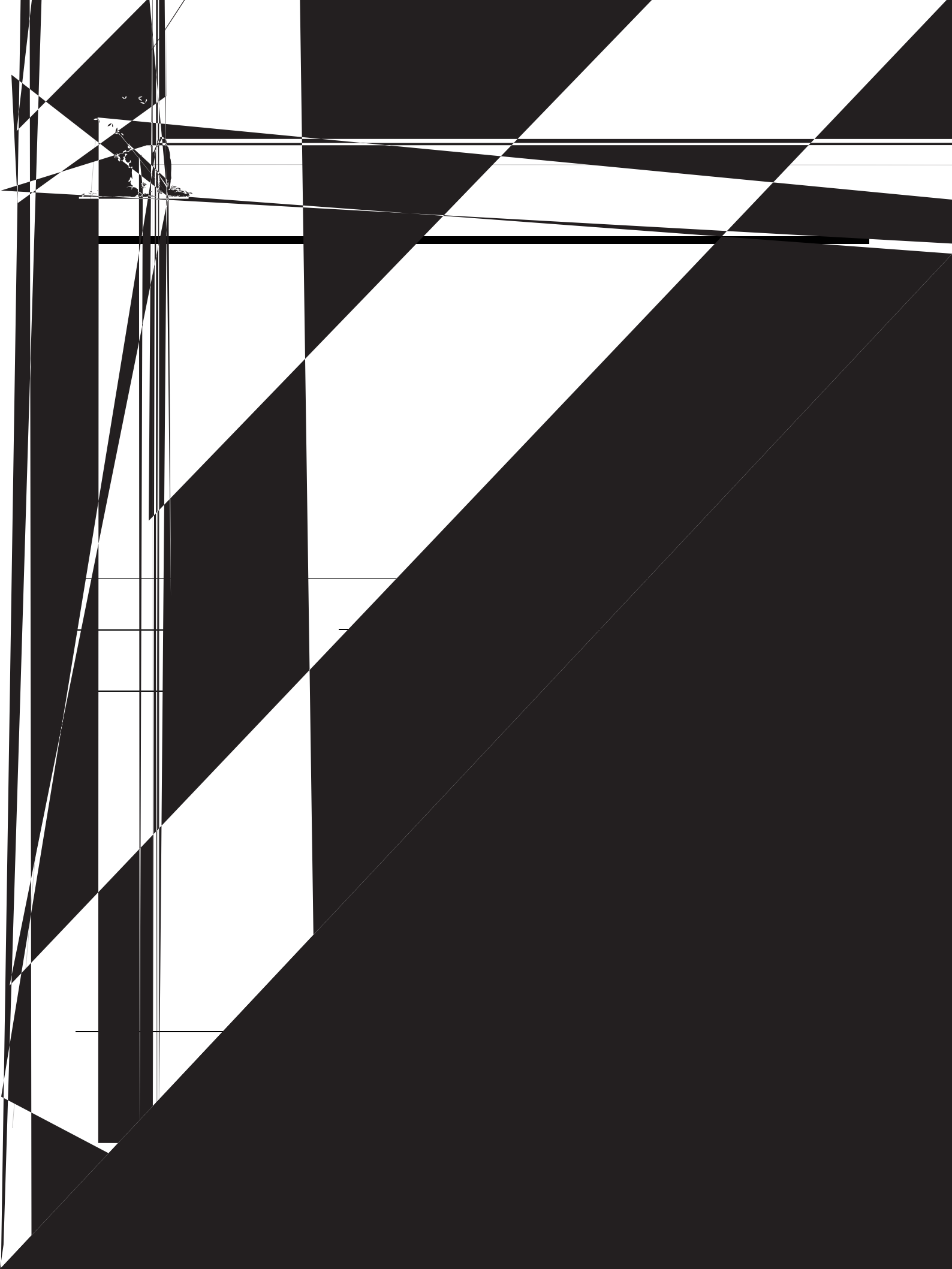

## 1. Introduction

Understanding an utterance involves rapidly combining elements of its meaning from multiple sources, including the

quantifiers rather than the downstream effects of expectations generated by pragmatic inferencing.

The experiments reported here were conducted in Mandarin Chinese, whereas previous online studies of scalar implicature have all used western languages. The characteristics of Mandarin scalar implicature, however, are not different from those of English (see Chi, 2000; Xie, 2003; Tsai, 2004; Rullman and You, 2006; Wu and Tan, 2009). The critical scalar quantifier in the present experiment is yǒu de (有的), which is partitive (Xie, 2003; Tsai, 2004) and has a strongly pragmatic interpretation (Wu and Tan, 2009) adult participants reported a pragmatic interpretation of yǒu de in 89% of trials). It is roughly equivalent in meaning to the English partitive some of, which robustly elicits a pragmatic interpretation (Grodner et al., 2010; Degen and TanenI

inconsistent all of

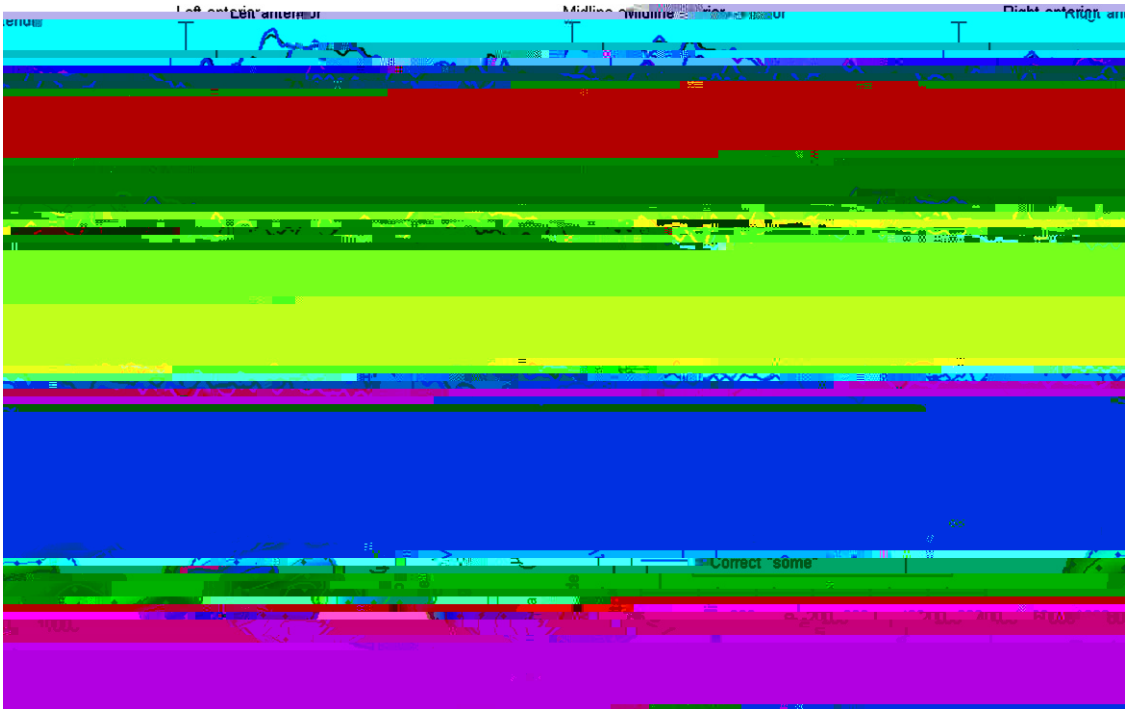


Fig. 2 – Event related potentials. Upper left: Grand average ERP (a 30 Hz low-pass filter) at the anterior electrodes. Lower left: The grand average topographical distribution of the ERP.

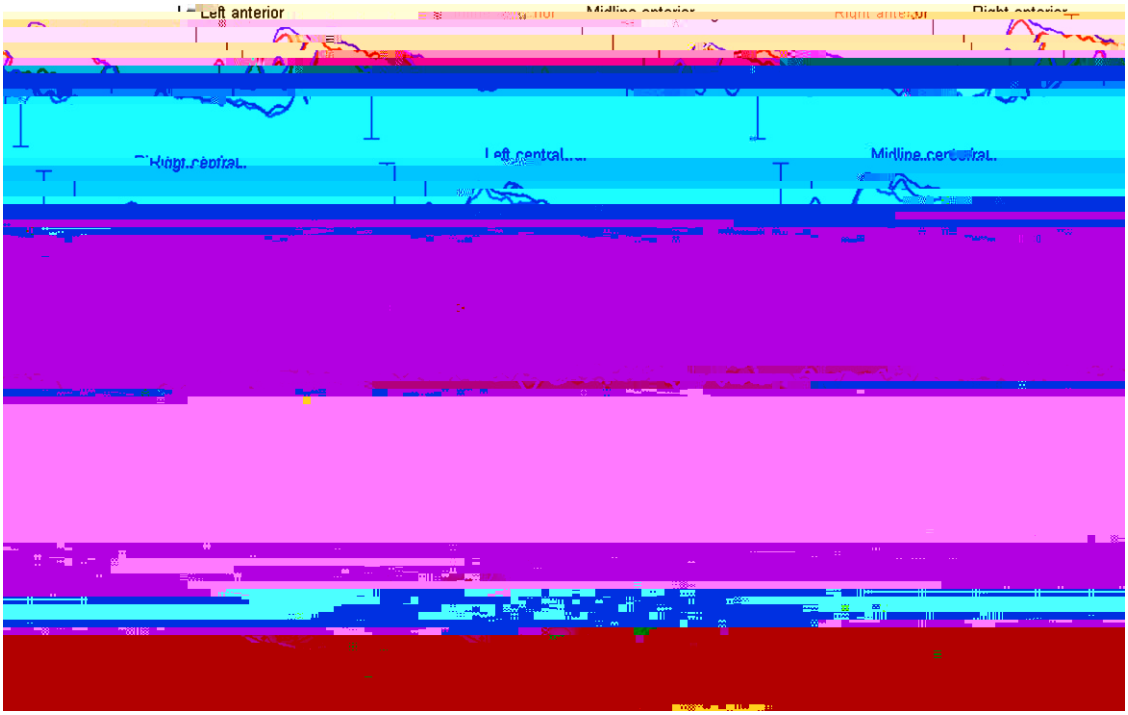


Fig. 3 – Event related potentials. Upper left: Grand average ERP (a 30 Hz low-pass filter) at the central electrodes. Lower left: The grand average topographical distribution of the ERP.

2.1.2.2. 500–1000 ms. In the later time window there was a

similar vein, our Experiment 2 tests whether revising an under-informative scalar inference interferes with lexico-semantic integration between the picture and the sentential object. We also include the same Quantifier by Consistency manipulation at the quantifier position as we had in Experiment 1, in order to test whether the effect obtained in that experiment would be replicated. (The pragmatically inconsistent “some” and correct “some” conditions were included in the critical items; items corresponding to the semantically inconsistent “all” and correct “all” of Experiment 1 were included in the fillers for this experiment.) While the primary motivation for Experiment 2 was to examine the interaction of pragmatic and lexical processing rather than effects of modality, we found that auditory presentation of sentences was both comfortable for participants and reduced the duration of each trial. For this reason, sentence stimuli were presented auditorily rather than visually in Experiment 2.

## 2.2. 2

### 2.2.1. Behavioral results

The participants’ task was to rate the consistency between the picture and the sentence using a 7-point scale. Average ratings were 6.3 for correct some of and 6.2 for correct all of sentences, 5.4 for pragmatic violations, 3.2 for lexical violations, 2.3 for double violations and 2.0 for semantically incorrect all of sentences. A

observed a highly significant effect of Lexical Consistency, reflecting the fact that both lexically inconsistent conditions (picture-sentence mismatch and double violation) elicited more negative ERPs than lexically consistent conditions (correct object, and correct object following a pragmatically inconsistent quantifier). The effect was broadly distributed (it did not interact significantly with Region). The effect of Pragmatic Consistency was not significant. Crucially, no interactions of Pragmatic Consistency and Lexical Consis-

analyze the sentence as "… all the girls are wearing bathing suits" or "… some of the girls are happy"). Because the structure of the verbs used in the present study varied (verbs were presented simultaneously with aspect markers that preceded or followed them and differed in length and other properties) as

with semantic meaning and that the generation of scalar implicatures is strongly affected by context and expectations about speakers.

The present study offers converging evidence with other emerging work in neurosemantics suggesting that the mechanisms by which the brain composes meaning may not be the same as those by which it accesses words from the lexicon, notices associations between words, or evaluates real-world plausibility (i.e., several of the processes reflected by the N400). Recent investigations suggest that the patterns of brain activation elicited by violations of real-world plausibility are not the same as those elicited by linguistically-motivated abstract operations such as semantic composition (Pylkkänen et al., 2011), licensing of negative polarity items (Steinhauer et al., 2010; Panizza, 2012) and semantic subcategorization (Kuperberg et al., 2000). In our experiments we found that quantifiers which were pragmatically inconsistent with a context elicited a qualitatively different ERP response than quantifiers which were semantically inconsistent, suggesting that they were processed by different mechanisms. We also found that costly pragmatic reanalysis of a quantifier's meaning did not modulate concurrent processing of lexico-semantic errors, providing further evidence that the processing of these two aspects of meaning are processed independently. We note, however, that while the qualitative differences in ERP responses found in the present study are consistent with distinct mechanisms of pragmatic and semantic meaning composition, it is difficult to infer the underlying sources of the ERP pattern. For this reason, localizing the neural generators of these effects using methods with high spatial resolution would be a valuable avenue

pictures paired with matching, felicitous some of sentences, and thirty-seven were All-type pictures paired with matching, correct all of sentences. The other seventy-four pictures were paired with sentences that had appropriate quantifiers but either an object that did not match any of the objects in the picture of a verb that did not match the activity shown. Several of these included verbs that yielded semantically anomalous sentences (e.g., "all the scientists are planting squirrels"), whereas most had verbs that were semantically plausible but not congruous with the picture (e.g., "all the boys are going for a walk with their classmates", after a picture in which all the boys are wrestling with their classmates). The filler sentences all included quantifiers that were not used in the critical sentences but were similar in meaning to all of or some of, or classifier phrases in place of quantifiers. None of the filler sentences used numbers in the place of quantifiers (for discussion of how the presence/absence of numbers and quantifiers in the experimental context may affect the perception of scalar implicature, see Degen, 2009; Grodner et al., 2010; Huang et al., 2010; and references therein). The set of fillers with mismatching pictures and sentences was included to distract participants from the quantifier manipulation in the critical sentences, and the remaining matching fillers were included to maintain a proportion of acceptable sentences of at least 50% during the experiment, assuming that pragmatically infelicitous stimuli are judged as unacceptable.

#### 4.1.3. Procedure

Participants were seated in a dimly-lit room about 1 m in front of a 41 cm CRT monitor. Stimuli were presented at the center of the screen using the Presentation software package (Neurobehavioral Systems). Each trial began with a fixation point presented for 500 ms, followed by a picture which remained on the screen for 4000 ms. The picture was followed by a fixation point of random duration (between 500 and 1500 ms), after which the sentence was presented region by region using the serial visual presentation paradigm. Regions were presented using a variable presentation procedure (see, e.g., Nieuwland et al., 2010), whereby each region was presented at a base duration of 425 ms per region, plus 80 ms for each character more than 3 in the region; because the critical quantifiers were all three characters or less, their presentation durations do not differ across conditions. The interstimulus interval was 400 ms for all regions.[7] Twenty percent of trials were followed by comprehension questions or acceptability judgments (see below), which were presented on the screen for 5000 ms or until the participant's response. Each trial was followed by a blank screen for 1500 ms before the start of the next trial. The experiment was divided into six blocks of approximately 50 sentences each, and participants were given short breaks between the blocks. Participants were instructed not to blink during the presentation of the sentences.

Participants performed a mixture of acceptability judgments and comprehension questions. On ten percent of trials, after the sentence ended, a question that probed information about the picture and was irrelevant to the sentence was presented (e.g., after the sentence "In this picture, some of the girls are sitting on blankets", the comprehension question "Are the girls wearing swimsuits?" appeared). In an additional ten percent of trials, the sentence was followed instead by an acceptability judgment (the question "

by a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Inc.) with a bandpass of 0.01 to 200 Hz, and digitized at a sampling rate of 1000 Hz.

The continuous EEG was re-referenced to the average of both mastoids and segmented into epochs from 1 s before to

using Praat (Boersma and Weenink, 2012), and the onset latencies of the quantifiers and objects were measured. The onset of the quantifier yǒu de (some of) was defined as the point of lowest intensity between the preceding syllable lǐ and the yǒu, which in most tokens also coincided with a perceptible change in phoneme quality and preceded, by 10–20 ms, a 200–400 Hz drop in frequency of the second through fourth formants. The onset of the quantifier suǒyǒu de (all of) was defined as the onset of high-frequency energy in the spectrogram. Onsets of objects were measured as the audible onset of the first consonant of the word (plosives were measured at the burst), except in two cases where the onset of the first consonant of the second syllable was measured since this was the point of disambiguation for the critical word. The latency between quantifier onsets and object onsets in the critical sentences was 1309 ms on average (sd = 203 ms, range 832–2127 ms).

The 400 trials (160 critical some of sentences, 80 all of fillers, and 160 other fillers) were arranged into four lists in a Latin square design. Each list contained 40 trials per object condition. For the all of sentences tested, each list contained 40 trials per condition (correct “all”, semantically inconsiste-291mnnt mfiall’F’

inconsistent), Lexical Consistency (consistent, inconsistent), and Region. The Huynh–Feldt correction was applied to F-tests with more than one degree of freedom in the numerator.

## Acknowledgments

This research was supported by the National Science Foundation East Asia and Pacific Summer Institutes (award ID #1015160) to SPA, the China Post-Doctoral Science Foundation (award IDs #20100480150, #2012T50005) to XJ, and the Natural Science Foundation of China (award ID #30970889) and Ministry of Science and Technology of China (award ID# 2010CB833904) to XZ. Experiment design, data analysis, and preparation of this manuscript was completed by the authors. The authors thank Liang Yan, Wu Chunping, and Wu Yue, Luo Yingyi, Zhu Mengyan, Wu Junru, and Lamar Hunt III for assistance in the construction of materials; Wu Yin and Zhou Yuqin for assistance with data collection; and Jamie Bost and Natalie Pak for assistance in the preparation of this manuscript.

## References

Baggio, G., van Lambalgen, M., Hagoort, P., 2008. Computing and recomputing discourse models: an ERP study. J. Mem. Lang. 59, 36–53.

Bezuidenhout, A., Cutting, J., 2002. Literal meaning, minimal propositions and pragmatic processing. J. Pragmat. 34, 433–456.

Boersma, P., & Weenink, D. (2012). Praat: Doing Phonetics by Computer [Computer program]. 〈http://www.praat.org/〉.

Bonin, P., Peereman, R., Malardier, N., Méot, A., Chalard, M., 2003. A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. Behav. Res. Methods Instrum. Comput. 35, 158–167.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Schlesewsky, M., 2011. Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. Brain Lang. 117, 133–152.

Bott, L., Bailey, T., Grodner, D., 2012. Distinguishing speed from accuracy in scalar implicatures. J. Mem. Lang. 66, 123–142.

Bott, L., Noveck, I., 2004. Some utterances are underinformative: the onset and time course of scalar inferences. J. Mem. Lang. 51, 437–457.

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

Chevallier, C., Noveck, I., Nazir, T., Bott, L., Lanzetti, V., Sperber, D., 2008. Making disjunctions exclusive. Q. J. Exp. Psychol. 61, 1741–1760.