Supplementary Material:
What we Know about Knowing:
Presuppositions generated by factive verbs influence downstream neural processing

Table of Contents

1	Brief review of previous ERP studies examining processes associated with accommodating presuppositions	p. 2
2	Traditional analyses of ERPs averaged across multiple spatiotemporal regions	p. 4
3	Additional discussion: The absence of N400 modulation on event-consistent versus event-inconsistent critical words	p. 8
4	References	p. 13

1. Brief review of previous ERP studies examining processes associated with accommodating presuppositions

In the present study, we were interested in the neural consequences of *violating* presuppositions. As discussed in the main text, when presuppositions are violated, the new input cannot be integrated into the comprehender's discourse model unless she revises this model by abandoning the original presupposition. In other cases, however, it may be possible to accommodate a presupposition, even if it is not immediately satisfied. There are several studies that have examined ERPs associated with the process of accommodating presuppositions. Most of these studies measured ERPs produced by definite noun-phrases (NP, e.g. "the girl"), which trigger the so-called *uniqueness presupposition* (see Heim, 1982) — the inference that the NP refers back to a unique referent in the previous context. In some of these studies, ERPs were time-locked to the determiner, which was presented separately from the noun (e.g. Kirsten et al., 2014; Anderson & Holcomb, 2005); in others, ERPs were time-locked to the definite noun-phrase, which was presented as a whole (Burkhardt, 2006; Domaneschi, Canal, Masia et al., 2018).

In these studies, the context either introduced a unique referent such that the presupposition was explicitly satisfied on the definite NP, or the context was manipulated in various different ways such that the presupposition could be accommodated. In some studies, the context introduced a group of possible referents and so the presupposition could be accommodated by selecting one of these referents (e.g. from Kirsten et al., 2014: "Antje visited Dusseldorf zoo yesterday and saw *some polar bears* in the bear enclosure. Antje noticed that <u>the polar bear</u> was very aggressive"; from Domaneschi, Canal, Masia et al., 2018: "In Paolo's office there are *many employees.* The office needs consultants for several branches of the activity. Due to overstaffing problems, about a month ago, <u>the graphic designer</u> was made redundant."). Some studies

included scenarios in which the context did not introduce any referents, and so, to accommodate the presupposition, comprehenders needed to infer new referents (e.g. from Burkhardt, 2006: "Tobias talked to Nina. He said that the conductor was very impressive."). Finally, Van Berkum et al. (1999, 2003) carried out a series of studies in which the context introduced two named referents of the same gender, and so, to accommodate the presupposition, the comprehender needed to temporarily restrict the domain reference to just one referent, with the assumption that the ambiguity in interpretation would later be resolved (e.g. "David had told *the two girls* to clean up their room before lunchtime. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time. David told the girl that had been on the phone to hang up").

Various different ERP effects have been reported in association with these manipulations. First, several studies report modulation of the N400 on definite NPs, with a smaller N400 evoked when the presupposition was explicitly satisfied by the prior context than when the presupposition was accommodated (Anderson & Holcomb, 2005; Burkhardt, 2006; Domaneschi, Canal, Masia et al., 2018). These N400 findings, however, are difficult to interpret because explicit presupposition satisfaction is confounded with repetition priming on the NP (see Anderson & Holcomb, 2005, for discussion). Second, some studies report larger late positivities on definite NPs when the presupposition had to be accommodated than when it was explicitly satisfied. This was interpreted as reflecting the selection of a unique referent within the context (Domaneschi et al., 2018; Kirsten et al., 2014) or the retrieval of a novel referent from long-term memory (Burkhardt, 2006). Finally, the studies by Van Berkum et al. (1999, 2003) report a larger late anterior prolonged negativity on ambiguous definite NPs referring back to two possible referents versus one possible referent. This late anterior negativity effect may have

reflected costs associated with maintaining the presupposition in working memory such that it could be accommodated by future inputs (see also Anderson & Holcomb, 2005). Interestingly, as noted in the Introduction, in a previous study of factive verbs, Ferretti et al. (2008, 2013) also reported a larger late negativity on words like "oranges" in factive than non-factive scenarios (e.g. "The coach determined (*versus* figured) that that it was oranges that…") when they followed events like "Ken and his brother ate some apples". It is possible that this too reflected similar working memory costs associated with accommodating the presupposition: in such scenarios, comprehenders could accommodate the presupposition if they assumed that the word, "oranges", served as the theme of an unstated future event (e.g. inferring that, despite the fact that Ken and his brother ate apples, they actually *wanted* to eat oranges).

2. Traditional analyses of ERPs averaged across multiple spatiotemporal regions

In an early analysis of this dataset, following a more traditional approach to ERP analysis (e.g., in our lab, Delaney-Busch & Kuperberg, 2013; Fields & Kuperberg, 2015), we carried out an omnibus repeated-measures ANOVA, which we followed up with simple effects ANOVAs. In these analyses, the dependent variable was the average ERP evoked across pre-specified temporal windows within five separate 3-electrode spatial regions along the anterior-posterior axis of the scalp (see Figure S1). Spatial region was included as a within-subject independent variable, and any interactions between spatial region and the other factors were followed up in the individual spatial regions.

As recently discussed by Luck and Gaspelin (2017), however, this type of ANOVA approach creates multiple opportunities to detect effects in different spatial regions (potentially, a

main effect, interactions with Region and additional effects at each region), which increases Type I error. This is why we chose to reanalyze our data with a mass univariate approach, which does not have this problem (and, when used in combination with regions of interest, does not sacrifice power to detect effects see Fields & Kuperberg, 2018). However, for completeness, here we report the pattern of findings of our original ANOVA, together with follow-ups.

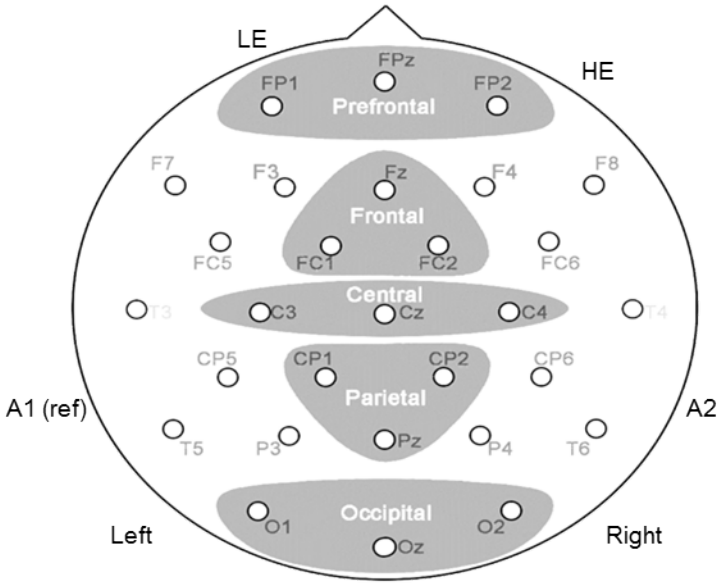

Figure S1. Scalp distribution of the regions used in the ANOVA analyses.

Methods

Three omnibus repeated measures ANOVAs were conducted in which the dependent variable was the ERP averaged across three time windows: 300-500ms, 500-700ms and 700-900ms (the same time-windows that we used for the mass univariate analyses reported in the main text). These ANOVAs included three within-participant variables: Verb factivity (two levels: factive and non-factive), Event consistency (two levels: consistent and inconsistent), and Region (five levels: prefrontal, frontal, central, parietal, and occipital, see Figure S1). Follow-up analyses were then conducted for the factive and non-factive scenarios separately, with Event consistency

and Region as within-subject variables. Significant interactions between Event consistency and Region were further followed up by testing the effects of Event consistency within each region. In all analyses, the Greenhouse and Geisser (1959) correction was applied to repeated measures with more than one degree of freedom, and a significance level of alpha = .05 was used for all comparisons.

Results

500-700ms. The omnibus ANOVA revealed a significant three-way interaction between Verb factivity, Event consistency and Region ($F(4, 92) = 4.25$, $p = .028$). To follow up this interaction, we carried out follow-up ANOVAs comparing the effects of Event consistency within each region in the factive and non-factive scenarios separately. In the factive scenarios, event inconsistent critical words evoked a larger positivity than event consistent critical words in the parietal region ($F(1, 23) = 7.1$, $p = .014$) and the occipital region ($F(1, 23) = 4.73$, $p = .04$) — that is, a larger late posterior positivity/P600. In the non-factive scenarios, no effects of Event consistency were seen in any region ($Fs < 0.72$, $p > 0.4$).

700-900ms. The omnibus ANOVA revealed a main effect of Event Consistency ($F(1, 23) = 5.30$, $p = .031$), which reflected larger widely distributed positivity on event inconsistent than consistent critical words, regardless of verb factivity. There was no main effect of Verb factivity, and no interaction involving these two factors (all $Fs < 1.3$, $ps > .27$).

300-500ms. The omnibus ANOVA revealed a significant three-way interaction between Verb factivity, Event consistency and Region ($F(4, 92) = 3.66$, $p = .035$). No other effects were found. We carried out follow-up ANOVAs examining the effects of Event consistency within each

region in the factive and the non-factive scenarios separately. In the factive scenarios, no effects were significant in any regions (Fs < 1.25, $p > 0.27$). In the non-factive scenarios, there were effects of event consistency that approached significance in frontal ($F(1, 23) = 3.98$, $p = .058$) and central ($F(1, 23) = 4$, $p = .057$) regions. As illustrated in Figure S2 (which shows this contrast at each 100ms time window), these effects appear to be driven by a more positive waveform on inconsistent than consistent critical words in the non-factive scenarios.

Discussion

The results of these ANOVA analyses revealed the same pattern of findings as the mass univariate analyses reported in the main manuscript: no significant effect of Event consistency on the centro-parietally distributed N400 component evoked by critical words in either the factive or the non-factive scenarios; a significant effect of event consistency on the posteriorly distributed late posterior positivity/P600 between 500-700ms in the factive scenarios but not in the non-factive scenarios, and a main effect of Event consistency in the 700-900ms time window.

In the non-factive scenarios, we also saw marginal effects of Event consistency between 300-500ms over frontal and central regions. This appeared to be driven by a larger frontal positivity in the non-factive scenarios on event inconsistent than event consistent critical words (see Figure S2). We are hesitant to interpret this effect given that it did not reach significance in either of these regions, even at a very liberal threshold of $p = 0.05$, which did not account for multiple comparisons (see Luck and Gaspelin, 2017). We also saw no evidence of a significant positivity effect when we carried out a post-hoc mass univariate analysis over frontal regions. Nonetheless, it will be interesting to see if an anterior positivity effect is reliably produced in future studies using similar designs with more power. This is because there are some previous

reports of anterior positivity effects, beginning within the N400 time window and sometimes extending into later time windows, on unpredicted critical words in plausible scenarios. Such anterior positivity effects are classically evoked by plausible words that violate strong lexical predictions in highly constraining contexts (DeLong, Urbach, Groppe & Kutas, 2011; Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007; Van Petten & Luka, 2012)). However, they can also sometimes be evoked by plausible words in less lexically constraining contexts, like in the present study (e.g. Thornhill & Van Petten, 2012), perhaps when such words provide new information that triggers comprehenders to update their discourse model.

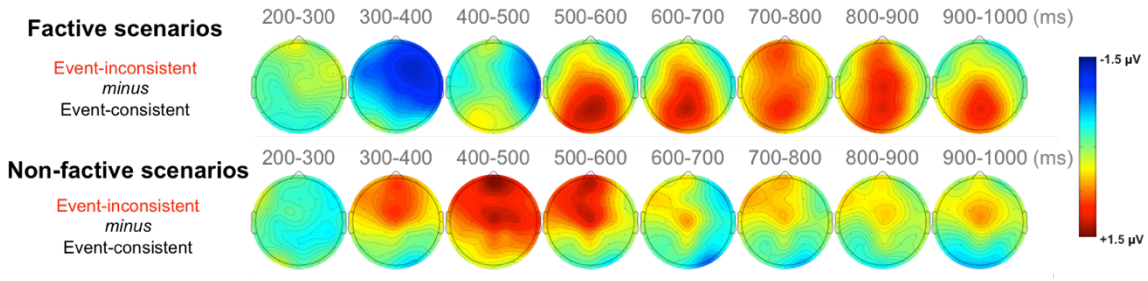

Figure S2. Effects of event consistency in the factive and non-factive scenarios at each 100ms time window. Mean voltage differences are shown. No effects reached significance within the 300-500ms time window (see text).

3. Additional Discussion: The absence of N400 modulation on event consistent versus event inconsistent critical words

In the present study, our main focus with regard to the N400 was whether or not there was an interaction between Verb factivity and Event consistency. As discussed in the main text, we saw no evidence of such an interaction. Indeed, we also saw no main effect of Event consistency on the N400. This was also the case when we examined the factive and non-factive scenarios separately. We discuss this latter finding here.

The absence of N400 modulation on words that are inconsistent versus consistent with their preceding context is not unusual. Although modulation on the N400 was first reported in response to words that were inconsistent versus consistent with their preceding sentential (Kutas & Hillyard, 1980) or discourse (van Berkum, Hagoort & Brown, 1999) contexts, and although the amplitude of the N400 often patterns with plausibility/coherence ratings and/or off-line measures of predictability like cloze probability, this is by no means always the case (for reviews, see Kuperberg, 2007 and Kuperberg, 2016). There are two broad sets of reasons why one may not see N400 modulation in comparing contextually consistent and inconsistent critical words: (A) no difference in semantic facilitation on contextually consistent versus inconsistent critical words; (B) a larger overlapping late positivity on contextually inconsistent than contextually consistent critical words.

A) <u>No difference in semantic facilitation on contextually consistent versus inconsistent critical words.</u>

(i) The related anomaly effect: A highly lexically constraining context can lead comprehenders to predict a specific upcoming word. This, in turn, can pre-activate semantic information within long-term semantic memory and 'prime' a semantically related critical word, even if it is incongruous with the context. This can lead to a partial attenuation of the N400 on contextually inconsistent critical words that are semantically related to a predicted word — the so-called 'related anomaly effect' (Federmeier & Kutas, 1999). In the present study, this is unlikely to explain the absence of N400 modulation on the event consistent versus inconsistent critical words (in either the factive or the non-factive scenarios). This is because the related anomaly

effect is only seen following very highly constraining contexts and not in medium-constraining contexts (Federmeier & Kutas, 1999), like those used in the present study.

(ii) A locally constraining context: A locally highly constraining context can sometimes lead to semantic facilitation on critical words, and a partial attenuation of the N400, even if such words are inconsistent with the global context (e.g. from Boudewyn, Long & Swaab, 2015: "Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some healthy and tasty veggies…"). This also cannot explain the pattern of N400 findings in the present study: even though there were some event inconsistent items that had greater-than-zero cloze probabilities, when we excluded these items and re-analyzed the data with only zero-cloze event inconsistent items, the results were the same: the N400 was still just as small on event inconsistent as on event consistent critical words.

(iii) No differential pre-activation of upcoming specific events over and above broad schema:

If broad schema-based semantic relationships are matched across conditions (e.g. using measures such as Latent Semantic Analysis, Landauer & Dumais, 1997; Landauer et al., 1998), and incoming event inconsistent critical words are consistent with these broad schema-based relationships, then one can sometimes fail to see semantic facilitation on contextually consistent than contextually inconsistent critical words. In these situations, whether or not one sees differential modulation on the N400 depends on whether, during comprehension, participants are able to *further* pre-activate specific upcoming specific events, over and above broad schema, before the semantic properties of an incoming critical word become available from the bottom-up input (e.g. Paczynski & Kuperberg, 2012; Xiang & Kuperberg, 2015; see Kuperberg, 2016 for a recent discussion). This, in turn, can depend on many different factors within the context that can influence the time it takes to predict upcoming specific events. This includes syntactic

complexity (e.g. Kolk, Chwilla, van Herten & Oor, 2003), the distance between a verb and its arguments (e.g. Chow, Lau, Wang & Phillips, 2018), the presence of a negation operator (e.g. Nieuwland & Kuperberg, 2008, the absence of strong pragmatic communicative cues (e.g. Xiang & Kuperberg, 2015), the rate at which the input unfolds (e.g. Camblin, Ledoux, Boudewyn, Gordon & Swaab, 2007; Ludtke, Friedrich, De Filippis, & Kaup, 2008), and individual differences (e.g. Nakano, Saron & Swaab, 2010, Kim, Oines & Miyake, 2017).[1]

In the present study, we suggest that in the event consistent scenarios, none of these factors led comprehenders to enhance the pre-activation of semantic features associated with upcoming specific events over and above features that were pre-activated by broad general schema in the event inconsistent scenarios.

B) Component overlap from a larger late positivity to event inconsistent than event consistent critical words

A second set of reasons why the amplitude of the N400 can sometimes appear to be just as small on contextually inconsistent as on contextually consistent critical words relates to the early onset of a larger late positive-going waveform on contextually inconsistent than contextually consistent critical words (e.g. Kuperberg, Sitnikova, Caplan & Holcomb, 2003; Kuperberg, 2007). If such a late positivity begins within the N400 time window, and its scalp distribution

[1] Nieuwland and Kuperberg (2008) showed that, in negated sentences, modulation on the N400 can depend on whether or not a sentence is pragmatically licensed. In the present study, the scenarios were not negated. However, because one of our reviewers speculated that some of these scenarios may have been pragmatically unlicensed, we gave 13 Tufts undergraduate students the event consistent scenarios (with factive and non-factive versions counterbalanced across the two lists), and we asked them to identify any scenario that seemed odd or unnatural. We interspersed these scenarios with eight probe scenarios that were purposefully pragmatically odd (adapted from the pragmatically odd sentences used in previous studies by Nieuwland & Kuperberg, 2008; Nieuwland, Ditman, & Kuperberg, 2010). Less than 0.1% of our scenarios were judged by at least three people as being odd or unnatural (0.03% of the factive scenarios and 0.08% of the non-factive scenarios). In contrast, all eight of the pragmatically odd probe scenarios were judged as odd by at least three people. When we re-analyzed our data excluding any scenario in which either the factive and/or the non-factive version was judged as odd by at least three raters, we found the same pattern of findings on both the N400 and the late posterior positivity/P600.

overlaps with that of the N400, then, even if contextually inconsistent critical words produced a large amplitude N400, this would not be detectable on the scalp surface because of spatiotemporal component overlap. Alternatively, the neurocognitive mechanisms reflected by the positivity effect might 'switch off' deep semantic processing of incoming words reflected by the N400 (e.g. Kuperberg, Sitnikova, Caplan & Holcomb, 2003; see also Delaney-Busch & Kuperberg, 2013; Lotze, Tune, Schlesewsky & Bornkessel-Schlesewsky, 2011).

Both these reasons could account for the absence of an effect of event consistency on the N400 in the present study. In the factive scenarios, the larger late posterior positivity/P600 on the event inconsistent critical words may well have started within the N400 time window. Although the non-factive event inconsistent critical words did not evoke a late posterior positivity/P600, they did appear to evoke a larger *anteriorly* distributed positivity between 300-500ms (although this was not statistically significant, see Section 2 above). Thus, in both the factive and the non-factive scenarios, positive-going waveforms may have obscured the appearance of an N400 effect on the scalp surface and/or 'switched off' deep semantic processing of the critical word.

While this study cannot disentangle these two possibilities, both accounts imply that the retrieval of semantic features associated with incoming words (as reflected by the N400) occurs in parallel with attempts to integrate the incoming word into its wider discourse context (as reflected by the late positivities).

4. References

Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language, 94*, 200-216.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 15*(3), 607-624.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159-168.

Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience, 33*(7), 803-828. doi:10.1080/23273798.2018.1427878

Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience, 13*(3), 473-490. doi:10.3758/s13415-013-0159-5

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203-1207. doi:10.1111/j.1469-8986.2011.01199.x

Domaneschi, F., Canal, P., Masia, V., Lombardi Vallauri, E., & Bambini, V. (2018). N400 and P600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics, 45*, 13-35. doi:10.1016/j.jneuroling.2017.08.002

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495. doi:10.1006/Jmla.1999.2660

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

Ferretti, T. R., Singer, M., & Harwood, J. (2013). Processes of discourse integration: evidence from event-related brain potentials. *Discourse Processes, 50*(3), 165-186.

Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition, 108*(3), 881-888. doi:10.1016/j.cognition.2008.06.002

Fields, E. C., & Kuperberg, G. R. (2015). Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias. *Social Cognitive and Affective Neuroscience, 10*(9), 1202-1209. doi:10.1093/scan/nsv004

Fields, E. C., & Kuperberg, G. R. (2018). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. *PsyArXiv preprint*. doi:10.31234/osf.io/qfkgc

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* (Doctor of Philosophy), University of Massachusetts, Amherst, MA.

Kim, A. E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language

comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 44*(3), 406-420. doi:10.1037/xlm0000457

Kirsten, M., Tiemann, S., Seibold, V., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29*(9), 1147-1162.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang, 85*(1), 1-36. doi:10.1016/S0093-934X(02)00548-5

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49. doi:10.1016/j.brainres.2006.12.063

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience, 31*(5), 602-616.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117-129. doi:10.1016/S0926-6410(03)00086-7

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205. doi:10.1126/science.7350657

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev, 104*(2), 211-240. doi:10.1037/0033-295x.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Process, 25*(2-3), 259-284. doi:10.1080/01638539809545028

Lotze, N., Tune, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia, 49*(13), 3573-3582. doi:10.1016/j.neuropsychologia.2011.09.009

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22*(12), 2886-2898. doi:10.1162/jocn.2009.21400

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. doi:10.1016/j.jml.2010.06.005

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: an event-related potential study on the pragmatics of negation. *Psychological Science, 19*(12), 1213-1218. doi:10.1111/j.1467-9280.2008.02226.x

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67*(4), 426-448. doi:10.1016/j.jml.2012.07.003

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392. doi:10.1016/j.ijpsycho.2011.12.007

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147-182.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*(2), 235-248. doi:10.1111/1469-8986.00025

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190. doi:10.1016/j.ijpsycho.2011.09.015

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience, 30*(6), 648-672. doi:10.1080/23273798.2014.995679