

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

Einat Shetreet^{1,2,3}, Edward J. Alexander¹, Jacopo Romoli⁴, Gennaro Chierchia⁵, and Gina Kuperberg^{1,6}

¹ Department of Psychology, Tufts University, Medford, Massachusetts, USA

²Department of Linguistics, Tel Aviv University, Tel Aviv, Israel

³Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

⁴Department of Linguistics, University of Ulster, UK

⁵Department of Linguistics, Harvard University, Cambridge, Massachusetts, USA

⁶Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital, Harvard Medical School, USA

Corresponding Author:

Einat Shetreet

Telephone: 972-3-6405025

Email: shetreet@post.tau.ac.il

Word count: 12,597

Abstract

Presuppositions convey information that comprehenders assume to be true, even when it is tangential to the communicator's main message. For example, a class of verbs called 'factives' (e.g. *realize, know*) trigger the presupposition that the events or states conveyed by their sentential complements are true. In contrast, non-factive verbs (e.g. *think, believe*) do not trigger this presupposition. We asked whether, during language comprehension, presuppositions triggered by factive verbs are encoded within the comprehender's discourse model, with neural consequences if violated by later bottom-up inputs. Using event-related potentials (ERPs), we examined neural activity to words that were either consistent or inconsistent with events/states conveyed by the complements of factive versus non-factive verbs while comprehenders read and actively monitored the coherence of short discourse scenarios. We focused on the modulation of a posteriorly-distributed late positivity or P600. This ERP component is produced when comprehenders constrain their discourse model such that it restricts predictions only to event structures that are compatible with this model, and new input violates these event structure predictions. Between 500-700ms, we observed a larger amplitude late posterior positivity/P600 on words that were inconsistent (versus consistent) with the events/states conveyed by the complements of factive verbs. No such effect was observed following non-factive verbs. These findings suggest that, during active discourse comprehension, the presuppositions triggered by factive verbs are encoded and maintained within the comprehender's discourse model. Downstream input that is inconsistent with these presuppositions violates event structure predictions and conflicts with this prior model, producing the late posterior positivity/P600.

Keywords: semantics, factive verbs, presuppositions, ERP, P600, N400

1 Introduction

During discourse comprehension, we are able to use certain linguistic cues to infer that certain events and states are true, regardless of whether their truth is central to the main message being conveyed. In theoretical semantics, such inferences are known as *presuppositions*, and they are triggered by particular words or phrases known as *presupposition triggers* (e.g., Karttunen 1973; Stalnaker 1974; Karttunen & Peters, 1979; Heim, 1982; for a detailed introduction of presupposition, see Chierchia & McConnell-Ginet, 2000; Beaver & Geurts, 2012; and Romoli & Sauerland, 2017). In the present study, we used event-related potentials (ERPs), a direct time-locked measure of brain activity, to study the effects of presuppositions that are triggered by a class of verbs — factive verbs — on downstream neural activity as comprehenders actively read and monitored the coherence of short discourse scenarios.

Factive verbs, such as “realize” and “know”, trigger the presupposition that the particular events or states conveyed by their sentential complements are true (Kiparsky & Kiparsky, 1971). For example, in the sentence, “John realized that the keys were on the shelf”, the factive verb, “realize” triggers the presupposition that the keys are, in fact, on the shelf. This presupposition holds even if this sentence is negated, e.g. “John did not realize that the keys were on the shelf”. This illustrates two important characteristics of presuppositions. First, they are inherited by more complex sentences — a behavior known as “presupposition projection” (e.g. Karttunen, 1973; Heim, 1983). Second, they are inferred even when they are not part of the communicator’s main point (not “at issue”, e.g., Simons, Tonhauser, Beaver & Roberts, 2011; Tonhauser, 2012): in the sentence examples above, the main point of the message is to convey John’s state of mind rather than the fact that the keys are actually on the shelf. Factive verbs can be contrasted with non-factive verbs such as “believe” and “assume”, which do not lead us to infer the truth of

information conveyed by their complements. For example, the sentence “John assumed that the keys are on the shelf” does not presuppose whether or not the keys are on the shelf.

Sentences that violate presuppositions are usually judged to be unacceptable (e.g. Amaral & Cummins, 2015; Cummins, Amaral, & Katsos 2013; Tiemann et al., 2011; Singh, Fedorenko, & Gibson, 2015; but see Bill, Romoli, Schwarz, & Crain, 2016). This is also true of sentences that violate the presuppositions induced by factive verbs (although most of these studies tested just one or two factive verbs — usually “regret”, “know” or “realize”, e.g., Chemla and Bott, 2013; Schulz, 2003; Tiemann et al., 2011). Such offline acceptability judgments, however, tell us little about the use of presuppositions during online comprehension.

Most psycholinguistic studies examining the use of presuppositions during online comprehension have used a violation paradigm to ask how quickly presuppositions are generated. In these studies, processing was measured at the presupposition trigger itself, and the preceding context was manipulated such that it was either consistent or inconsistent with the resulting presupposition. For example, Tiemann et al. (2015) measured reading times on the presupposition trigger, “again”, following contexts that either conveyed that a particular event had occurred (context consistent) or not (context inconsistent), e.g., “Last week, Linda bought Judith a pink lamp ... Two days ago, {(a) Judith / (b) Linda} received a pink lamp *again*...”). Self-paced reading and eye tracking studies show evidence of a relative slow-down on (or straight after) the region containing an inconsistent (versus an consistent) presupposition trigger (Clifton, 2013; Schwarz, 2007; Schwarz & Tiemann, 2017; Tiemann et al., 2015). In addition, work using the visual world paradigm suggests that comprehenders are able to detect a mismatch between prior contextual information and a presupposition trigger as early as 200ms following the trigger’s onset (Chambers & San Juan, 2008; Romoli, Khan, Snedeker, & Sudo, 2015;

Schwarz, 2015). Finally, a recent ERP study reported a larger positive-going waveform from 350-450ms following the presupposition trigger, “again”, when it was inconsistent versus consistent with information conveyed by the context (Jouravlev et al., 2016). This earlier positivity effect was followed by a later positivity effect between 450-750ms (see Discussion).¹

Taken together, these findings suggest that the brain can generate presuppositions very quickly as language unfolds in real time. However, they do not tell us about whether these presuppositions are encoded and maintained within the comprehender’s discourse model, and whether they are used to constrain expectations of subsequent information during comprehension. One way of addressing this question is to use factive verbs. Because such verbs precede their complements, it is possible to examine processing of downstream input that is either consistent or inconsistent with the presuppositions that they trigger. As we discuss next, there is some preliminary evidence that inputs violating these presuppositions are associated with costs in processing.

In an early eye-tracking study, Inhoff (1985) reported longer gaze durations, but no effects on first fixations, on complements communicating false common-knowledge facts following factive versus non-factive verbs (e.g., “He *knew/said* that two and two equaled three”). This was taken as evidence that factivity modulates the integration of incoming words into the discourse model (indexed by gaze duration), but that it has no effect on the initial lexical retrieval of these words (indexed by first fixations). However, given the well-established finding that it takes longer to process words in sentences that are incongruous (versus congruous) with

¹ Other studies have examined ERPs associated with the so-called uniqueness presupposition, triggered by definite noun-phrases (e.g., Heim, 1982). Rather than examine the consequences of violating this presupposition, these studies focused on the neurocognitive mechanisms engaged in its accommodation. We provide a brief overview of this group of studies in the Supplementary Material.

real-world knowledge (e.g. Marslen-Wilson, Brown, & Tyler, 1988), these findings are difficult to interpret.

In other work using ERPs, Ferretti and colleagues (Ferretti, Singer, & Patterson, 2008; Ferretti, Singer & Harwood, 2013) explored the effects of factivity as participants read short stories. These stories introduced a context that communicated a particular event, e.g. "...Ken and his brother ate some oranges/apples..." (sentence 2 in a story), which was later followed by either a factive or a non-factive verb and a complement that described an event that was either consistent or inconsistent with the initial event, e.g. "... the coach determined/figured that it was oranges that Ken ate..." (sentence 5). Critical nouns (e.g. "oranges") that followed both factive and non-factive verbs elicited a larger negativity between 300-500ms (N400) when they were inconsistent versus consistent with the event described in the preceding context. However, this negativity effect continued into a later 600-1000ms window only when critical nouns followed factive verbs. Aspects of the design of this study, however, limit interpretation. Specifically, critical words that were consistent with the event described in the context were always repeated (e.g. "oranges" following "oranges"), whereas those that were inconsistent with the event described were not (e.g. "oranges" following "apples"). Since the amplitude of the N400 is attenuated by word repetition (e.g., Karayanidis, Andrews, Ward, & McConaghy, 1991; Rugg, 1985), even at long lags (e.g., Bentin & Peled, 1990; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991), this introduced a confound between the consistent and inconsistent conditions. In addition, the authors do not report the process by which they selected the factive and non-factive verbs, or how many time any specific verb was repeated. Notably, some examples given in their papers (e.g. "determine" in Ferretti et al., 2008, and "was certain" in Ferretti et al., 2013) suggest that some verbs may have been incorrectly assigned to the factive class (e.g. the

compound “is certain” is usually classified as non-factive based on linguistic diagnostics of presupposition).

Building on this previous work, the current study used ERPs to determine whether presuppositions triggered by factive verbs (i.e., that their complements are true) are encoded within the comprehender’s discourse model such that it restricts expectations of upcoming events/states only to those that are compatible with the presupposition. For example, consider the discourse context, “Calvin needed to meet with his team members in the conference room. He was aware that it was busy.”. Here, the factive verb, “aware”, triggers the presupposition that the conference room was actually busy. The question we asked is whether this presupposition is incorporated and maintained within the comprehender’s discourse model such it restricts predictions of upcoming events and states *only* to those that are compatible with a busy conference room, excluding the possibility of later encountering incompatible events/states (e.g. a vacant conference room).

To address this question, we measured ERPs on words that were either consistent or inconsistent with a prior event/state that was conveyed by the complement of either a factive or a non-factive verb (e.g. “Calvin needed to meet with his team members in the conference room. He *was aware that/presumed that* it was *unused/busy*. He checked and it was vacant...”). Thus, our study crossed Verb factivity (factive versus non-factive verbs) and Event consistency (an event or state that was consistent versus inconsistent with the information conveyed by the verb’s complement).

Our main focus was on the modulation of a late positive-going ERP component with a posterior scalp distribution, otherwise known as the P600.² The precise neurocognitive mechanisms that are indexed by the late posterior positivity/P600 are debated, and we consider this further in the Discussion section (4.1). At this stage, we note that this ERP component is produced when comprehenders constrain their discourse model such that it restricts predictions only to sets of events and states (referred to as ‘event structures’) that are compatible with this model, and new input violates these event structure predictions (Kuperberg, 2013 and Xiang & Kuperberg, 2015). At the very least, the late posterior positivity/P600 reflects the detection of a conflict between the new input and the prior mental model (see Kuperberg & Wlotko, 2018 for recent discussion; see also Kuperberg, 2007; Paczynski & Kuperberg, 2012; Kim & Osterhout, 2005; van de Meerendonk, et al., 2009).

It therefore follows that if the presuppositions triggered by factive verbs are incorporated into the discourse model, thereby restricting predictions only to upcoming events/states that are compatible with the presupposition, then event inconsistent incoming words should evoke a larger late posterior positivity/P600 than event consistent incoming words (e.g. in the example above, a larger late posterior positivity/P600 should be evoked by “vacant” following “...busy” than following “...unused”). No such modulation should be seen in non-factive scenarios because these do not trigger presuppositions. If, however, the presuppositions triggered by factive verbs are not encoded or maintained within the comprehender’s discourse model, then there should be no differences between the factive and non-factive scenarios in the effect of event consistency on the late posterior positivity/P600.

² Here we refer primarily to the late posterior positivity/P600 that is evoked by semantic violations (see Kuperberg, 2007 for a review). Some of the ideas discussed in this paper are also relevant to understanding the late posterior positivity/P600 evoked by words that are syntactically anomalous or dispreferred (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; 1993).

In addition to asking whether factive verbs lead comprehenders to restrict their predictions to particular event structures, we were also interested in whether they lead comprehenders to generate stronger predictions about specific individual events and their associated semantic features/properties. For example, does the presupposition that a conference room is busy, and the prediction of an event structure that excludes incompatible events/states (e.g. a vacant conference room) lead comprehenders to generate predictions about possible individual events/states that are compatible with busy conference rooms, along with their associated semantic properties and features (e.g. <noisy>, <lights on>, <full of people>)?

To address this question, we examined the N400 ERP component (Kutas and Hillyard 1984; Federmeier & Kutas, 1999; Kuperberg, 2016). 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) context. However, the N400 is not always or necessarily modulated by contextual consistency, particularly if schema-based relationships between content words are matched across conditions (see Kuperberg, 2016 and references therein). This is because the N400 does not directly reflect the difficulty of computing a coherent higher-order representation of meaning. Rather, it is thought to reflect the degree to which the semantic features of an incoming word match those that were pre-activated by the preceding context (see Supplementary Materials and Kuperberg, 2016, for further discussion). Most relevant to the present study, contextual cues that lead comprehenders to restrict their expectations to a particular event structure do not always or necessarily enhance the pre-activation of semantic features associated with specific events. When they do, both an N400 and a late posterior positivity/P600 effect are observed on contextually inconsistent (versus consistent) critical words (e.g. Experiment 1: Xiang & Kuperberg, 2015). When they do not,

only a late posterior positivity/P600 effect, but no N400 effect, is observed on contextually inconsistent (versus consistent) critical words (see Kuperberg, 2007 for a review). Thus, in the present study, if factive verbs lead comprehenders not only to constrain their predictions to an upcoming event structure, but also to enhance their predictions of upcoming specific events and their associated semantic features, then event inconsistent (versus event consistent) words in the factive scenarios should produce not only a larger late posterior positivity/P600 but also a larger N400 effect than in the non-factive scenarios. If, however, factive verbs do not lead comprehenders to enhance their predictions of specific upcoming events/semantic features, then the effect of event consistency should not differ between the factive and the non-factive scenarios.

2 Methods

2.1 Development of stimuli

In order to develop and characterize the stimuli, we carried out a series of norming studies and used data collected from various linguistic databases, as described below. For the norming studies, participants were recruited through Amazon Mechanical Turk (AMT). Informed consent was obtained in all participants, and they were compensated for their time. Participants who had early exposure to a language other than English, and who had psychiatric or neurological illness, past neurological damage including stroke and concussion, or current treatment with psychoactive medication, were excluded. Prior to carrying out the norming tasks, participants completed a guided practice. In addition, “catch” questions were used to identify and exclude bots or inattentive participants.

2.1.1 Selection and classification of verbs: We began with an initial list of attitude verbs (e.g. “realized”) or verb compounds (e.g., “found out”). We then narrowed the list down to 91 verbs that we classified as either factive or non-factive, based on linguistic diagnostics of presupposition heritability (e.g., that the presupposition is maintained under negation). Then, to further assess the factivity of these verbs, we carried out a rating study. We embedded each verb into 16 different sentences that comprised a subject, the verb (with or without negation, in the past or present tense), a complementizer (“that”), and an ending that described an event or a state, e.g. “John realized that his wife was waiting in the car”. The same event/state was paired with four different verbs, with the exception of three events that were paired with three different verbs. The sentences were then counterbalanced across eight different lists so that different endings never appeared in the same list.

Seventy-two participants (nine per list) were instructed to rate how certain they were that the event being described in the sentence actually happened (or was actually happening) on a scale of 1 to 5. Based on these certainty judgments, we selected 27 verbs from sentences that were given ratings of above 3.4 and classified these as *factive verbs* (mean rating: 3.8; SD: 0.23), and 27 verbs from sentences that were given ratings of below 2.9 and classified these as *non-factive verbs* (mean rating: 2.8; SD: 0.17). There was no significant difference in the frequency of the factive and non-factive verbs ($t(44) = 0.61; p = .54$).³

2.1.2 Construction of experimental scenarios: Using the selected factive verbs and non-factive verbs, we constructed 152 quadruplets of three-sentence scenarios — four different versions of each scenario that corresponded to each of the experimental conditions. In all scenarios, the first

³ Only verbs, and not compounds (e.g., “is amazed”), were included in this analysis.

sentence introduced the characters and the general situation; the second sentence included a human subject followed by either a factive or a non-factive verb, the complementizer “that”, and an ending that described an event or a state (similar to those used in the sentence judgment task used to select the verbs, described above). Finally, a critical word in the third sentence was either consistent or inconsistent with the event or state that was described by the complement of the verb in the second sentence.

This resulted in a 2x2 design with Verb factivity (factive, non-factive) and Event consistency (event consistent, event inconsistent) as factors, creating four experimental conditions: (1) a factive verb in the second sentence with a critical word in the third sentence that was consistent with the event/state described (factive consistent); (2) a factive verb in the second sentence with a critical word in the third sentence that was inconsistent with the event/state described (factive inconsistent); (3) a non-factive verb in the second sentence with a critical word in the third sentence that was consistent with the event/state described (non-factive consistent), and (4) a non-factive verb in the second sentence with a critical word in the third sentence that was inconsistent with the event/state described (non-factive inconsistent). Because the third sentences, including the critical words, were identical in all versions of same scenario, the events/states described in the second sentence differed between the Event consistent and inconsistent conditions. Examples of scenarios, along with the full list of verbs, are given in Table 1.

2.1.3 Plausibility ratings: We gathered plausibility ratings for the initial set of experimental scenarios from 36 participants (9 per list), using AMT. The three-sentence scenarios were presented up until and including the critical word, with an ellipsis to indicate that the scenario could continue after this word. They were counterbalanced across four balanced and randomized

lists. Participants were told that they were seeing ‘beginnings of scenarios’ and were asked to rate how much the scenarios made sense to them on a scale of 1 to 5 (1 for scenarios that did not make sense at all and 5 for scenarios that made complete sense). Due to technical problems, we collected ratings for 144 (out of 152) scenarios.

Table 1. Example stimuli and full list of factive and non-factive verbs used in the study.

Factive		Non-factive	
Consistent	Inconsistent	Consistent	Inconsistent
Bruce taught a class on quantum physics. He saw that his students had <i>mastered</i> the material. Almost all of them scored <u>perfectly</u> on every test.	Bruce taught a class on quantum physics. He saw that his students were <i>confused</i> by the material. Almost all of them scored <u>perfectly</u> on every test.	Bruce taught a class on quantum physics. He was sure that his students had <i>mastered</i> the material. Almost all of them scored <u>perfectly</u> on every test.	Bruce taught a class on quantum physics. He was sure that his students were <i>confused</i> by the material. Almost all of them scored <u>perfectly</u> on every test.
Calvin needed to meet with his team members in the conference room. He was aware that it was <i>unused</i> . He checked and it was <u>vacant</u> and dark.	Calvin needed to meet with his team members in the conference room. He was aware that it was <i>busy</i> . He checked and it was <u>vacant</u> and dark.	Calvin needed to meet with his team members in the conference room. He presumed that it was <i>unused</i> . He checked and it was <u>vacant</u> and dark.	Calvin needed to meet with his team members in the conference room. He presumed that it was <i>busy</i> . He checked and it was <u>vacant</u> and dark.
Melanie and Billy were taking a walk together. She resented that he walked too <i>quickly</i> . She was always <u>behind</u> by several steps.	Melanie and Billy were taking a walk together. She resented that he walked too <i>slowly</i> . She was always <u>behind</u> by several steps.	Melanie and Billy were taking a walk together. She suggested that he walked too <i>quickly</i> . She was always <u>behind</u> by several steps.	Melanie and Billy were taking a walk together. She suggested that he walked too <i>slowly</i> . She was always <u>behind</u> by several steps.
Lila is very open with her emotions. Her boyfriend grasped that she was <i>upset</i> . She was <u>crying</u> all day long.	Lila is very open with her emotions. Her boyfriend grasped that she was <i>happy</i> . She was <u>crying</u> all day long.	Lila is very open with her emotions. Her boyfriend assumed that she was <i>upset</i> . She was <u>crying</u> all day long.	Lila is very open with her emotions. Her boyfriend assumed that she was <i>happy</i> . She was <u>crying</u> all day long.
Factive verbs		Non-factive verbs	
accept, acknowledge, care, discover, figure out, find out, forget, grasp, is amazed, is aware, is bothered, is informed, is shocked, is surprised, know, mind, notice, realize, recognize, regret, remember, resent, reveal, see, spot, take into account, take into consideration.		allege, assume, believe, claim, decide, estimate, expect, feel, figure, guess, hope, hypothesize, imagine, infer, is sure, postulate, predict, presume, reckon, sense, speculate, suggest, suppose, suspect, theorize, think, trust	

In the example stimuli, factive and non-factive verbs are marked in bold; the word within the context that determines whether the subsequent critical word will be consistent or inconsistent with the previous context is marked in italics, and the critical word itself (to which ERPs were time-locked) is underlined.

Table 2: Stimulus characteristics.

Scenario Type	Plausibility ratings#	Cloze*	Constraint**	LSA^
Factive consistent	4.38 [0.32]	22% [28%]	45% [22%]	0.062 [0.089]
Factive inconsistent	1.64 [0.43]	7% [16%]	35% [18%]	0.061 [0.089]
Non-factive consistent	4.32 [0.37]	26% [29%]	46% [22%]	0.059 [0.087]
Non-factive inconsistent	1.75 [0.49]	9% [20%]	37% [21%]	0.058 [0.088]

Means are shown with standard deviations in square parentheses.

#Plausibility ratings on a 5-point Likert scale up until and including the critical word.

*Cloze probability of the critical words was calculated as the proportions of participants in the norming study who responded with this word.

**Constraint of the critical words was calculated as the proportion of most common completion out of the total number of responses for each stem, regardless of whether or not it matched the critical word.

^LSA: Latent Semantic Analysis (Landauer & Dumais, 1997; Landauer et al., 1998). Semantic Similarity Values (SSVs) between the critical words and their preceding contexts are given.

The ratings are shown in Table 2. A 2 (Verb factivity) x 2 (Event consistency) ANOVA revealed a significant main effect of Event consistency ($F(1,143) = 4068.5, p < .0001$), confirming that, at the point of the critical word, participants did indeed rate the event inconsistent scenarios as more implausible than the event consistent scenarios. There was no main effect of Verb factivity ($F(1,154) = 0.76, p = .39$). As expected, there was a significant interaction between Verb factivity and Event consistency ($F(1,143) = 7.15, p = .008$), due to lower plausibility ratings in the inconsistent factive scenarios than in the inconsistent non-factive scenarios ($t(143) = 2.68, p = .008$). This finding reflects the assumption that factive verbs, unlike non-factive verbs, create an expectation of truth via their presupposition, and so inconsistent factive scenarios received lower plausibility ratings than inconsistent non-factive scenarios.

2.1.4 Cloze norming of experimental scenarios: We aimed to match the cloze probability of critical words between the event consistent factive and non-factive scenarios, and between the event inconsistent factive and non-factive scenarios. To do this, we collected cloze probabilities for the initial set of experimental scenarios from 40 participants (10 per list) using AMT. For each scenario, the critical word and all other words until the end of the third sentence were replaced with an ellipsis (e.g., “The court held a trial yesterday. At the end, the defendant realized that he won. The jury decided to...”). These scenario stems were then counterbalanced across four balanced and randomized lists. Participants were asked to read each scenario stem and then, in a free-response box, to type a single word that most likely followed the context. The cloze probabilities of the critical words in the four conditions were calculated as the proportions of participants who responded with that word. Then, in order to match cloze probabilities of critical words in the factive and non-factive scenarios within each level of Event consistency, we changed selected critical words and re-clozed the corresponding scenarios (40 participants) in order to come up with our final stimulus set.

Cloze values of the final stimulus set are shown in Table 2. A 2 (Verb factivity) x 2 (Event consistency) ANOVA revealed, as expected, a significant difference in the cloze probability of critical words in the consistent and inconsistent scenarios (a significant main effect of Event consistency, $F(1,151) = 83.14, p < .001$).⁴ There was also a marginal effect of Verb factivity ($F(1,151) = 3.46, p = .065$) due to very slightly lower cloze values of critical words in the factive than the non-factive scenarios. Importantly, there was no interaction between Verb factivity and Event consistency ($F(1,151) = 0.72, p > .39$).

⁴ A few of the event inconsistent items had greater-than-zero cloze probabilities because of constraining local (rather than global) contexts. We reanalyzed the ERP dataset after excluding these items (that is, including only zero-cloze event inconsistent items). The pattern of results was the same.

We also calculated the lexical constraint of each scenario stem: the proportion of most common completion out of the total number of responses for each stem, regardless of whether or not it matched the critical word, see Table 2. The average constraint of the contexts across all conditions was 0.4. Scenario stems in which the critical word was consistent were more lexically constraining than those in which the critical words were inconsistent (effect of Event consistency ($F(1,151) = 29.26, p < .001$)). There was no main effect of Verb factivity ($F(1,151) = 1.1, p < .29$) and no interaction between these two factors ($F(1, 151) = 0.11, p > .74$).

2.1.5 Latent Semantic Analysis (LSA): While our main focus was on the late posterior positivity/P600 component, we were also interested in whether the contexts in the different conditions would lead to differential prediction of specific events, thereby modulating the N400 on critical words. It was therefore important to rule out the possibility that the conditions differed in the degree to which they pre-activated more general schemas (see Myers & O'Brien, 1998; Gerrig & McKoon, 1998; see also Duffy, Henderson & Morris, 1989; Foss & Ross, 1983; Carroll & Slowiaczek, 1986; Morris & Folk, 1998), which can also modulate the N400 without necessarily entailing differential pre-activation of specific events (see Paczynski & Kuperberg, 2012, Kuperberg, Paczynski & Ditman, 2011, and Kuperberg, 2016 for recent discussion). We therefore used LSA, which can capture knowledge about multiple different semantic relationships between words and concepts, including schema-based relationships, to extract Semantic Similarity Values (SSVs) between the critical words and their preceding contexts in all four conditions (obtained from CU Boulder at lsa.colorado.edu: Landauer & Dumais, 1997; Landauer et al., 1998). We found no differences in SSVs across conditions (see Table 2), as

reflected by a 2 (Event confirmation) x 2 (Verb factivity) ANOVA, which revealed no main effects or interactions (all $F_s < 2.16$, $p_s > .14$).⁵

2.1.6 Setup of lists for the ERP experiment: The final set of experimental scenarios was divided into four lists, which were randomized and counterbalanced using a Latin Square design such that each participant would see only one version of each scenario. We added 70 coherent filler scenarios to each list in order to introduce some variation in the stimuli and to increase the rate of ‘acceptable’ judgments for the task used in the ERP experiment (as described below). Like the experimental scenarios, all fillers included three sentences and told a short story. However, unlike the experimental scenarios, the events/states introduced in the third sentences of the filler scenarios were new, continuing the scenarios (e.g., “Wesley spent the day painting his kitchen. His wife confirmed that the room looked much better. She wanted him to paint the bathroom as well.”). None of the experimental verbs were included in the filler scenarios in any position. Thus, each of the four lists comprised 152 experimental scenarios (with 38 scenarios per condition) and 70 filler scenarios. Participants were randomly assigned to one of the four lists (six participants per list).

⁵ Because of the way the stimuli were constructed, each pair of conditions differed by only a single word within the context (aside from the factive/non-factive verb itself). Thus, an alternative possibility was that, during comprehension, the lexical representation of this single word would linger in its relatively raw form, leading differential lexico-semantic ‘priming’ of the critical word across conditions (e.g. Forster, 1981; Norris, 1986; see also Fodor, 1983). For example, “unused” (in the event consistent scenarios) might be more likely than “busy” (in the event inconsistent scenarios) to prime the critical word, “vacant”. However, behavioral (Foss & Ross, 1983; Morris, 1994, Experiment 2; Traxler & Foss, 2000) and ERP (Camblin, Gordon, & Swaab, 2007; Coulson, Federmeier, Van Petten, & Kutas, 2005; Van Petten, 1993) studies suggest that this type of single word lexical priming plays a minimal role in sentence and discourse comprehension.

2.2 ERP Experiment

2.2.1 Participants: We report data from 24 individuals. Twenty-six individuals from the greater Boston area initially participated, but the datasets of two of the original participants were subsequently excluded due to excessive artifacts, leaving 24 datasets in the final analyses (10 females; mean age: 21.8 years; range: 18-35 years). All participants were right-handed (as assessed by the Edinburgh handedness inventory; Oldfield, 1971), native English speakers (having learned no other language before the age of 5), with no history of psychiatric or neurological disorders. All participants had normal or corrected-to-normal vision and no history of head trauma. They provided informed consent in accordance with the Institutional Review Board of Tufts University and were compensated for their time.

2.2.2 Experimental procedures: Participants sat in a quiet, dimly lit room facing a computer monitor. Stimuli were presented in white font centered on a black background. Each trial started with the word “READY” and participants pressed a button to initiate the trial. The first two sentences in each scenario were each presented as a whole; participants read them at their own pace, pressing a button to move on from the first to the second sentence to the next. After pressing a button following the second sentence, a fixation cross (‘+’) appeared in the middle of the screen for 800ms, followed by a blank screen for 200ms. Then, the last sentence of each scenario was presented word-by-word (each word presented for 400ms, followed by a black screen interstimulus interval of 200ms). The last word of the final sentence appeared with a period, and was followed by an 800ms black screen. After this, a question mark (‘?’) appeared, remaining on the screen until participants responded, at which point the next trial began.

Participants were asked to decide whether or not the scenario, as whole, made sense, and were instructed to give their responses by pressing one of two buttons (“yes” or “no”, counterbalanced across participants). They were instructed to wait until the “?” cue before responding, to reduce contamination of the ERP waveform by response sensitive ERP components.

2.2.3 EEG recording: A set of 29 tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH). Electrodes were also placed below the left eye and at the outer canthus of the right eye to monitor vertical and horizontal eye movements, and on the left and right mastoids. The electroencephalography (EEG) signal was referenced to the left mastoid online. The EEG signal was amplified by an Isolated Biometric Amplifier (SA Instrumentation Co., San Diego, CA) with a band pass of 0.01–40 Hz. It was continuously sampled at 200 Hz and the impedance was kept below 5k Ohm.

2.2.4 ERP analysis: ERPs were averaged offline at each electrode site in each experimental condition. Trials contaminated by eye artifact or amplifier blockage were excluded from analyses. Of the participants who were included in the final analysis, artifact contamination from eye movement or amplifier blocking led to the rejection of 6.7% of trials (on average). A 2 x 2 within-subjects ANOVA revealed no differences across the four experimental conditions in the rate of artifact rejection (no main effects of Verb factivity or Event confirmation or interactions between these two factors, $F_s < 0.73, p > .4$). For ERP analyses, we included all trials, regardless of participants’ end-of-sentence judgments, because these judgments may not have necessarily

reflected neural activity evoked at the point of the critical words themselves, and because we aimed to retain counterbalancing of our critical stimuli.

Here, we report the results of mass univariate analyses in which tests are carried at multiple time points and electrode sites and a permutation-based cluster mass test is used to account for multiple comparisons (Groppe, Urbach & Kutas, 2011; Luck, 2014). While this method has sometimes been viewed as a way of exploring uncharacterized ERP effects across the whole brain, recent simulations in our lab show that when such tests are carried out across broad spatial regions that correspond to components of interest, this approach does not sacrifice power to detect ERP effects (Fields, 2017a; Fields & Kuperberg, In Preparation), and it does not inflate Type 1 error rate.⁶ To carry out this analysis, we used the Mass Univariate ERP Toolbox (Groppe, Urbach & Kutas, 2011) and the Factorial Mass Univariate ERP Toolbox (Fields, 2017b) using a -100-0ms pre-stimulus baseline. To probe the late posterior positivity/P600, we carried out tests within a broad parietal-occipital spatial region of interest (CP1, CP2, P3, Pz, P4 O1, Oz, O2), which captures the full set of electrode sites where numerous previous studies have reported this effect. We carried out tests at all sampling points within two time windows: 500-700ms and 700-900ms to capture the earlier and later portions of the effect (following previous work that has emphasized that the late posterior positivity/P600 effect is not a monolithic component and that it is likely to reflect neurocognitive mechanisms with different latencies, see Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Gouvea, Phillips, Kazanina, &

⁶ In an earlier analysis of this dataset, we carried out an omnibus repeated-measures ANOVA, which we followed up with simple effects ANOVAs. In these ANOVAs, the dependent variable was the average ERP evoked across pre-specified temporal windows within several spatial ‘regions’ (groups of electrode sites) across the scalp. Region was included in the ANOVAs as a within-subject independent variable. However, as recently discussed by Luck and Gaspelin (2017), this approach creates multiple opportunities to detect effects in different spatial regions (potentially, a main effect, interactions with Region and additional effects at each region). It therefore increases the likelihood of Type I error. This is why we chose to report the results of a mass univariate analysis of our data here. Nonetheless, we report the results of our original analysis in Supplementary Materials for completeness. The pattern of findings was the same.

Poeppel, 2010, and see Discussion, section 4.1). To examine the N400 component, we carried out a similar analysis on centroparietal sites (C3, Cz, C4, CP1, CP2, P3, Pz, P4) between 300-500ms.

Within each of these broad spatial regions of interest, we carried out 2 x 2 repeated measures ANOVAs, which crossed Event consistency and Verb factivity (both within-participant factors) at all sampling points at each electrode site in each participant. Consecutive data points at electrodes within 8cm of one another (assuming a head diameter of 56cm) that exceeded a pre-set uncorrected p-value of 0.05 or less were considered clusters. The individual F-statistics within each cluster were summed to yield a cluster mass statistic. Next, we randomly re-assigned the values across the four conditions at each sampling point at all electrode sites within each participant, and calculated cluster-level statistics as described above. This was repeated 10,000 times. For each randomization, we took the largest cluster mass statistic, and in this way created a null distribution for the cluster mass statistic. Then we compared our observed cluster-level test statistic against this null distribution. Any clusters falling within the top 5% of the distribution were considered significant.

Any significant interaction between Event consistency and Verb factivity were followed up with planned repeated measures ANOVAs that directly compared ERPs to event consistent and inconsistent critical words in the factive and non-factive scenarios separately, once again using a mass univariate approach with similar parameters (see Fields, 2017b for discussion for why an F- rather than a t-test is more appropriate for this follow-up).

3 Results

3.1 Behavioral findings: We classified the event consistent scenarios in both verb categories as making sense (requiring a “yes” response), and the event inconsistent scenarios in both verb categories as not making sense (requiring a “no” response). We treat these classification as ‘correct’ responses in our analysis of accuracy. However, we note that these judgments are somewhat subjective, and because they were made at end of each sentence, they may not necessarily reflect neural activity at the point of the critical words. Overall, participants’ judgments matched our prior categorizations, with an average of 90% correct responses (Table 3).

A 2 x 2 repeated measures ANOVA revealed no significant difference in accuracy between the factive and non-factive scenarios ($F(1, 95) = 20.6, p = .09$). However, there was a significant main effect of Event consistency ($F(1, 95) = 20.6, p < .001$) due to greater accuracy in classifying the event consistent scenarios than the inconsistent scenarios. There was also a near-significant interaction between Event consistency and Verb factivity ($F(1, 95) = 20.6, p = .06$), which was driven by a greater tendency to judge the event inconsistent factive scenarios than the event inconsistent non-factive scenarios as not making sense ($t = 2.05, p = .05$), but no significant difference between the factive and the non-factive event consistent scenarios ($t = 0.59, p = .56$). These findings mirror the plausibility norms obtained at the point of the critical word itself.

Table 3. The mean percentage of responses that were classified as ‘correct’ based on our prior categorizations. In the event consistent scenarios, we coded “yes” responses as correct responses. In the event inconsistent scenarios, we coded “no” responses as correct responses.

Scenario Type	Mean correct responses
Factive consistent	93.0%
Factive inconsistent	86.7%
Non-factive consistent	93.5%
Non-factive inconsistent	82.3%

3.2 ERPs

Our primary question was whether the magnitude of the late posterior positivity/P600 effect on event inconsistent (versus event consistent) critical words would be larger in the factive than the non-factive scenarios.

3.2.1 Earlier portion of the late posterior positivity/P600 (500-700ms) We found evidence to support this hypothesis. Within the parietal-occipital region, between 500-700ms, there was a cluster that showed a significant interaction between Verb factivity and Event consistency ($p = .04$, Table 4). The cluster analysis revealed no significant main effect of either factor ($ps > .1$). To explore the source of the interaction, we examined the effect of Event consistency in the factive and non-factive scenarios separately within the same spatiotemporal region. In the factive scenarios, a cluster showed a main effect of Event consistency ($p = .01$), reflecting a larger late posterior positivity/P600 on event inconsistent than event consistent critical words (Figure 1, Table 4). In the non-factive scenarios, however, there were no clusters showing a significant effect of Event consistency.

3.2.2 Later portion of the late posterior positivity/P600 (700-900ms) Between 700-900ms, there was a cluster that showed a main effect of Event consistency ($p = .02$) due to a larger positivity on event inconsistent than event consistent critical words (Figure 1, Table 4). There were no clusters showing either a main effect of Verb factivity or an interaction between the two factors. Nonetheless, we carried out a post-hoc analysis examining the effect of Event consistency in the factive and non-factive scenarios separately in the same spatiotemporal region. In the factive

scenarios, a cluster showed a marginally significant main effect of Event consistency ($p = .06$), reflecting a larger late posterior positivity/P600 within this time window on event inconsistent than event consistent critical words. In the non-factive scenarios, there were no clusters showing a significant effect of Event consistency.

3.2.3 N400 (300-500ms)

A secondary question was whether we would also see a larger effect of event consistency on the N400 in the factive than in the non-factive scenarios. We found no evidence for this: while a marginally significant cluster was identified for the interaction between the two variables ($p = .08$), follow-up ANOVAs comparing the effects of Event consistency in the factive and non-factive scenarios separately failed to reveal any significant clusters (all $ps > .1$). Moreover, there were no significant clusters in the centroparietal region that showed either a main effect of Verb factivity or Event consistency (all $ps > .1$).

Table 4. Characterization of significant clusters in the Mass Univariate Analysis.

Time window	Effect	Cluster p -value	Spatial extent	Temporal extent	Spatial cluster mass peak	Temporal cluster mass peak
500-700ms	Verb factivity X Event consistency	0.041	CP1, CP2, P3, Pz, O1, Oz	550-620ms	P3	580ms
	Factive event consistent vs. factive event inconsistent	0.015	CP1, CP2, P3, Pz, P4, O1, Oz	530-615ms	Pz	600ms
700-900ms	Event consistency	0.023	CP1, CP2, P3, Pz, P4, O1	825-890ms	CP1	855ms
	Factive event consistent vs. factive event inconsistent	0.062	CP1, CP2, P3, Pz, O1	770-880ms	CP2	855ms

Spatial extent: All electrode sites in which at least one time-point was included in the cluster. *Temporal extent:* All time points in which at least one electrode site was included in the cluster. Note that mass univariate approaches are not designed to determine the beginning or the end of an effect and so these time windows underestimate the true extent of the effects. *Spatial cluster mass peak:* The electrode site with the largest spatial cluster mass. At each electrode site, all F values across all time points appearing at

that site were summed. The spatial cluster mass peak is the electrode site at which this summed F value is largest. *Temporal cluster mass peak*: The electrode site with the largest temporal cluster mass. At each time point, all F values across all electrode sites appearing at that time point were summed. The temporal cluster mass peak is the time point at which this summed F value is largest.

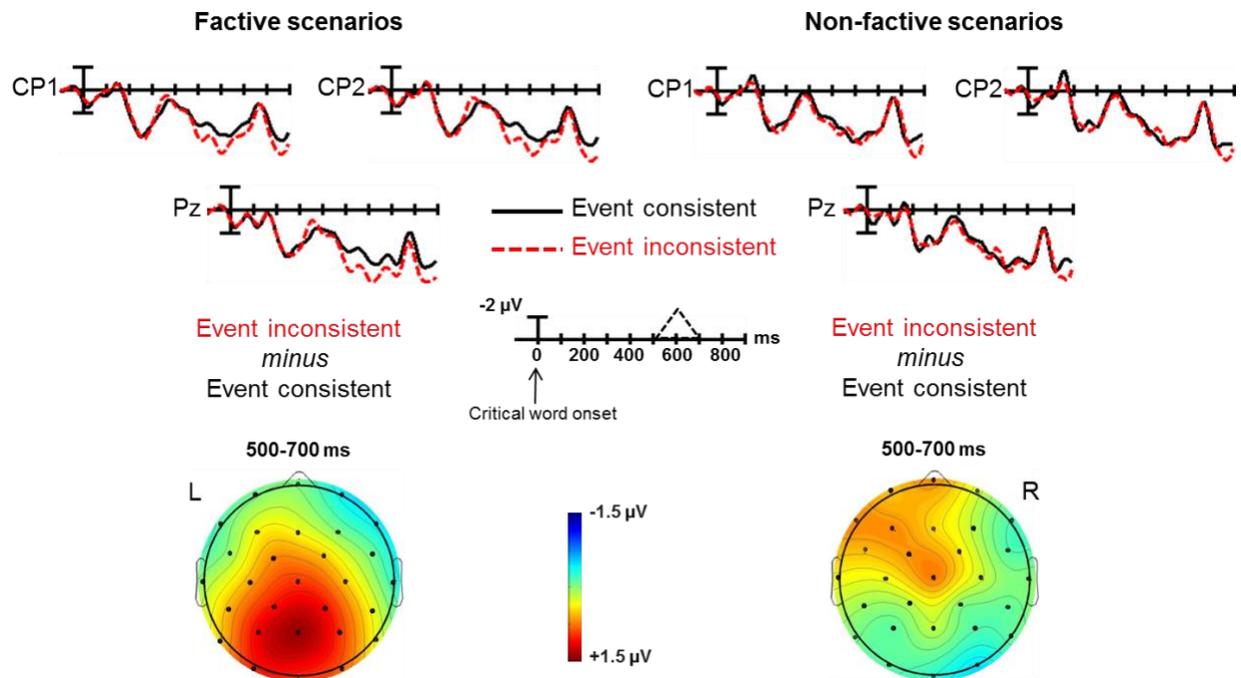


Figure 1. Grand-averaged waveforms evoked by event consistent and event inconsistent critical words in factive and non-factive scenarios. Waveforms evoked by critical words that were consistent with the event or state conveyed by the preceding context are shown with black solid lines; waveforms evoked by critical words that were inconsistent with the event or state conveyed by the preceding context are shown with red dashed lines. Voltage maps show differences between ERPs evoked by event inconsistent and event consistent critical words in the factive and non-factive scenarios in the earlier portion of the late positivity time window (500-700ms). Note that although the voltage map for the non-factive scenarios suggests that a larger late anteriorly distributed positivity was evoked by event inconsistent than event consistent critical words, a post-hoc analysis showed that this effect was not significant (see Supplementary Materials).

4 Discussion

In the linguistics literature, factive verbs are argued to trigger the presupposition that the events or states conveyed by their sentential complements are true (see Kiparsky & Kiparsky, 1971, and much subsequent work). In the present study, we asked whether, under task conditions that

encouraged coherence monitoring, these presuppositions are incorporated into the comprehender's discourse model, thereby influencing the pattern of neural activity evoked by later incoming information. To this end, we compared the effects of factive and non-factive verbs on processing incoming words that were either consistent or inconsistent with the events or states conveyed by the complements of these verbs (Table 1). Our behavioral data suggested that scenarios that were inconsistent with these prior events/states were rated as more implausible when they followed factive verbs than non-factive verbs. This was true both at the critical word (as reflected by our norming data) and at the end of each scenario (as reflected by the behavioral data collected during the ERP experiment). Our ERP data, which indexed neural activity that was time-locked to the onset of critical words, showed that between 500-700ms, the magnitude of the late posterior positivity/P600 effect evoked by critical words that were inconsistent (versus consistent) with prior events/states was larger in the factive than in the non-factive scenarios. We saw no differential modulation across conditions on the N400. Below, we discuss each of these findings in relation to the questions outlined in the Introduction and the previous literature, and outline open questions for future research.

4.1 The late posterior positivity/P600

We suggest that the larger late posterior positivity/P600 effect between 500-700ms produced by critical words in the factive inconsistent (versus factive consistent) scenarios was evoked by the violation of the presupposition triggered by factive verbs and incorporated into the comprehender's prior discourse model. For example, after reading the context, "Calvin needed to meet with his team members in the conference room. He was aware that it was busy.", comprehenders incorporated the presupposition that the conference room was busy into their

discourse model. They maintained this information over time and restricted their predictions to a particular event structure – a set of upcoming events/states that were compatible with this presupposition. New bottom-up evidence that the room was “vacant” thus violated this event structure prediction and conflicted with the prior discourse model. Comprehenders were therefore unable to initially incorporate the new input into the prior model (Xiang & Kuperberg, 2015 and Kuperberg & Wlotko, 2018 for recent discussion). In contrast, no significant modulation of the late posterior positivity/P600 in this time window was observed following the non-factive verbs (e.g. “He *presumed* that it was busy”) because these did not trigger presuppositions and so comprehenders did not restrict their predictions to a particular event structure.

This interpretation of the late posterior positivity/P600 as reflecting the detection of conflict between alternative representations is broadly in keeping with previous characterizations of this component (e.g. Kuperberg, 2007, van de Meerendonk, Kolk, Chwilla & Vissers, 2009; Kim & Osterhout, 2005; Paczynski & Kuperberg, 2012; see Kuperberg, 2013, Xiang & Kuperberg, 2015 and Kuperberg & Wlotko, 2018 for recent discussion). In previous studies, comprehenders constrained their mental models and predicted upcoming event structures based on strong contextual cues, such as the animacy-based selection restrictions of verbs (e.g. Paczynski & Kuperberg, 2011, 2012; Kuperberg & Wlotko, 2018), strong semantic attraction between verbs and arguments (e.g. Kolk, Chwilla, van Herten & Oor, 2003; Kim & Osterhout, 2005, Experiment 1), or particular discourse coherence markers (e.g. Xiang & Kuperberg, 2015, Experiment 1). Similarly, in the present study, factive verbs provided a strong cue for comprehenders to incorporate presuppositions about the truth of the information conveyed by

their complements into their discourse model, and to restrict predictions of future events/states to those that were compatible with the presuppositions.

Our findings are also consistent with those of another recent ERP study on presupposition: Jouravlev et al. (2016) reported a late widely distributed positivity effect that was evident between about 300-1000ms following the onset of the presupposition trigger “again” when it was inconsistent (versus consistent) with the event explicitly stated in the prior context (“Jake had never tipped a maid at the hotel before. Today he tipped a maid at the hotel *again ...”). In that study, the positivity effect began earlier than in our study. This may be because Jouravlev et al. measured ERPs on the presupposition trigger itself and so the violation of the presupposition became evident more quickly — as soon as the presupposition was computed. In contrast, in the present study, we measured ERPs well after the presupposition trigger and so conflict between discourse model and the new incompatible event could only be detected once comprehenders had integrated the incoming word into its local context to infer the new event, which may have taken a little longer.⁷ Importantly, our findings extend Jouravlev et al.’s study by showing that truth presuppositions, generated at an earlier point in the discourse, were incorporated and maintained within the discourse model such that they influenced downstream processing with neural consequences when they were violated by new inputs.

Although the late posterior positivity/P600 is triggered by violations of predicted event structures, the precise neurocognitive processes that it reflects remain unclear. As noted above, at the very least, it is likely to reflect the *detection of conflict* between the comprehender’s prior discourse model and the input, resulting in an initial failure to incorporate this word into this

⁷ An alternative possibility is that, in the present study, the positivity did in fact begin earlier than 500ms, but was obscured on the scalp surface by an earlier opposite-polarity N400 component. This component overlap may have been less likely to occur in Jouravlev et al. (2016)’s study because the N400 evoked by function words (like “again”) is smaller than to content words and because the N400 would have been reduced by multiple repetitions of the same word (“again”).

prior model and initially establish coherence (see Kuperberg & Wlotko, 2018 for recent discussion). It may additionally reflect *prolonged attempts to re-establish coherence* (cf. Kuperberg, 2007; Kuperberg, Caplan, Sitnikova, Eddy & Holcomb, 2006; Paczynski & Kuperberg, 2012). This may involve a reanalysis of preceding words in the context (e.g. the factive verb itself) to check whether they were accurately perceived (cf van de Meerendonk, et al., 2009) and/or a re-evaluation of the prior discourse model. Re-analysis and re-evaluation may sometimes lead comprehenders to successfully re-establish coherence by *revising/repairing* the prior model (see Kuperberg, 2013).⁸ For example, in the present study, encountering an event that is inconsistent with the predicted event structure may have led comprehenders to revise their prior discourse model by inferring that the protagonist's knowledge was unreliable (e.g. that Calvin's knowledge was based on false evidence) and abandoning the presupposition (that the conference room is actually busy). This, in turn, would have allowed them to incorporate the new input ("vacant") into the revised model and re-establish coherence.⁹

Finally, we distinguish between processes engaged when presuppositions are violated (reflected by the late posterior positivity/P600) and processes engaged in *accommodating*

⁸ Reanalysis, re-evaluation and/or revision may engage domain-general mechanisms, including those indexed by the well-known P3b ERP component, which has been linked to more general 'contextual updating' processes (Donchin & Coles, 1988) and is functionally related to the posterior late positivity/P600 (Coulson, King, & Kutas, 1998). These domain-general mechanisms may include a re-allocation of attentional resources (Sassenhagen, Schlesewsky & Bornkessel-Schlesewsky, 2014), the retrospective evaluation of the contents of working memory (Kuperberg, 2013), the retrieval of information stored in episodic memory (Van Petten & Luka, 2012).

⁹ This type of revision/repair may, in turn, be closely linked to *adaptation* over the course of the experiment, and it may have been reflected by the later part of the late positivity observed between 700-900ms, which had a more widespread scalp distribution (for a similar functional interpretation of this later portion of the late posterior positivity/P600, see Kuperberg, 2016, footnote 5, discussing results of a study reported by Chow, Smith, Lau & Phillips, 2015). It is also possible that participants adapted in this way to the non-factive scenarios. On this account, as the experiment progressed and comprehenders received increasing evidence that in 34% of scenarios the protagonists' knowledge (factive scenarios) or beliefs (non-factive scenarios) would be disconfirmed, they became increasingly more likely to re-evaluate the reliability of the protagonists' knowledge. Consistent with this idea, the main effect of Event consistency between 700-900ms was driven mainly by modulation in the second half of the experiment (as shown in a post-hoc analysis of this time window where no effect of Event consistency was found in the first half of the experiment, but a near-significant main effect of Event consistency was seen in the second half of the experiment).

presuppositions. When a new input *violates* a prior presupposition, it cannot be incorporated into the comprehender's discourse model unless she revises/repairs her prior model by abandoning the presupposition. In other cases, however, a new input may not immediately satisfy a presupposition, but it can still be incorporated into the discourse model without abandoning the presupposition. This is known as presupposition *accommodation*. Mechanisms of accommodation will vary, depending on the type of presupposition trigger and on the information provided by the context. They are therefore likely to be reflected by a variety of different ERP effects. Indeed, previous studies examining ERPs on definite noun-phrases, which trigger the so-called uniqueness presupposition — the inference that the NP refers back to a unique referent in the prior context (see Heim, 1982) — report modulation on several different ERP components (see Supplementary Materials for a discussion).

4.2 The N400

In addition to asking whether factive verbs restrict predictions to event structures that are compatible with the truth of the states/events described by their complements, we also asked whether they enhance predictions of upcoming semantic features. We did not find clear evidence for this: there was no significant interaction between Event consistency and Verb factivity on the N400 component. Indeed, we saw no effect of Event consistency on the N400 at all (this was also true when we examined the factive and non-factive scenarios separately).

The absence of an N400 effect to event inconsistent versus consistent critical words is not uncommon, particularly when broad schema-based semantic relationships are matched across conditions (for discussions, see Kuperberg, 2016, and Supplementary Materials). For the purpose of the question asked here, this pattern of results suggests that while factive verbs restricted

predictions to particular event structures (as reflected by the late posterior positivity/P600 effect when the event structure was violated), they did not additionally enhance the pre-activation of specific upcoming events and their semantic features (over and above those that were pre-activated by more general schema-based knowledge). This is in line with previous reports that a posterior late positivity/P600 effect but no N400 effect can sometimes be produced by contextually inconsistent versus consistent critical words (see Kuperberg, 2007, for a comprehensive review). It also distinguishes factive verbs from other types of contextual cues that *both* constrain predictions to upcoming event structures and enhance the pre-activation of upcoming semantic features (resulting in both an N400 and a late posterior positivity/P600 effect when these predictions are violated, e.g., Xiang & Kuperberg, 2015, Experiment 1).

4.3 Open Questions

Our findings raise important open questions. The first concerns the role of task. In the present study, participants were asked to actively monitor coherence by judging whether or not each scenario made sense. As we have previously discussed, task is one of several factors that can bias towards a late posterior positivity/P600, none of which is necessary or sufficient to produce the effect (Kuperberg, 2007, section 3.7). A late posterior positivity/P600 effect can be evoked in the absence of active coherence monitoring (e.g. Nakano, Saron & Swaab, 2010; Wang, Choi & Kuperberg, 2010; Nieuwland & Van Berkum, 2005). Indeed, the widespread positivity effect described by Jouravlev et al. (2016) on presupposition violations was seen when participants carried out a more passive comprehension task. Nonetheless, there is clear evidence that a requirement to monitor coherence enhances the magnitude of the posterior positivity/P600 effect (e.g., Xiang & Kuperberg, 2015). We have argued that this is because it both encourages

deep processing of the context and the construction of a rich discourse model, as well as the detection of conflict when the bottom-up input violates constraints of this discourse model.

Our assumption is that this type of active coherence monitoring is a component of deep and successful reading (and oral) discourse comprehension (for early evidence, see Wagoner, 1983; Garner, 1980; for more recent discussion, see Cain, 2016; Kim, 2014; van de Meerendonk, Kolk, Chwilla & Vissers, 2009). Without a goal of monitoring coherence, discourse processing can sometimes be dominated by passive memory-based processes (Myers & O'Brian, 1998; van den Broek & Helder, 2017), resulting at times in shallow interpretations. This is particularly likely to occur in an artificial experimental environment after reading hundreds of short unrelated discourse scenarios, including many incoherent scenarios. Indeed, under passive reading conditions, comprehenders can sometimes fail to detect anomalies at all (see Sanford et al. 2011 for evidence that the detection of incoherence can play an important role in triggering a late posterior positivity/P600 effect during passive reading comprehension).

On the other hand, with the type of binary coherence judgment task used in this study, we cannot tell whether the late posterior positivity/P600 *only* reflects the initial detection of conflict and the initial failure to incorporate the critical word into the prior discourse model, or whether it *also* reflects prolonged attempts to re-establish coherence through reanalysis/re-evaluation and revision/repair, as described above. One way of disentangling these possibilities in future work will be to explore how the posterior late positivity/P600 (and other ERP components) are modulated in tasks that not only require comprehenders to detect initial incoherence, but that also encourage them to actively repair and make sense of the scenarios.

A second open question concerns the relationship between presupposition and entailment. In the particular constructions used in the present study, the factive verbs did not

only trigger the *presupposition* that the particular events/states conveyed by their sentential complements were true; they also *entailed* that these events/states were true (e.g., Chierchia & McConnell-Ginet, 2000 among many others). Entailments are strong inferences that cannot be canceled by subsequent information. For example, in the sentence, “He was aware that [the conference room] was unused.”, the factive verb “aware that” not only induces a presupposition that “the conference room is unused”; it also entails that the room is unused: any attempt to cancel the presupposition results in an ill-formed sentence (e.g. “*He was aware that [the conference room] was unused because it wasn’t in fact unused”). It is therefore possible that, in the present study, the ERP response evoked by the violation of these presuppositions *also* reflected the violation of this entailment.

One way in which a future study might test this hypothesis is to include a set of negated factive stimuli, e.g. “He wasn’t aware that [the conference room] was unused”. Like its affirmative counterpart, this sentence also presupposes the truth of its complement (<the conference room is unused>). However, unlike its affirmative counterpart, it does not entail the truth of this complement: it is possible to cancel the presuppositions (e.g. the sentence, “He wasn’t aware that [the conference room] was unused because it wasn’t, in fact, unused” is not ill-formed). Thus, if violations of these negated factive scenarios also produce a late posterior positivity/P600 effect (under the same task demands), this would provide evidence that the effect can be produced by violations of presupposition in the absence of entailment.

4.4 Conclusion:

Our findings suggest that, during active discourse comprehension, factive verbs trigger presuppositions that are encoded and maintained within the comprehender’s discourse model.

When downstream input violates these presuppositions, it conflicts with this prior model, producing a larger late posterior positivity/P600. More generally, our results add to a growing body of evidence that the brain is able to quickly draw upon the rich semantic properties encoded in single words or phrases to compute and encode information within a mental model, even when such information is tangential (“not at issue”) to the central message being conveyed by the discourse.

Acknowledgements

This work was funded by the National Institute of Child Health and Human Development (R01 HD082527 G.R.K.), an Alon Fellowship (to ES) and the Israeli Science Foundation (1824/17 to ES), as well as by the Sidney R. Baer, Jr. Foundation, which supported undergraduate students, Arjun Dutta and Simon Riley, who contributed to stimuli creation. We are grateful to Ben Stillerman and Connie Choi who helped with data collection.

References

- Amaral, P., & Cummins, C. (2015). A cross-linguistic study on information backgrounding and presupposition projection. In *Experimental perspectives on presuppositions* (pp. 157-172): Springer.
- Beaver, D., & Geurts, B. (2012). Presupposition. *Maienborn, C.; Heusinger, K. von; Portner, P.(ed.), Semantics*, 2432-2460.
- Bentin, S., & Peled, B. (1990). The contribution of task-related factors to ERP repetition effects at short and long lags. *Memory & Cognition*, *18*(4), 359-366.
- Bill, C., Romoli, J., Schwarz, F., & Crain, S. (2016). Scalar Implicatures vs. Presuppositions—The view from Acquisition. *TOPOI*, *25*(1), 57-71.
- Cain, K. (2016). Reading comprehension development and difficulties: an overview. *Perspectives on Language and Literacy*, *42*(2), 9.
- Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*(1), 103-128.
- Carroll, P. S., M. L. . (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory and Cognition*, *14*(6), 509-522.
- Chambers, C. G., & Juan, V. S. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26-50.
- Chemla, E., & Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, *28*(3), 241-260.
- Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics*: MIT press.

- Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2015). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, 31(5), 577-596.
- Clifton Jr, C. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 487.
- Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 129-147. doi:10.1037/0278-7393.31.1.129
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21-58. doi:10.1080/016909698386582
- Cummins, C., Amaral, P., & Katsos, N. (2013). *Backgrounding and accommodation of presuppositions: an experimental approach*. Paper presented at the Proceedings of Sinn und Bedeutung.
- Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(3), 357-427. doi:10.1017/s0140525x00058027
- Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 791-801.
- 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

- 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. (2017). *Chapter 3: The problem of evidentiary value in the statistical analysis of ERP data and mass univariate statistics as a solution (Event-related potential and functional MRI studies of emotion and self-relevance)*. (Doctoral dissertation), Tufts University, Medford, MA.
- Fields, E. C. (2017). Factorial Mass Univariate ERP Toolbox [Computer software].
<https://github.com/ericcfields/FMUT/releases>: Github.
- Fields, E. C., & Kuperberg, G. R. (In Preparation). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data.
- Fodor, J. A. (1983). *The modularity of mind: an essay on faculty psychology*. Cambridge, M.A.: MIT Press.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology. A: Human Experimental Psychology*, 33(4), 465-495.
doi:10.1080/14640748108400804
- Foss, D. J., & Ross, J. R. (1983). Great expectations: Context effects during sentence processing. In G. B. Flores d'Arcais & R. J. Jarvella (Eds.), *The Process of Language Understanding* (pp. 169-191). Chichester: John Wiley & Sons.

- Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(2), 305-323.
- Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior, 12*(1), 55-63.
- Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes, 26*(2-3), 67-86.
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*(2), 149-188.
doi:10.1080/01690960902965951
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology, 48*(12), 1711-1725.
doi:10.1111/j.1469-8986.2011.01273.x
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439-483.
doi:10.1080/01690969308407585
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. (Doctor of Philosophy), University of Massachusetts, Amherst, MA.
- Heim, I. (1983). On the projection problem for presuppositions. In *Formal semantics: The essential readings* (pp. 249-260): Blackwell Publishers Ltd.
- Inhoff, A. W. (1985). The effect of factivity on lexical retrieval and postlexical processes during eye fixations in reading. *Journal of Psycholinguistic Research, 14*(1), 45-56.

- Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: an event-related potential study. *Language, Cognition and Neuroscience, 31*(10), 1245-1256.
- Karayanidis, F., Andrews, S., Ward, P. B., & McConaghy, N. (1991). Effects of Inter-Item Lag on Word Repetition: An Event-Related Potential Study. *Psychophysiology, 28*(3), 307-318.
- Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry, 4*(2), 169-193.
- Karttunen, L., & Peters, S. (1979). Conventional implicature. In C. K. Oh & D. A. Dinneen (Eds.), *Syntax and semantics 11: Presuppositions*. New York, NY: Academic Press.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225. doi:10.1016/J.Jml.2004.10.002
- Kim, Y., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49*(3), 269-281.
- Kiparsky, P., & Kiparsky, C. (1971). Fact. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: an Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (pp. 143-171). Cambridge.
- 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 and Language, 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. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components* (pp. 176-192). Baltimore, MD: Paul Brookes Publishing.
- 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., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489-530. doi:10.1080/01690960500094279
- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: an ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230-1246. doi:10.1162/jocn.2010.21452
- Kuperberg, G. R., & Wlotko, E. (2018). A Tale of Two Positivities (and the N400): Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Bioarxiv*. doi:10.1101/404780
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205. doi:10.1126/science.7350657
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163. doi:10.1038/307161a0
- 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. *Psychological Review*, *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 Processes*, 25(2-3), 259-284. doi:10.1080/01638539809545028
- Luck, S. J. (2014). *Online Chapter 13: The Mass Univariate Approach and Permutation Statistics (An introduction to the event-related potential technique)*. MIT Press Online: MIT press.
- Marslen-Wilson, W. D., Brown, C., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes*, 3, 1-17.
doi:10.1080/01690968808402079
- Morris, R., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory and Cognition*, 26, 1313-1322.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 92-103.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26(2&3), 131-157. doi:10.1080/01638539809545042
- 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., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691-701.

- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, 22(2), 93-136.
doi:10.1016/s0010-0277(86)90001-6
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113. doi:10.1016/0028-3932(71)90067-4
- Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785-806. doi:10.1016/0749-596x(92)90039-Z
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. In S. M. Garnsey (Ed.), *Language and Cognitive Processes. Special Issue: Event-related brain potentials in the study of language* (Vol. 8 (4), pp. 413-437). Hove: Lawrence Erlbaum Associates.
- Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes*, 26(9), 1402-1456. doi:10.1080/01690965.2011.580143
- 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
- Romoli, J., Khan, M., Sudo, Y., & Snedeker, J. (2015). Resolving temporary referential ambiguity using presupposed content. In *Experimental perspectives on presuppositions* (pp. 67-87): Springer.

- Romoli, J., & Sauerland, U. (2017). Presupposition and accommodation. In A. Barron, Y. Gu, & G. Steen (Eds.), *The Routledge Handbook of Pragmatics* (pp. 257-256). Milton Park, Oxon, UK: Routledge.
- Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22, 642-647.
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. . (2011). Anomalies at the borderline of awareness: an ERP study. *Journal of Cognitive Neuroscience*, 23, 514-523.
- Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29-39. doi:10.1016/j.bandl.2014.07.010
- Schulz, P. (2003). *Factivity: Its nature and acquisition* (Vol. 480): Walter de Gruyter.
- Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics*, 24(4), 373-416.
- Schwarz, F. (2015). Presuppositions vs. asserted content in online processing. In *Experimental perspectives on presuppositions* (pp. 89-108): Springer.
- Schwarz, F., & Tiemann, S. (2017). Presupposition projection in online processing. *Journal of Semantics*, 34(1), 61-106.
- Singh, R., Fedorenko, E., Mahowald, K., & Gibson, E. (2016). Accommodating presuppositions is inappropriate in implausible contexts. *Cognitive Science*, 40(3), 607-634.
- Stalnaker, R. (1974). Pragmatic presuppositions. In *Semantics and philosophy* (pp. 197-213). New York: New York University Press.
- Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2011). What projects and why. Paper presented at the Semantics and Linguistic Theory.

- Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., & Rolke, B. (2015). Presupposition processing and accommodation: An experiment on wieder ('again') and consequences for other triggers. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions* (Vol. 45, pp. 39-65).
- Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., . . . Beck, S. (2011). *Psycholinguistic evidence for presuppositions: On-line and off-line data*. Paper presented at the Proceedings of Sinn & Bedeutung.
- Tonhauser, J. (2012). Diagnosing (non-)at-issue content. *Proceedings of Semantics of Underrepresented Languages of the Americas*.
- Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(5), 1266-1282. doi:10.1037/0278-7393.26.5.1266
- 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 de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211-1224. doi:10.1111/j.1749-818X.2009.00163.x
- van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372.

- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. Special Issue: Event-related brain potentials in the study of language. *Language and Cognitive Processes*, 8, 485-531.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & Melsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3, 131-150.
- 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
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 328-346.
- Wang, S., Ditman, T., Choi, A., & Kuperberg, G. R. (2010). *The effects of task on processing real-world, animacy and syntactically violated sentences*. Paper presented at the Annual Meeting of the Cognitive Neuroscience Society, Montreal, QC.
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition, & Neuroscience*, 30(6), 648-672.
doi:10.1080/23273798.2014.995679