overall, this pattern suggests that in these contexts the scalar term is processed somewhat superficially, with the context influencing the weighting of different interpretations, and then when the interpretation is encountered it is processed more deeply in order to finalize the meaning of the scalar term.

In contrast to the scalar term, we did observe a difference in a P300 for the subsequent interpretation of the term. Specifically, participants displayed a larger P300 in response to the semantic interpretation than to the pragmatic interpretation, which suggests that the semantic interpretation required additional resources for context updating that were not necessary for the pragmatic interpretation. Consistent with this finding, an increased positivity in the same time window has also been observed in response to the evaluation of conditional arguments, wherein arguments with illogical conclusions showed an increased positivity versus logical conclusions (Bonnefond & Van der Henst, 2013). At the same time, however, our finding contrasts with prior research demonstrating that the semantic interpretation is the default, and that generation of the pragmatic interpretation is effortful and time-consuming (Bott & Noveck, 2004; Breheny et al., 2006). This different pattern of results is likely a function of different contexts and utterances, as well as different dependent measures (see e.g., Degen & Goodman, 2014). As noted by Shetreet et al. (2014), the processing of underinformative statements containing a scalar term may be more time-consuming, not because the semantic interpretation is the default, but because they are logically mismatched with the context. That is, they are logically true but pragmatically incorrect, and it is this mismatch that accounts for the longer response time. In the present study, the scalar terms occurred in conversational contexts when there was no such mismatch.

At the same time, our results suggest that the interpersonal context does matter. We manipulated the degree of face-threat in the context, and hence the extent to which there was a potential politeness motive for the use of a scalar expression. When the context was more face-threatening, the P300 difference between the semantic and pragmatic meanings was larger than when the context was less face-threatening. This suggests that in more face-threatening contexts there is an expectation of politeness, and the clarification of the scalar term via the semantic interpretation violates that expectation; hence, the enhanced difference between the two interpretations when the context is face-threatening. In this interpretation, context updating as indexed by the P300 would require more neural resources when faced with a semantic interpretation, but only when politeness is included as a part of the context. Note that this effect occurred not for the scalar term itself, but rather the subsequent interpretation, again supporting the idea that politeness considerations have their effect post-comprehension of the scalar term (Mazzarella, 2015).

Prior research has focused almost exclusively on the quantity expression “some”, even though there are many different types of scalar terms. Hence, an important issue examined in this research was the question of whether similar processes would be involved in the processing of different scalar types. Although there were some differences among the different scalar types (to be noted below), the comprehension of the different scalar terms exhibited a remarkably similar pattern. Specifically, the enhanced P300 in response to the semantic interpretation relative to the pragmatic interpretation was consistent over scalar types, as was the interaction between face-threat and interpretation type. This consistent pattern provides important evidence regarding the generalizability of this effect across different scalar expressions.

In this research we also considered the possibility of hemispheric differences in the processing of scalar terms. For the interpretation there were no consistent hemispheric differences although there were some differences in response to the presentation of the scalar term. For the scalar term, the P200 was larger in the left hemisphere than in the right hemisphere at frontal sites, when the context was threatening. This finding is consistent with past research which has found increased left frontal P200 amplitudes in response to emotional words (Kanske, Plitschka, & Kotz, 2011). It may be that situations which invoke politeness are more emotionally salient than situations where the probability of offending someone is low.

This hemispheric difference was reversed at parietal sites however, such that the subsequent P300 was significantly larger in the right hemisphere than in the left hemisphere, a pattern suggesting that the right hemisphere may play an important role in processing the meaning of scalar terms, especially in conversational contexts. Such a finding is consistent with prior research on the processing of other ambiguous expressions such as metaphors (e.g., Klepousniotou, Pike, Steinhauer, & Gracco, 2012) and is worthy of additional research.

The major finding in this study – the demonstration of an enhanced P300 in response to a semantic interpretation – contrasts with other research focusing on an N400 in pragmatically uninformative sentences (Nieuwland et al., 2010), as well as other research focusing on a late-term positivity to critical phrases following a scalar term (Hartshorne et al., 2015). There are multiple differences in procedure and content between those studies and the present research that could account for this difference. The major difference, however, is that participants in this study both were given a larger context for each sentence and were required to respond as they processed the utterance. Hence, their neural responses reflect a mixture of comprehension and response processes, a situation in which P300s are likely to be observed. It is also possible that individual sentences lacking a broader context do not require additional context updating during online processing, and that the need for politeness enhances this effect. Note also that the P300-N400 difference may be an important one as it is relevant for the nature of the processes involved in the comprehension of scalar terms. First, the extent to which the P300 and N400 are totally independent is in some dispute (e.g., Arbel, Spencer, & Donchin, 2011). Still, an N400 is typically viewed as indexing the degree of semantic incongruity of an incoming lexical item (Kutas & Hillyard, 1984). A P300, in contrast, indexes a broader reaction to any unexpected stimuli, as well as the process of integrating the unexpected stimuli into the broader context (Polich, 2007). In addition, whereas an N400 has been shown to reflect the activation of long-term semantic memory as a means of making sense of an incoming semantic incongruity, a P300 reflects a more local updating of the unfolding

discourse context, precisely the situation for participants in this experiment as they encountered new information relevant for their comprehension of a scalar term they had viewed just a few words earlier. Hence, the P300 in this experiment can be taken as an indicator of difficulty integrating new information into a previously acquired context (Van Petten & Luka, 2012). Also, the P300 observed in this study occurred in the parietal region, which is typically where a P300b is reported (Polich, 2007).

We note several limitations of this research. First, the lexical items representing the pragmatic and semantic meaning were not identical. Instead, the pragmatic meaning contained a negation (e.g., not always) while the semantic meaning did not (e.g., always). Note, however, that even though it was the pragmatic meaning that included the "negation" (and hence was the more complex construction as evidenced by the longer reaction times), it was the semantic meaning that was surprising and produced the enhanced P300. If the finding was reversed (enhanced P300 for the pragmatic meaning) one could argue that it was the presence of the negation that was causing the enhanced P300. But that's not what happened here. Note also that this limitation is not relevant for the effect of face-threat (as the items were constant across levels of face-threat). Second, the different scalar expressions (and their interpretations) were obviously not identical. However, this limitation is mitigated by the fact that we did not observe differences between them. Third, our findings should be regarded as exploratory in the sense that we did not have specific, a priori predictions. Instead, our analyses were based upon visual inspection of the neural responses. Additionally, due to the complex situations we examined and material that we used, it is possible that multiple processes underlie the pattern of results that we report. For example, the enhanced P300 in response to the semantic meaning may indicate surprise at the face-threatening information.

Another possible limitation of this research is the use of a self-paced reading paradigm. While past research has shown that this type of paradigm is suitable for measuring sentence comprehension (Dittman et al., 2007), it is conceivable that subjects or trials with a significantly shorter reading time than 300 ms may show different response related processes which could account for the differences between conditions. Future studies may test this possibility by either using paradigms which are not self-paced so that the SOAs of each utterance are static, or by varying the location of the interpretation to see if any spillover effects occur during the subsequent utterance (e.g. Nieuwland et al., 2010).

In sum, scalar expressions play an important role in interpersonal interactions and our results suggest that the processing of scalar terms in conversational contexts may differ from their processing in nonconversational contexts. Our understanding of natural language comprehension will be enhanced by future systematic investigations of these terms in conversational contexts.

Author notes

The research was supported by a grant from the National Science Foundation (BCS-1224553) awarded to the first author. The assistance of Chelsea Cadle, Richard Ward, Julie Matsen, Tony Kanazeh, and Zoe Leichty-Wireman in conducting this research is gratefully acknowledged.

Conflicts of interest

No conflicts of interest are present for any of the authors of this manuscript.

Appendix A

Stimulus Materials for all Critical Trials

Note: The non-face-threatening version was created by substituting the material in parentheses for the italicized material. To create the implicature condition, the negation in parentheses [not or didn't] was presented; it was not presented in the implicature rejection.

Quantity1 John couldn't make it to Susan's party. To make up for it, he made her some cookies and brought them over for the party. After the party was over, *John* (Susan's sister) asked *Susan* (her) if any of his cookies were left over. Susan says:

There were some left over, specifically, they were [not] all left over.

Quantity2 Nina was a graduate student and taught her first class. After the class was over, *Nina* (a student who had missed the class) asked *one of the students named James* (James, a student who had attended the class,) if any students were using their phone during class. James says:

There were some students using their phone, specifically, they were [not] all using their phone.

Quantity3 Hannah, an incoming freshman, attended a Welcome Week party for first year students in her residence hall. Later, *the person who organized the party* (a friend from another residence hall) asked her whether any of the first year students left the party early. Hannah replied:

There were some that left early, specifically, they [did not] all left [leave] early.

Quantity4 Ryan went with his family to his sister, Nancy's, dance recital. Afterwards, *Nancy* (his friend Jordan) asked him whether the family was bored at the recital. Ryan said:

There were some that were bored, specifically, they were [not] all bored.

Quantity5 Owen got his girlfriend Robyn a coupon book for Valentine's Day. After a couple of months, *he* (Robyn's friend Lauren) asked *Robyn* (her) if she still had any coupons. Robyn replied:

I still have some coupons left, specifically, they are [not] all left.

Quantity6 Hugo, Sun, and Jin were camping when they noticed their food supplies were gone. They also saw that there were no

animal markings around the food. Sun asked Jin if he (thought Hugo) had eaten the food. Jin replied:

I (Hugo) ate some of the food, specifically, *I* (Hugo) [did not] ate [eat] all of the food.

Quantity7 *On her way home from work, Leah calls her husband Clif. Leah asks Clif if he washed the dishes like he promised he would that morning.* (Clif's neighbor knows that Clif promised his wife that he would wash the dishes while she was at work. At the end of the day, Clif's neighbor asks him if he washed the dishes like he promised.) Clif says:

Right now some of the dishes are dirty, specifically, the dishes are [not] all still dirty.

Quantity8 Fifteen year old Brody was proud that his father trusted him with grilling the hot dogs at the holiday cook-out. When his *father* (cousin) noticed him kneeling on the patio with the platter of hot dogs, he asked Brody if everything was alright. Brody said:

I dropped some of the hot dogs on the ground, specifically, *I dropped* [I did not drop] all of the hot dogs on the ground.

Quantity9 Hector was in charge of booking a band for the fraternity's end of the year bash. He tried his best but his plans fell through. The band he booked cancelled at the last minute. *Hector asked his best friend in the fraternity how much the other brothers were blaming him for the band not showing up. Hector's friend* (One of the fraternity brothers, Doug, was talking about the party with his sister the next day. She asked him how much the other brothers were blaming Hector for the band not showing up. Doug) said:

They think thatsome of it was *your* (his) fault, specifically, they *think it was [not] all your* ([do not] think that all of it was his) fault.

Quantity10 An eager group of girls sang and auditioned for a national talent show. Proud of their performance and out of breath, they awaited the *judge's* (judges') decision. *The judge said* (While privately discussing the performance, one judge said to another):

You (They) were off key some of the time, specifically, *you* (they) were off key [not] all of the time.

Quantity11 Anton fancied himself a chili gourmet. He decided to cook up a pot of chili for his co-workers. Later in the *day he asked Fred, one of his co-workers,* (day, two co-workers, Mark and Fred, were conversing when Mark asked Fred) if he had heard what people thought of *his* (Anton's) chili. Fred replied:

I heard that some people disliked *your* (his) chili, specifically, I heard that [not] everyone disliked *your* (his) chili.

Quantity12 Rachel gave a difficult presentation to a group of co-workers. Later in the day, *she asked Marcie, one of her co-workers who had attended the presentation,* (two co-workers Lynn and Marcie were talking when Lynn asked Marcie) if she had heard what people had thought of *it* (Rachel's presentation). Marcie replied:

I heard that some people disliked it, specifically, I heard that [not] everyone disliked it.

Frequency1 *Beth and John* (Beth, Laura, and John) are classmates and often study together. One day, *Beth* (Laura) asked John if he thought *she* (Beth) was a poor student. John replies:

You are (Beth is) sometimes a poor student, specifically, *you are* (she is) [not] always a poor student.

Frequency2 *Rose and Leo* (Rose, April, and Leo) are friends. *Rose* (April) and Leo were talking when *Rose* (April) asked Leo if he thought *she* (Rose) was moody. Leo responded:

You are (I think Rose is) sometimes moody, specifically, *you are* (I [don't] think Rose is) [not] always moody.

Frequency3 *John and Jake* (John, Jake, and Kevin) decided to move into an apartment together off campus. After living together for about a month, *Jake* (Kevin) asked John if he thought *he* (Jake) was lazy. John responds:

You are (Jake is) sometimes lazy, specifically, *you are* (he's) [not] always lazy.

Frequency4 *Ron and Sarah* (Ron, Linda, and Sarah) were enrolled in Driver's Education. *They are talking together one day* (Linda and Sarah were talking) when *Ron asks* (Linda asked) Sarah if she *thinks he is* (thought Ron was) a bad driver. Sarah replied:

You are (Ron is) sometimes a bad driver, specifically, *you are* (Ron is) [not] always a bad driver.

Frequency5 Jack and James recently went on a beach vacation. When they returned, Jack asked his friend, Kate, how noticeable *his* (James') sunburn was. Kate replied:

Your (His) sunburn is sometimes noticeable, specifically, it is [not] always noticeable.

Frequency6 *Tony overheard his six year old son, Milo, telling* (Tony's six year old son, Milo, told) a friend how annoying his dad is. *Tony* (Milo then) asked his eight year old *daughter, Layla,* (sister) if she is annoyed by *him* (their dad). *Layla* (Tony's daughter) said:

You (He) sometimes *annoy* (annoys) me, specifically, *you* (he) [does not] always annoy me.

Frequency7 At the restaurant where Felicity works, employees are not allowed to take food home with them. One day, *her boss catches her boxing up a salad after clocking out.* (she brings a salad home from work.) Felicity's *boss* (roommate) asks her how often she takes food home (from work). Felicity says:

I sometimes take food home, specifically, I [do not] always take food home.

Frequency8 John and Elizabeth like to kiss. One day, *John* (Elizabeth's friend Rita) asked *her if he* (if John) was a bad kisser. Elizabeth replied:

You are (He is) sometimes a bad kisser, specifically, *you are* (he is) [not] always a bad kisser.

Frequency9 Crissola was a massage therapist. One day Greg, a regular customer with back pain, came into her office for a massage. Afterwards, *Crissola* (Greg's wife) asked him if *her* (Crissola's) massages were ineffective. Greg said:

Your (Her) massages are sometimes ineffective, specifically, they are [not] always ineffective.

Frequency10 (Five year old) London is a student in Alaina's gymnastics class. Today the class is doing somersaults. Afterward, *London* (another instructor) asks Alaina about *her* (London's) somersaults. Alaina says

Your (Her) somersaults are sometimes poor, specifically, they are [not] always poor.

Frequency11 Mary and Jim are married. Whenever they go anywhere Jim insists on driving. *Jim and Mary are driving one evening when Jim* (One day Mary's sister Jane) asks *Mary* (her) whether she thinks *he sometimes* (Jim) drives too fast. Mary replies:

You (He) sometimes drive too fast, specifically, *you* (he) [*do* (does) not] always drive too fast.

Frequency12 Gabe and Leah have been married for 10 years. One night *Gabe* (Leah's sister, Michelle,) asks *Leah* (her) how good *he* (Gabe) is at remembering to do his chores. Leah replies:

You (He) sometimes forgets to do his chores, specifically, *you* (he) [*do not* (doesn't)] always forget to do his chores.

Preference1 Vincent invited Susan (and Natasha) to see his art exhibit at a museum. Later, *Vincent* (Natasha) and Susan were talking when *Vincent* (Natasha) asked Susan what she thought of *his* (Vincent's) art exhibit. Susan responds:

I liked it, specifically, I [didn't] love it.

Preference2 Cameron rented his favorite action movie to watch with his girlfriend, Phoebe. After the movie, *he* (her friend, Cristiana,) asked her if she liked it. Phoebe replied:

I liked it, specifically, I [didn't] love it.

Preference3 Jeff bought his wife Kathy a tennis bracelet for their anniversary. Later *Jeff* (Kathy's friend) asked her if she liked it. Kathy replied:

I liked it, specifically, I loved [didn't love] it.

Preference4 Allie's *boyfriend, Brett,* (friend James) asked her if she likes Skittles. *Allie knows Skittles are Brett's favorite candy.* Allie says:

I like them, specifically, I [don't] love them.

Preference5 Donald bought his girlfriend Sandy a new necklace for her birthday. Later that night, *Donald* (Sandy's friend Cindy) asked Sandy if she liked it. Sandy replied:

I like the necklace, specifically, I [don't] love it.

Preference6 Brice gave Amanda a stylish new haircut. Afterwards, *Brice* (Amanda's friend Danielle) asked *Amanda* (her) if she liked it. Amanda replied:

I like it, specifically, I [don't] love it.

Preference7 Bubba painted the kitchen for his girlfriend Karen's birthday. Later that night, *Bubba* (Karen's friend, Sherry,) asked Karen if she liked it. Karen said:

I like it, specifically, I [don't] love it.

Preference8 Ben spent a good part of the afternoon making his favorite meal for his wife, Julie. After the meal, *Ben* (Julie's friend, Libby,) asked *her how she liked it.* (if she liked the meal.) Julie *replied* (said):

I liked the meal, specifically, I loved [didn't love] it.

Preference9 Heather brought her neighbor, Sam, homemade banana bread as thanks for plowing her driveway. The next day, *Heather* (Sam's coworker) asked Sam what he thought of *her* (Heather's) banana bread. Sam said:

I liked it, specifically, I loved [didn't love] it.

Preference10 After singing in the church choir for a few years, William decided to sing a solo piece at the Christmas service. Feeling nervous about the upcoming performance, he asked his fellow choir member, Janette, what she thought of his singing. Janette (Two other choir members, Janette and Susie, were talking about William's upcoming solo. Janette asked Susie what she thought of William's singing. Susie) said:

I like the way *you* (he) sings, specifically, I [don't] love the way *you* (he) sings.

Preference11 After Carter's many complaints about his old flip phone, Carter's mom picked out a smart phone and bought it for him. *Grinning from ear to ear, she surprised him with it at home one day.* Two days *later she* (later, Carter's friend) asked him what he thought of his new phone. Carter said:

I like my new phone, specifically, I [don't] love my new phone.

Preference12 Mitch was an aspiring documentary film maker. He was having a screening for his first film and he invited his best friend Daniel to watch it. *After it was over he asked Daniel what he thought.* (Later, Daniel's roommate asked him what he thought of Mitch's film.) Daniel said:

I liked it, specifically, I loved [didn't love] it.

Probability1 Mrs. Taylor was cleaning the room of her two sons, Jon and Joe. While doing so, she discovered a pack of cigarettes. While Joe was at soccer practice, she asked Jon whether the cigarettes were *his* (Joe's). Jon replied:

It's possible that the cigarettes are *mine* (Joe's), specifically, it's [not] certain that the cigarettes are *mine* (Joe's).

Probability2 Dr. Jones recently discovered that two of his students, James and Ethan, scored very similarly on a challenging exam, and he suspected that cheating had taken place. But he was unsure who was guilty. So, he asked Ethan if *he* (James) was the one that had cheated. Ethan replied:

It's possible *I* (James) cheated on the exam, specifically, it's [not] certain that *I* (James) cheated on the exam.

Probability3 Allie's ice cream was missing from the freezer, and she suspected that one of her roommates had taken it without asking. But which one? That evening, she asks Katie if *she* (Brianna) ate her ice cream. Katie replies:

It's possible that *I* (Brianna) ate your ice cream, specifically, it's [not] certain that *I* (Brianna) ate your ice cream.

Probability4 Trevin, Caleb, and Aaron moved into a new apartment off campus. When they were finally all unpacked, Trevin discovered that his flat screen TV would no longer work. He suspected that one of his roommates dropped it on the way in. He decides to ask Caleb if *he* (Aaron) dropped the TV. Caleb replies:

It's possible that *I* (Aaron) dropped your TV, specifically, it's [not] certain that *I* (Aaron) dropped your TV.

Probability5 Nathaniel lost his cat. He suspects that one of his roommates left a door open and the cat escaped. Nathaniel asked his roommate Jordan if he *left* (saw their roommate, Corbin, leave) the door open. Jordan replies:

It's possible *I* (Corbin) left the door open, specifically, it's [not] certain that *I* (Corbin) left the door open.

Probability6 Steve just returned home from his first semester of college and noticed that his expensive cologne bottle was empty. He suspected his brother Fred of using it while he was away. *When he confronts Fred, Fred* (He asked his mother if she had seen Fred use his cologne. She) replies:

It's possible *I* (he) used your cologne, specifically, it's [not] certain that *I* (he) used your cologne.

Probability7 Ronald thinks that his neighbor Chris has been throwing trash on his lawn. *When Ronald asks Chris about the trash on his lawn, Chris* (He decides to ask his neighbor Shawn if he saw Chris throw trash on his lawn. Shawn) replies:

It's possible *I* (Chris) threw trash on your lawn, specifically, it's [not] certain that *I* (Chris) threw trash on your lawn.

Probability8 Thomas believes that his roommate Reginald has been taking his food from the refrigerator. When Thomas asked Reginald about this Reginald (their roommate Brent if he also thinks Reginald is taking his food from the fridge, Brent) replies:

It's possible *I've* (he's) been taking your food, specifically, it's [not] certain that *I've* (he's) been taking your food.

Probability9 Kim suspects that Kyle's dog has been going to the bathroom on her front lawn. When she asks *Kyle* (her neighbor Jeff) about this *Kyle* (Jeff) replies:

It's possible *he's* (Kyle's dog has) gone to the bathroom on your lawn, specifically, it's[not] certain *he's* (Kyle's dog has) gone to the bathroom on your lawn.

Probability10 Two hours from home, while driving down the highway, Bill asked Trudy if *she* (their son had) closed the garage door like *she said she* (he said he) would. Trudy replies:

I possibly (It's possible he) closed the garage door, specifically, it's [not] certain *I* (he) closed the garage door.

Probability11 Kyler and Nathan were taking care of their pet hedgehog when they noticed it was not looking well. Nathan noticed the medicine bottle was still full and he asked Kyler if *he had given* (their roommate Andre could have forgotten to give) the hedgehog its medicine. Kyler replies:

It's possible *I* (he) forgot to give him his medicine, specifically, it's [not] certain *I* (he) forgot to give him his medicine.

Probability12 After hearing a rumor about her boyfriend, Ian, at last night's party, Karina asked *Ian* (Ian's friend) if *he* (Ian) kissed another girl at the party. *Ian* (Ian's friend) replies:

It's possible *I* (he) kissed her, specifically, it's [not] certain *I* (he) kissed her.

Evaluation1 Grace made her roommate, Lucy, a pumpkin cheesecake for her birthday. Later on, *Grace* (their neighbor Dan) asked Lucy what she thought of it. Lucy stated:

It was good, specifically, it was [not] excellent.

Evaluation2 Lena, an aspiring interior designer, decorated the living room of her neighbors, the Smiths, while they were on a weekend vacation. When they returned, *Lena* (their friends, the Jones,) asked them what they thought of it. The Smiths responded:

It looked good, specifically, it looked [not] excellent.

Evaluation3 Thomas took Bethany out on a date. At the end of the night, *Bethany* (Thomas' roommate) asked Thomas how he thought the date went. Thomas responded:

It was good, specifically, it was [not] excellent.

Evaluation4 Jackie had to read poetry to her English class in order to complete an assignment. Later, *Jackie* (Brian) asked Jane, a fellow classmate, what she thought of *her* (Jackie's) poetry reading. Jane says:

It was good, specifically, it was [not] excellent.

Evaluation5 James and Lilly attended a potluck dinner. Their friends, the Smiths, made many dishes for the dinner. At the end of the dinner, *the Smiths* (James) asked Lilly what she thought of *their* (the Smith's) food. Lilly said:

The food was good, specifically, it was [not] excellent.

Evaluation6 Mary recommended a new restaurant downtown to Zach. After Zach dined there, *Mary* (Zach's friend Rachel) asked him what he thought of the restaurant. Zach replied:

I thought it was good, specifically, I thought it was [not] excellent.

Evaluation7 Giovanni made his prized chili dogs for his girlfriend Stephanie. After she ate, *he* (Stephanie's friend Tanya) asked her if she liked them. Stephanie replied:

I thought they were good, specifically, *but* (I thought) they were [not] excellent.

Evaluation8 Tyler asked Julie to go with him to see his favorite movie that was playing in theaters. After they saw it, *Tyler* (Julie's friend Anna) asked *Julie* (her) if she liked it. Julie replied:

It was good, specifically, it was [not] excellent.

Evaluation9 Anna gave her friend, Echo, a scarf she had made. Later, *Anna* (Echo's friend, Maurice,) asked Echo if he liked the scarf. Echo said:

I thought it was good, specifically, it was [not] excellent.

Evaluation10 Laura watched a performance of the play Fiddler on the Roof. Later, *the cast of the play stood in the theater's lobby to mingle with audience members. Laura was on her way through the lobby when Greg, one of the actors*, (her friend, who watched the play with her,) asked her what she thought of the performance. Laura said:

I thought the play was good, specifically, it was [not] excellent.

Evaluation11 Alan was looking to buy a used car that was reliable. His friend Damon convinced him to buy his car, promising it would not let him down. After a while, *Damon* (Alan's dad) asked *Alan* (him) how the car was. Alan said:

The car is good, specifically, it is [not] excellent.

Evaluation12 Sam was working hard on his first novel. When he finished he asked his older brother, Guy, to read *it and let him know what he thought. When Sam asked Guy for his opinion,* (it. After reading it, Guy's wife asked him what he thought of it and) Guy said:

The novel is good, specifically, it is [not] excellent.

Sample Filler trials

FILL Andy wanted his sister Crystal to meet his girlfriend, Ashley. Later, Andy asks Crystal what she thought of Ashley. Crystal says:

I liked her, specifically, I hated her.

FILL Kevin wrote Samantha a song and sang it to her. Later, someone asks Samantha what she thought of Kevin's song. Samantha says:

I liked it, specifically, I didn't hate it.

FILL Alice, Bill, and Daniel are friends. Bill and Daniel were talking when Bill asks Daniel if he thought Alice was lazy. Daniel replies:

She is sometimes lazy, specifically, she is never lazy.

FILL Jill wrote a poem and read it to Ashley. Later, Jill asks Ashley what she thought of her poem. Ashley says:

I thought it was good, specifically, I thought it was terrible.

FILL Kevin came home to find his last slice of pizza missing. When his roommate Bill came home, Kevin asked him if he had eaten his last slice of pizza. Bill replied:

It's possible that I ate your last slice of pizza, specifically, it's not certain I ate your last slice of pizza.

FILL Jim sold some of Carla's jewelry during a garage sale. At the end of the garage sale, Carla asked Jim if there was any jewelry left. Jim said:

There was some jewelry left, specifically, there was no jewelry left.

FILL Dr. Johnson suspects that one of his students, Keith, is cheating on tests. He asked another student, Kayla, if Steve is a cheater. Kayla said:

Steve sometimes cheats on tests, specifically, he never cheats on tests.

References

Arbel, Y., Sp[illegible]K. M., & Donchin, E. (2011). The N400 and P300 are not all that independent. *Psychophysiology, 48*, 861–874.

Bergen, L., & [illegible]er, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1450–1460. http://dx.doi.org/10.1037/a0027850.

Bonnefon, J. F., De Neys, W., & Feeney, A. (2011). Processing scalar inferences in face- threatening contexts. In C. H. L. Carlson, & T. Shipley (Eds.). *Proceedings of the 33rd annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Bonnefond, M., & Van der Henst, J. B. (2013). Deduction electrified: ERPs elicited by the processing of words in conditional arguments. *Brain and Language, 124*(3), 244–256.

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

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

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

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*(3), 434–463.

Degen, J., & Goodman, N. (2014, January). Lost your marbles? The puzzle of dependent measures in experimental pragmatics. *Proceedings of the cognitive science society: Vol. 36* No. 36.

Degen, J., & Tanenhaus, M. K. (2011). Making inferences: The case of scalar implicature processing. *Paper presented at the Proceedings of the 33rd annual conference of the cognitive science society*.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.

Dittman, T., Holcomb, P., & Kuperberg, G. (2007). An investigation of concurrent ERP and self-paced reading. *Psychophysiology, 44*(6), 927–935.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. *Behavioral and Brain Sciences, 11*(3), 357–427.

Greischar, L. L., Burghya, C. A., van Reekum, C. M., Jackson, D. C., Pizzagalli, D. A., et al. (2004). Effects of electrode density and electrolyte spreading in dense array electroencephalographic recording. *Clinical Neurophysiology, 115*, 710–720.

Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. *Foundations of Language,* 225–242.

Hartshorne, J. K., Snedeker, J., Liem Azar, S. Y. M., & Kim, A. E. (2015). The neural computation of scalar implicature. *Language, Cognition and Neuroscience, 30*(5), 620–634.

Holtgraves, T. (2014). Interpreting uncertainty terms. *Journal of Personality and Social Psychology, 107*(2), 219.

Horn, L. R. (1976). *On the semantic properties of logical operators in English.* University Linguistics Club.

Horn, L. R. (1989). *A natural history of negation.* University of Chicago Press Chicago.

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia, 49*(11), 3121–3129.

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language, 123*(1), 11–21.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 30*, 161–163.

Levinson, S. (1983). *Pragmatics.* Cambridge, MA: MIT Press.

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

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences, 94*(20), 10979–10984.

Mazzarella, D. (2015). Politeness, relevance, and scalar inferences. *Journal of Pragmatics, 79*, 93–106.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R., Koppenhaver, A. M., Greischar, L. L., et al. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage, 49*(3), 2416–2432.

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*(3), 324–346.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203–210.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience & Biobehavioral Reviews, 30*(6), 808–822.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Politzer-Ahles, S., & Fiorentino, R. (2013). The realization of scalar inferences: Context sensitivity without processing cost. *PLoS One, 8*, 1–6.

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*, 134–152.

Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of prefrontal cortex in scalar implicatures: Evidence from magnetoencephalography. *Language, Cognition, and Neuroscience, 30*, 853–866.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide.* Psychology Software Incorporated.

Shetreet, E., Chierchia, G., & Gaab, N. (2014). When some is not every: Dissociating scalar implicature generation and mismatch. *Human Brain Mapping, 35*(4), 1503–1514.

Sperber, D., & Wilson, D. (1995). *Relevance* (2nd ed.). Cambridge, MA: Cambridge University Press.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*(1), 18–35.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190.