"Maybe" not all scalar implicatures are created equal

Stephen Politzer-Ahles[*]

LSA Annual Meeting, Minneapolis, January 2–5, 2014

1. Introduction. An utterance with an informationally weak expression is often understood as meaning that a proposition with a stronger expression is false (e.g., "I ate some of the cookies" implies "I did not eat all of the cookies", and "I might eat cookies" implies "It is not certain that I will eat cookies"). This phenomenon is known as *scalar implicature* (Horn, 1972), and is a type of quantity implicature (Geurts, 2010). Expressions triggering scalar implicatures typically have both a core semantic meaning (e.g., *some* meaning "more than none", and *might* meaning "there is a more than 0% chance") and an upper-bounded, enriched meaning based on conversational implicature (Grice, 1975) or on semantic exhaustification (Chierchia, Fox, & Spector, 2012). While there is strong linguistic evidence that these aspects of meaning are distinct (for instance, the implicature or enriched meaning is cancellable, while the core semantic meaning is not), an important question in psycholinguistics and neurolinguistics is whether there are observable cognitive and neural correlates of this distinction.

Neurolinguistic methods provide a good means to investigate the time course and nature of the realization of the different aspects of meaning for scalar expressions. Recently, event-related potentials (ERPs), which provide millisecond temporal resolution, have been used to examine brain responses elicited by infelicitous scalar expressions (e.g., "some of the girls", when actually "all of the girls" is true). While content words downstream of the implicature-triggering expression robustly elicit a negative brain potential called the N400 when the previously-computed scalar inference makes these words unexpected (Noveck & Posada, 2003; Nieuwland, Ditman, & Kuperberg, 2010; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Sikos, Tomlinson, & Grodner, 2013; Spychalska, Kontinen, & Werning, 2014), the neural response elicited by an infelicitous scalar expression itself is less clear. Politzer-Ahles, Fiorentino, Jiang, and Zhou (2013) tested cases in which a pragmatic violation is noticeable as soon as the scalar expression is heard, and found that these expressions elicit a sustained negative wave, which may be related to inhibiting the dominant pragmatic interpretation. On the other hand, Panizza and Onea (2014) did not replicate this effect when testing a similar paradigm (in German, as opposed to Chinese). Both of these studies used picture-sentence verification paradigms, in which a picture is first presented (e.g., a picture of five girls who are all sitting on chairs) and then is followed by a sentence (e.g., "Some of the girls are sitting on chairs") which is rendered inconsistent with the picture by the pragmatic interpretation of its quantifier.

Furthermore, most neurolinguistic research on scalar implicatures focuses on the expression *some* (or its equivalent in other languages) and its "not all" interpretation; one study has exampled the <*or*, *and*> scale in French (Chevallier, Bonnefond, Van der Henst, & Noveck, 2010), and no other quantity implicatures have been tested. There are many other types of expressions that trigger quantity inferences more or less robustly than "some" (van Tiel et al., 2014; Doran et al., 2009, 2012). It therefore remains an open question how much the previous ERP findings extend to scalar inferences in general, as opposed to scalar inferences in nominal quantifiers specifically. Likewise, we do not yet know whether or not behavioral differences in the rates at which different expressions trigger scalar inferences reflect qualitative differences in how these different inferences are realized.

[*] Author contact information: Stephen Politzer-Ahles, New York University Abu Dhabi (spa268@nyu.edu).

The present experiment had two goals. First, to examine whether a sustained negativity would again be observed for infelicitous scalar quantifiers (in a different language than the previous experiments that used this paradigm); and second, to test the neural processing of a previously untested scale. We thus focused on the <*maybe*, *definitely*> scale.

2. Methods. 17 participants provided informed consent and took part in the task. All methods were approved by the IRB of NYU Abu Dhabi, and participants received payment. Each critical trial began with a picture and a question about the picture (see Figure 1 for examples), followed by a fixation point with a random duration (following a normal distribution with mean=500 ms and sd=200 ms), followed by the word *maybe*, which was either a correct answer to the question (e.g., in the left portion of Figure 1, the case in which it is not clear whether or not the soda in the pitcher will fit in the cup), a semantically false answer (as in the middle portion of Figure 1, where it is obvious that the soda will *not* fit in the cup), or a pragmatically infelicitous answer (as in the right portion of Figure 1, where it is obvious that the soda *will* fit in the cup).

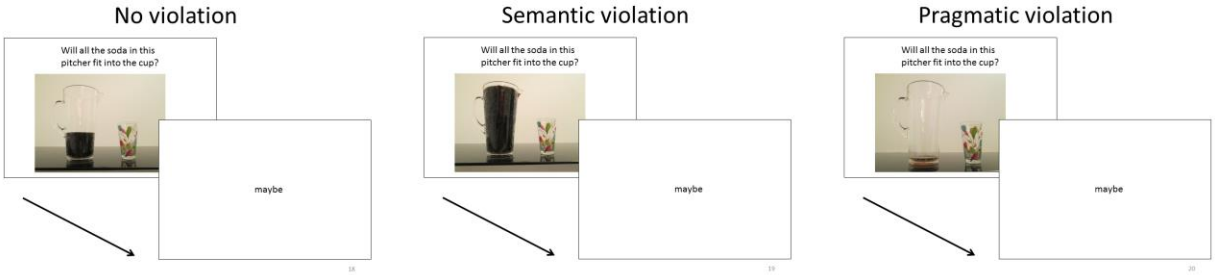

Figure 1. Sample stimuli

Fillers followed the same structure but used *definitely* and *definitely not* as answers. The correspondence between the extreme pictures and correctness was reversed for half of the trials (e.g., half of the trials asked questions like "Is this too much soda to fit in the cup?" rather than "Will all the soda in this pitcher fit in the cup?"). Five different picture scenarios were used, and all versions of each picture were normed on Mechanical Turk to ensure that participants agreed the "definitely" and "definitely not" versions were obvious and the "maybe" versions were not. The number of trials per condition is shown in Table 1 (critical conditions are in bold):

		Picture context		
		DEFINITELY	DEFINITELY NOT	MAYBE
Critical word	*definitely*	30	20	30
	definitely not	20	30	30
	maybe	**40**	**40**	**40**

Table 1: Number of trials per condition

EEG and MEG were concurrently recorded while participants viewed the trials. At the end of each trial (after the critical word was on-screen for 800 ms) the participant's task was to indicate whether she agreed or disagreed with the answer (e.g., the word *maybe*) with respect to the question that had been asked about the picture. EEG data were re-referenced to the average of both mastoids, high-pass filtered at 0.1 Hz and epoched from -300 to 1300 ms, baseline-corrected, cleaned for ocular artifacts using ICA, and manually artifact-rejected. Three participants were excluded from the analysis because of excessive artifacts.

3. Results. Behaviorally, participants indicated they agreed with 79.85% of the correct answers, 29.85% of the pragmatically infelicitous answers, and 15.91% of the semantically incorrect answers. The differences between all pairs of conditions were significant (*p*s < .001) in a generalized linear mixed-effects model.

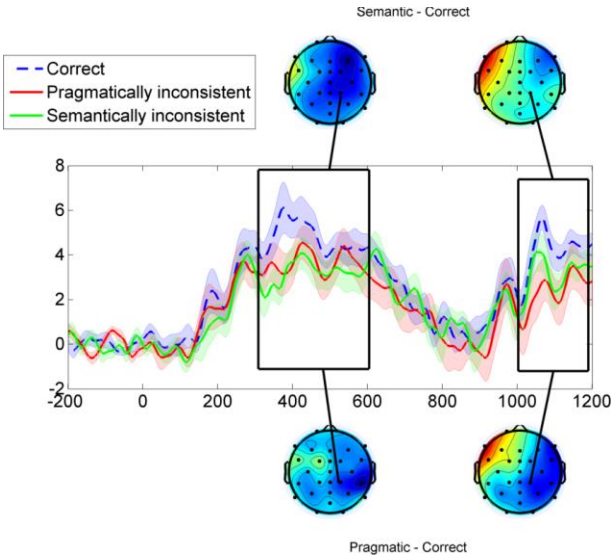

Figure 2. ERP results at electrode CP2. Ribbons represent ±2 SE.

Figure 2 shows the ERPs at electrode CP2. In the N400 time window, semantic violations elicited an N400 relative to correct trials, which was significant at $\alpha = .05$ both in ANOVA on the mean amplitude of the time window and in non-parametric spatiotemporal clustering. In this same window, pragmatic violations appear to elicit an N400 relative to correct trials, although this effect is marginal in the time window ANOVA and non-significant in spatiotemporal clustering. The semantic and pragmatic violations did not differ in this time window.

In a late time window (1000-1200 ms after the appearance of the critical word, and thus 200-400 ms after the appearance of the overt judgment prompt), the pragmatic violations (relative to correct trials) elicited a right-posterior negativity in both the spatiotemporal clustering and ANOVA analyses, whereas the semantic violations elicited only a left-anterior positivity, and only in the ANOVA analysis. In this time window, semantic and pragmatic violations significantly differed from one another when compared directly in the ANOVA analysis, although not in the spatiotemporal clustering analysis.

4. Discussion. The present experiment was an attempt to replicate the sustained negativities observed for pragmatic violations in Politzer-Ahles et al. (2013) using the <*maybe*, *definitely*> scale rather than the <*some*, *all*> scale. While a late negativity was indeed elicited, it was much later and weaker than what was seen in the previous studies; furthermore, the pragmatic violations showed a trend towards eliciting an N400-like effect, which was not observed in the previous studies. Methodologically, the present study differed from the previous studies in that the critical word was presented in isolation (as the one-word answer to a question) rather than embedded within a sentence, and this study used a smaller amount of visual contexts, each of which was repeated several times.

As the results are somewhat equivocal, it would be valuable to compare *some* and *maybe* within one experiment, using the exact same paradigm. If differences were still observed, what

might they mean? First of all, the picture-sentence verification task may induce non-linguistic processing differences between *maybe* and *some*. For example, a participant can easily see whether *some* and *all* are true in a picture, whereas evaluating the truth of *maybe* requires mentalizing possible events. Secondly, there may also be linguistic differences that influence how the respective scalar implicatures are processed. For instance, intuition suggests that the "not definitely" interpretation of *maybe* is less easy to cancel than the "not all" interpretation of *some*. Likewise, *some* and *maybe* may differ in terms of the number and nature of relative alternatives (i.e., other numeral quantifiers, other adverbs).

In summary, the results suggest some interesting differences between the processing of *some* and *maybe*, but more importantly they highlight the need for controlled studies comparing the two within the same paradigm.

References

Chevallier, Coralie, Mathilde Bonnefond, Jean-Baptiste Van der Henst, & Ira Noveck. 2010. Using ERPs to capture inferential processes guided by prosodic cues. *Italian Journal of Linguistics, 22*, 125-152. http://www.italian-journal-linguistics.com/wp-content/uploads/chevallier-et-al..pdf

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In Maienborn, von Heusinger & Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning, Vol. 3*, pp. 2297-2331, Berlin: Mouton de Gruyter.

Doran, Ryan, Rachel Baker, Yaron McNabb, Meredith Larson, & Gregory Ward. 2009. On the non-unified nature of scalar implicature: an empirical investigation. *International Review of Pragmatics, 1*, 211-248.

Doran, Ryan, Gregory Ward, Meredity Larson, Yaron McNabb, & Rachel Baker. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language, 88*, 124-154.

Geurts, Bart. 2010. *Quantity Implicatures*. Cambridge: Cambridge University Press.

Grice, Herbert Paul. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics 3: Speech acts*, pp. 41-58, New York: Academic Press.

Horn, Laurence. 1972. *On the semantic properties of logical operators in English*. Los Angeles: University of California dissertation.

Hunt, Lamar III, Stephen Politzer-Ahles, Linzi Gibson, Utako Minai, & Robert Fiorentino. 2013. Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture-sentence verification. *Neuroscience Letters, 534*, 246-251.

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

Noveck, Ira, & Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*, 203-210.

Panizza, Daniele, & Edgar Onea. 2014. Some implicatures take their time: an ERP study on scalar implicatures with 'sentence-picture vs. picture-sentence' verification task. Talk presented at *27th CUNY Conference on Human Sentence Processing*. Columbus, OH.

Politzer-Ahles, Stephen, Robert Fiorentino, Xiaoming Jiang, & Xiaolin Zhou. 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.

Sikos, Les, Sam Tomlinson, Hilary Traut & Daniel Grodner. .2013. Incremental computation of scalar implicatures: An ERP study. Poster presented at *26th CUNY Conference on Human Sentence Processing*. Columbia, SC.

Spychalska, Maria, Jarmo Kontinen, & Markus Werning. 2014. Electrophysiology of pragmatic processing: exploring the processing cost of the scalar implicature in the truth-value judgment task. In P. Bello et al. (eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pp. 1497-1502, Austin, TX: Cognitive Science Society.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina, & Bart Geurts. 2014. Scalar diversity. *Journal of Semantics*. Advance access. http://jos.oxfordjournals.org/content/early/2014/12/22/jos.ffu017.full