The Truth Before and After: Brain Potentials Reveal Automatic Activation of Event Knowledge during Sentence Comprehension

Mante S. Nieuwland

Abstract

■ How does knowledge of real-world events shape our understanding of incoming language? Do temporal terms like "before" and "after" impact the online recruitment of real-world event knowledge? These questions were addressed in two ERP experiments, wherein participants read sentences that started with "before" or "after" and contained a critical word that rendered each sentence true or false (e.g., "Before/After the global economic crisis, securing a mortgage was easy/harder"). The critical words were matched on predictability, rated truth value, and semantic relatedness to the words in the sentence. Regardless of whether participants explicitly verified the sentences or not, false-after-sentences elicited larger N400s than true-after-sentences, consistent with the well-established finding that semantic retrieval of concepts is facilitated when they are consistent with real-world knowledge. However, although the truth judgments did not differ between before- and after-sentences, no such sentence N400 truth value effect occurred in before-sentences, whereas false-before-sentences elicited an enhanced subsequent positive ERPs. The temporal term "before" itself elicited more negative ERPs at central electrode channels than "after." These patterns of results show that, irrespective of ultimate sentence truth value judgments, semantic retrieval of concepts is momentarily facilitated when they are consistent with the known event outcome compared to when they are not. However, this inappropriate facilitation incurs later processing costs as reflected in the subsequent positive ERP deflections. The results suggest that automatic activation of event knowledge can impede the incremental semantic processes required to establish sentence truth value. ■

INTRODUCTION

In many ways, history is marked as "before" and "after" Rosa Parks.

Rev. Jesse Jackson (2005)

One of the major challenges of ongoing language comprehension is to interpret the unfolding input in light of what we already know about the world. For any given proposition that we encounter during conversation or reading, the mapping of incoming linguistic representations onto our existing world knowledge determines whether the input makes any sense to us, whether it contains novel information deemed worth remembering, and whether or not we agree with it. An important source of world knowledge is our knowledge about specific events, how they unfold in time, and the changes that those events bring about (e.g., McRae & Matsuki, 2009; Zacks & Tversky, 2001; Zwaan & Radvansky, 1998)—in other words, knowledge of the state of affairs before an event and knowledge of how the event resulted in a different state of affairs after the event (e.g., Evans, 2013; Dowty, 1986). In language, people often refer to those states of affairs by using temporal terms like "before" and

"after" (e.g., Beaver & Condoravdi, 2003; Anscombe, 1964). This is exemplified by Rev. Jesse Jackson's tribute to Rosa Parks, whose arrest for refusing to give up her bus seat to a white passenger is associated with a sea change in the momentum of the civil rights movement.

The combined knowledge of the meaning of "before" or "after" with our knowledge of real-world events can therefore dictate whether we consider a given sentence to be true or false. For example, people with knowledge of the recent global economic crisis generally believe that securing a mortgage is harder, not easier, after the crisis compared to before. However, the question arises; given this knowledge about the outcome of the crisis, do people find it difficult to evaluate a sentence about the state of affairs before the crisis, compared to a sentence about after the crisis? The temporal terms "before" and "after" provide a direct test of how language comprehenders balance their real-world event knowledge with the value of the incoming input. The current study addressed this issue by examining electrical brain activity (N400 ERPs) evoked by critical words that render a sentence starting with "before" or "after" either true or false (e.g., "Before/After the global economic crisis, securing a mortgage was easy/harder"). Before presenting the rationale and predictions of the current study, I will review relevant experimental studies in separate sections on the activation of event

University of Edinburgh

knowledge during language comprehension, the online comprehension of "before" and "after," and on the impact of truth value on online sentence comprehension.

The Activation of Event Knowledge during Language Comprehension

A substantial body of literature suggests that people automatically activate world knowledge associated with narrated events (e.g., Metusalem et al., 2012; Bicknell, Elman, Hare, McRae, & Kutas, 2010; McRae & Matsuki, 2009; Ferretti, Kutas, & McRae, 2007; Gerrig & O'Brien, 2005; Zacks & Tversky, 2001; Zwaan & Radvansky, 1998; Graesser, Millis, & Zwaan, 1997; Kintsch, 1988). In general, recruitment of world knowledge will facilitate understanding, as it will activate concepts that are relevant to or implicitly present in the described event and may be mentioned in the unfolding discourse (e.g., Kutas & Federmeier, 2011; Altmann & Mirković, 2009; McRae & Matsuki, 2009). A recent demonstration comes from an ERP study by Metusalem et al. (2012), which tested the activation of event knowledge using the N400, a negative voltage deflection whose amplitude peaks approximately 400 msec post-stimulus and indexes the extent to which retrieval of semantic memory associated with a word is facilitated by the context (Kutas & Hillyard, 1980, 1984; for a review, see Kutas & Federmeier, 2011). In the Metusalem et al. study, participants read small discourse contexts (e.g., "A huge blizzard ripped through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big") that continued with a highly expected target word ("snowman"), an event-related implausible word ("jacket") or an event-unrelated implausible word ("towel"). Event-related words were not lexical associates of the target word, but the objects most often named (in a pretest) to be physically present in the described event. Both implausible words elicited enhanced N400 effects compared to the target words, but N400 amplitude was smaller for event-related words than for unrelated words. The authors did not rule out that event-related words were also considered less implausible than event-unrelated words, such that event-related words led to smaller semantic integration difficulties (e.g., for discussion, Federmeier & Kutas, 1999), yet these findings suggest that generalized event knowledge is activated during comprehension (see McRae & Matsuki, 2009), even at points at which this knowledge would constitute an anomalous continuation of the linguistic stream.

The above results could have interesting implications for how people understand sentences about a state of affairs before an event took place. In such sentences, concepts that are associated with the known outcome of an event may be automatically activated despite rendering a sentence false with respect to real-world knowledge. For example, a sentence about the financial crisis and the obtaining of mortgages activates event knowledge that the crisis made it more difficult for people to secure a mort-

gage. This current or "event outcome" knowledge can be said to compete for activation with representations about the state of affairs before the crisis (Cook & O'Brien, 2014; Altmann, 2013; Hindy, Solomon, Altmann, & Thompson-Schill, 2013; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Cook, 2005; Gerrig & O'Brien, 2005). Given that successful language comprehension involves the construction of an adequate mental representation of the described state of affairs (e.g., Zwaan & Radvansky, 1998; Kintsch, 1988), the activation of concepts that are consistent with outcome knowledge ("harder" activated by "Before the global economic crisis, securing a mortgage was") is "inappropriate." Such inappropriately activated concepts may even act as a "lure" when they render a sentence false. Because the activation of event knowledge is in accordance with a sentence that describes the outcome of an event ("After the global economic crisis, securing a mortgage was harder"), a difference between comprehension of sentences with "before" and "after" could emerge naturally from the way in which people activate world knowledge to comprehend narrated events. Whereas many studies have examined the ways in which people understand "before" and "after," however, the impact of these temporal terms on the online recruitment of real-world event knowledge during comprehension is unknown.

The Comprehension of "Before" and "After"

Extant research on comprehension on "before" and "after" focuses on how people use these terms to establish the chronological order of events (e.g., "Before [Event2], [Event1]"). Results from behavioral studies suggest that people are slower and less accurate when reading sentences that start with "before" compared to "after" (Mandler, 1986). Such results have been explained as reflecting people's default expectation that narrated events occur in chronological order, called the iconicity assumption (Zwaan, Madden, & Stanfield, 2001), which is flaunted by sentence-initial "before." Moreover, a classic study by Münte, Schiltz, and Kutas (1998) suggested that "before" incurs immediate and long-lasting processing costs compared to "after." Participants in their ERP study read sentences such as "Before/After the author submitted the paper, the journal changed its policy." The sentence-initial term "before" elicited a gradually increasing negativity at left-frontal electrodes compared to "after" that lasted throughout the sentence, reminiscent of ERP effects associated with sentences that are more demanding of working memory compared to sentences that are less demanding (e.g., King & Kutas, 1995). Accordingly, the ERP effect elicited by "before" and "after" sentences in the Münte et al. study was positively correlated with individual working memory capacity as measured through the Reading Span test (Just & Carpenter, 1992), ruling out that the observed ERP differences arose solely from comparing two different lexical items. These findings were

taken as further evidence that "before" requires readers to "mentally rearrange" the input to match temporal order (e.g., Ye, Kutas, et al., 2012; Ye, Milenkova, et al., 2012; Zwaan et al., 2001; Trosborg, 1982), via working memory operations. Münte et al. argued that these operations are initiated immediately upon reading "before" and have long-lasting effects on sentence comprehension.

Does this indeed mean that a sentence starting with "before" is always more difficult to understand than a sentence starting with "after"? An alternative explanation of the Münte et al. results could be that "after" is more difficult than "before," expressed by a greater positivity to "after" sentences (see also Hoeks, Stowe, & Wunderlink, 2004). Although this interpretation would not be compatible with well-established behavioral results (e.g., Mandler, 1986; Clark, 1971), it would be consistent with the patterns they observed: Compared to individuals with low working memory scores, individuals with high working memory scores showed more positive ERP responses to after-sentences, rather than more negative ERP responses to before-sentences, giving rise to the difference between the ERPs elicited by "before" and "after" being with correlated working memory capacity. That is, ERP responses to before-sentences did not visibly differ between individuals with low or high working memory capacity, which is not consistent with the interpretation that the effects arose because of the demanding nature of before-sentences. Another possible explanation for the ERP effect between "before" and "after" sentences is that the paraphrase task that participants performed impacted processing of the temporal terms differently. The participants needed to explicitly track temporal relations between clauses to perform a paraphrase task (on subsequently presented sentences with different temporal connectives, different clause order, different position of the connective, or all three). It is possible that the observed ERP difference resulted from the demands of this particular task and might not show the same pattern without a task, as would be the case in natural language settings. The relation between task demands and the observed ERP difference could be addressed by examining the resulting behavioral data, which were not reported. If comprehension accuracy was indeed lower in before-sentences than in after-sentences and lower in individuals with low working memory capacity than in individuals with high working memory capacity, then the behavioral data would yield strong support for the interpretation that "before" imposes long-lasting demands on sentence comprehension. Further strong evidence would be to show that the ERP effect is predictive of comprehension accuracy.

Another related issue is whether "before" always leads to comprehension difficulties or whether it depends on the syntactic function or word category that "before" is assigned (as a preposition or as a conjunction). Children initially start understanding and using "before" as a preposition that relates events to a fixed time point (e.g., "be home before dinner"; Coker, 1978). Very little is known about how adults process the preposition "before" during language comprehension. Most studies on comprehension of "before," including Münte et al., have examined comprehension of "before" as a subordinate conjunction that connects two events. The potential difficulty with "before" could hinge on the description of arbitrarily connected events in which the mental model of the temporal order needs to be construed online. Mandler (1986) showed that before-sentences that describe causally connected events (e.g., "Before the hiker could find shelter, he got soaking wet") do not incur costs, at least not as measured in whole-sentence reading times.

A final point is that differences between "before" and "after" may also arise from their respective truth conditions as specified in semantic theory (for discussion, see Baggio, Van Lambalgen, & Hagoort, 2012, 2015; Beaver & Condoravdi, 2003; Lascarides & Oberlander, 1993). The term "before" may lead to increased processing load because it may herald a counterfactual event (i.e., the event described in the before-clause need not actually have happened), whereas "after" always enforces a veridical reading (i.e., the event in the after-clause must have happened). Baggio and colleagues therefore argued that the observed ERP difference at "before" need not reflect the mental rearrangement of two described events, but rather the uncertainty that accompanies "before."

It is an open question whether and why "before" always leads to sustained processing costs and possibly to reduced comprehension accuracy In the literature on discourse comprehension, people have been shown to routinely construct temporally accurate representations of the discourse (the temporal dimension of situation models; e.g., Zwaan & Radvansky, 1998) and that the meaning of "before" is incrementally used to successfully update the situation model. Findings from behavioral and neuroimaging research suggest that readers accurately represent temporal aspects of the discourse, even when temporal information is only implicitly available (e.g., Becker, Ferretti, & Madden-Lombardi, 2013; Therriault & Raney, 2007; Claus & Kelter, 2006; Ferstl, Rinck, & Cramon, 2005; Rinck, Gámez, Díaz, & De Vega, 2003; Rinck, Hähnel, & Becker, 2001). This usually plays out as an increased processing cost for information that is inconsistent with the temporal structure of a described event. For example, story inconsistencies involving "before" and "after" lead to similar processing costs as other types of story inconsistencies (e.g., emotional or spatial; Ferstl et al., 2005). Some of these studies involved whole-sentence reading times or BOLD-fMRI responses and lacked the temporal resolution to pick up on processing differences of short duration. However, several studies provided evidence for a strong version of situation model theory of language comprehension, in which readers apply a temporally accurate representation of the preceding discourse immediately when they understand incoming words (e.g., for a review, see Zwaan & Radvansky, 1998). For example, an event that is described as to be ongoing is

relatively active in working memory as compared to an event that is described as finished ("the girl was skating" vs. "the girl skated"). Therefore, if people successfully incorporate the meaning of "before" to update their temporal situation model, such that the situation model only represents what happened before the described event, "before" need not incur sustained difficulties with language comprehension.

In summary, the evidence for processing difficulties associated with before-sentences is mixed. The results of several studies suggest that before-sentences may be cognitively demanding (Ye, Kutas, et al., 2012; Ye, Milenkova, et al., 2012; Münte et al., 1998; Mandler, 1986). However, there are several indications that costs associated with "before" are limited to specific circumstances, such as when "before" is used as a conjunction that connects two causally unrelated clauses, and the immediate neural effects of "before" may rely on an explicit evaluation task. The Münte et al. results constitute the only available set of findings on the immediate and lasting impact of "before" on sentence processing. Previous research on temporal term therefore do not lead to clear predictions on how people comprehend sentences about commonly known real-world events such as the global economic crisis. To examine the impact of "before" on online comprehension more directly, the current study takes a different approach from previous studies by examining the downstream consequences as measured in online effects of propositional truth value. To put this approach in the appropriate context, I will provide a brief overview of ERP studies on the online impact of sentence truth value.

Sentence Truth Value N400 Effects

The language comprehension system is thought to be highly incremental by relating incoming words to the widest interpretive background as early as possible (e.g., Altmann & Mirković, 2009; Van Berkum, 2009; Hagoort & Van Berkum, 2007). A well-known demonstration of incrementality is the effect of high-level, real-world knowledge on the N400. Regardless of whether participants are explicitly evaluating sentences or not, words that render a sentence true elicit reduced N400s compared to words that render a sentence false (e.g., Nieuwland, 2013, in press; Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008; Van Berkum 2009; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Fischler, Bloom, Childers, Roucos, & Perry, 1983). Observed "sentence truth value N400 effects" may not directly reflect the online computation on truth value but seem to reflect people's use of real-world knowledge to generate expectancies about upcoming words. This interpretation is based on observations that N400 amplitude is not so much a direct function of propositional plausibility or truth value but instead a function to what extent the incoming word shares semantic features with information that people may be expecting to appear (e.g., Kutas & Federmeier, 2011). When an incoming word is consistent with these knowledge-based predictions, the semantic retrieval of relevant information is facilitated, leading to smaller N400s compared to words that are inconsistent with world knowledge (Nieuwland, in press; Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008; Hagoort et al., 2004).

Whereas sentence truth value N400 effects have often been observed for straightforward affirmative sentences (e.g., Nieuwland, 2013; Nieuwland & Martin, 2012; Hagoort et al., 2004; Kounios & Holcomb, 1992; Fischler et al., 1983), a substantial body of literature suggest that these effects do not always occur in sentences that contain negation operators or negative quantifiers such as "no" or "few" (e.g., Nieuwland, in press; Urbach & Kutas, 2010; Nieuwland & Kuperberg, 2008; Kounios & Holcomb, 1992; Fischler et al., 1983). In negative sentences like "A robin is not a tree," N400 amplitude is not reduced for true sentences compared to false sentences, and some older studies have reported N400 amplitude is not sensitive to negation at all (e.g., Kounios & Holcomb, 1992; Fischler et al., 1983). More recent findings, however, suggest that the often observed lack of truth value N400 effects in negative sentences arises from their pragmatically infelicitous or underinformative meaning (i.e., negating a proposition that makes no sense to begin with like "a bird is a tree"). In pragmatically meaningful negative sentences like "With proper equipment, scuba-diving is not dangerous," truth value N400 effects that are identical to effects in affirmative sentences are found (e.g., Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008).

A recent study on quantifier comprehension specifically linked N400 truth value effects in affirmative and negative sentences to predictive processing (Nieuwland, in press). In that study, participants read negative and positive quantifier sentences matched on offline predictability (cloze value) and on truth value (e.g., "Most/Few gardeners plant their flowers during the spring/winter for best results"). Whereas true-positive quantifier sentences elicited reduced N400s compared to false-positive quantifier sentences, no difference was observed between true-negative and false-negative quantifier sentences, which both elicited larger N400s than true-positive sentences. However, a single trial regression analysis revealed that the interaction between quantifier and truth value only occurred for low cloze sentences and that N400 truth value effects became more similar for positive and negative quantifier sentences with higher cloze values. The online impact of truth value thus depends on the incorporation of quantifier meaning into a knowledge-based prediction for upcoming words.

In summary, the reviewed ERP studies suggest that sentence truth value can impact the N400 ERP, even in what are considered complex sentences (i.e., sentences containing counterfactuals, negation, or quantifiers; e.g., Clark & Chase, 1972). Sentence truth value N400 effects can thus be employed as a tool to investigate whether people successfully and rapidly incorporate the meaning

of "before" and "after" during language comprehension and to index people's ability to generate online expectancies about upcoming information based on their real-world knowledge about event-induced changes.

The Present Study

The present study examined electrical brain activity (N400 ERPs) evoked by critical words that render a sentence starting with "Before" or "After" either true or false (e.g., Before/After the global economic crisis, securing a mortgage was easy/harder"). Pretests established that before- and after-sentences were associated with identical truth value ratings and with equally strong expectations for the true critical word (see Table 1; the pretests are described in the Methods section). The main hypotheses focused on the sentence truth value N400 effects elicited by the critical words. Of note, sentence truth value N400 effects are observed in advance of and without the principled need for explicit evaluation (e.g., Nieuwland, in press; Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008). Here, I examine truth value N400 effects both when participants engage in explicit verification and when they do not. Evidence for differences between before- and after-sentences can be considered stronger when this evidence is obtained despite explicit instruction to evaluate sentence truth value. In addition, effects that replicate across different instructions cannot solely be ascribed to strategic task effects.

Under a strong and fully incremental version of situation model theory (e.g., Zwaan & Radvansky, 1998), "before" and "after" are both used to construct a temporally specific, fully updated situation model, without activation of information that is inappropriate for the described time period. These temporally specific representations can then be used to generate appropriate online expectancies about upcoming words, leading to similar sentence truth value N400 effects in before- and after-sentences. However, there are two alternative hypotheses to consider in which the sentence truth value effects differ in before- and after-sentences, which are not mutually exclusive with regard to the observed N400 effects. First, if sentence-initial "before" incurs long-lasting effects on sentence comprehension (e.g., Münte et al., 1998) such that incremental semantic processes are impeded relative to after-sentences, this may result in less facilitation of semantic retrieval for true words in before-sentences than in after-sentences. This could lead an interaction pattern of smaller N400 truth value effects in before-sentences than in after-sentences, with critical words in true-before-sentences eliciting N400s more similar to those observed for false-after-sentences and false-before-sentences than to N400s observed for true-after-sentences. The second alternative prediction is based on the literature on event comprehension and involves a difference between understanding the description of a previous state of affairs that is currently no longer true ("before") and the description of a current state of affairs ("after"). Concepts that are consistent with outcome knowledge ("Before the global economic crisis, securing a mortgage was harder") may be "inappropriately" activated and thereby act as a "lure" when they render a sentence false. If event outcome representations indeed impact comprehension of before-sentences, a reduced truth value N400 effect is predicted. If facilitation of semantic retrieval for true words is weaker in before-sentences than in after-sentences, critical words in true-before-sentences elicit larger N400s than those in true-after-sentences. If facilitation of semantic retrieval of

Table 1. Example Sentences with Results from the Independent Cloze Value and Truth Value Pretests and Characteristics of the Critical Words

Condition	Example Sentences	Cloze Value (%)	Truth Value Pre-rating	Length in Letters	Log Frequency	Semantic Relatedness (LSA-SSV)
True-before	Before the global economic crisis, securing a mortgage was <u>easy</u>, commonly.	38.3 (23.8)	3.8 (0.3)	6.2 (1.9)	1.54 (0.77)	0.18 (0.07)
True-after	After the global economic crisis, securing a mortgage was <u>harder</u>, commonly.	36.9 (21.7)	3.8 (0.3)	6.2 (2.0)	1.56 (0.84)	0.18 (0.06)
False-before	Before the global economic crisis, securing a mortgage was <u>harder</u>, commonly.	0.0 (0.0)	2.2 (0.3)	6.2 (2.0)	1.56 (0.84)	0.18 (0.06)
False-after	After the global economic crisis, securing a mortgage was <u>easy</u>, commonly.	0.0 (0.1)	2.1 (0.3)	6.2 (1.9)	1.54 (0.77)	0.18 (0.07)

Number of words per sentence and position of the critical word was identical across all 120 sentences.

Standard deviations are given in parentheses. Critical words are underlined for expository purposes. For truth value preratings, 1 = *false*, 5 = *true*. Log frequency is based on the Celex corpus (celex.mpi.nl). Semantic relatedness is indexed with semantic similarity values obtained with latent semantic analysis (lsa.colorado.edu).

false words is stronger in before-sentences than in after-sentences, critical words in false-before-sentences elicit smaller N400s than those in false-after-sentences.

Under the first alternative hypothesis, but not under the second alternative hypothesis, "before" would also elicit the sustained ERP effects compared to "after" as reported by Münte et al. Therefore, an additional comparison was performed to examine the differential impact of the words "before" and "after" that may last throughout a sentence (Münte et al., 1998). Although there are important differences between the current study and the Münte et al. study, as outlined in the previous sections, the possibility exists that "before" elicits immediate processing costs compared to "after," as reflected in a left anterior negativity that lasts throughout the sentences and is dependent on working memory. Therefore, sentence ERPs that started at the temporal preposition and that lasted up to the critical words were computed. Following Münte et al., participants were also tested for Reading Span capacity to test for a relation between span size and the effect of "before."

The current study also examined the post-N400 window because reduced N400 effects can be accompanied by subsequent processing difficulty as reflected in enhanced positive ERPs. In particular, semantically anomalous words that are considered hard to detect because of their strong superficial relatedness to the described scenario (e.g., "Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year sentence was given to the victim") elicit a reduced N400 effect but enhanced subsequent positive deflections (Sanford, Leuthold, Bohan, & Sanford, 2011; Nieuwland & Van Berkum, 2005). Such effects could indicate that critical words that are not immediately detected as being anomalous nevertheless elicit a second, more elaborate interpretive process upon detection (see also Van Herten, Chwilla, & Kolk, 2006; for a review, see Brouwer, Fitz, & Hoeks, 2012).

METHODS

Participants

Sixty right-handed Edinburgh University students (21 men) between 19 and 35 years old gave written informed consent. All were native English speakers, and none had neurological or psychiatric disorders or participated in the pretests. The first half of the participants read the sentences under for explicit verification, whereas the second half did not, such that each instruction was tested using a sample size similar to relevant previous studies (Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008).

Development and Pretest of Materials

An initial 215 sentence quadruplets were constructed that ended with critical word pairs (predicates, nouns or verbs). One critical word rendered the before-sentence true and the after-sentence false, and the reverse for the other word. Each sentence contained 10 words in two clauses separated by a comma (the first clause always contained four or five words). The items covered a wide range of world knowledge topics that native English-speaking Edinburgh University students were assumed to be familiar with, as was assessed in two pretests: (1) In a cloze probability pretest, 28 participants completed one of two counter balanced lists with one version of each item truncated before the critical word. They were instructed to complete the sentence with the first sensible word coming to mind. Cloze value was computed as the percentage of participants who used the intended critical word. (2) In a truth value rating pretest, 40 different participants evaluated one of four counterbalanced full-sentence lists containing only one condition per quadruplet and decided whether each sentence was true (1 = *false*, 5 = *true*), skipping sentences that they could not evaluate.

Subsequently, 120 quadruplets with a varied cloze value were selected by excluding quadruplets with true sentences receiving average ratings below 3.2 or false sentences receiving ratings over 2.9. In this final set, the true-before and true-after conditions were matched on cloze probability and truth value ratings, as were the false-before and false-after conditions (see Table 1). Critical word pairs were also matched on length and lexical frequency. In addition, latent-semantic analysis was performed to assure that the critical words were equally semantically related to the context words (the LSA-SSV measure based on lexical co-occurrence; lsa.colorado.edu).

In the ERP experiment, critical words were presented with a right-attached comma and followed by one additional word. These words were mostly adverbs (e.g., "generally," "typically," "usually") and were chosen to be as neutral as possible with respect to the sentence context to minimize differences in the ratings from the pretest (without sentence-final word) and the ratings in the ERP experiment (with sentence-final words). The fact that these differences were indeed very minimal (details are provided below) suggests that there was little impact of the sentence-final word on sentence evaluation.

Four counterbalanced lists were created so that each sentence appeared in only one condition per list, but in all conditions equally often across lists. Within each list, items were pseudorandomly mixed with 220 filler sentences (128 of which were true) to limit succession of identical conditions while matching sentence conditions on average list position.

Procedure

Participants silently read sentences, presented word by word and centered on a computer monitor, while minimizing movement. Word duration was 300 msec, with an additional 300 msec for critical words (presented with the comma to mark the clause boundary) and for sentence-final words (presented with a full stop). Commas were inserted so that the clause boundary would be clear,

but to avoid having the critical word as the sentence-final word, as the N400 modulations may then be clouded by sentence wrap-up effects. The extended duration of the critical words with the comma was based on the fact that readers slow down at words that mark clause or sentence boundaries. All interword intervals were 200 msec. Following sentence-final words, a blank screen was presented for 1800 msec.

In the verification instruction, a response display followed, showing the response options 1-2-3-4-5 centered on the screen and "Strongly disagree" and "Strongly agree" below the 1 and 5, respectively. Participants were asked to respond as accurately as possible, using the right hand to press the response option on the keyboard and to take as much time as needed. Accuracy was stressed because the planned analyses used sentences to which participants correctly responded (1 or 2 for false sentences, 4 or 5 for true sentences). Upon the response, a fixation mark appeared, indicating the opportunity to start the next sentence by pressing the space bar. In truth value N400 analyses, only sentences where participants gave condition-consistent responses were included (true-before, $M = 26.6$, $SD = 0.39$; false-before, $M = 26.9$, $SD = 0.37$; true-after, $M = 25.9$, $SD = 0.55$; false-after, $M = 25.4$, $SD = 0.48$). More before-sentences were included than after-sentences ($F(1, 28) = 4.2$, $p = .05$), but this did not depend on sentence truth value. Analysis of the average responses per condition (true-before, $M = 4.70$, $SD = 0.15$; false-before, $M = 1.25$, $SD = 0.17$; true-after, $M = 4.75$, $SD = 0.15$; false-after, $M = 1.20$, $SD = 0.12$) revealed a different evaluation of truth value in before- and after-sentences ($F(1, 28) = 9.8$, $p < .005$, $\eta_p^2 = .259$, reflecting the fact that true-before-sentences received slightly lower agree responses than true-after-sentences (true-before minus true-after, $M = -0.05$, $SD = 0.02$, $F(1, 28) = 4.8$, $p = .005$, $\eta_p^2 = .253$), whereas the false-before-sentences did not receive significantly higher or lower disagree responses than false-after-sentences (false-before minus false-after, $M = 0.05$, $SD = 0.03$, $F(1, 28) = 2.8$, $p = .1$, $\eta_p^2 = .09$).

In the no-verification instruction, the postsentence blank screen was followed either by a fixation mark or by a yes/no world knowledge question to which participants answered by button-press (followed by a fixation mark). These questions were orthogonal to the experimental manipulation and were included to keep participants alert (e.g., "After sunrise each summer morning, city streets are rather dark, usually" question: Does the sun rise in the east?). At the fixation mark, participants self-paced on to the next sentence with the space bar.

Participants in both instruction conditions were given several short breaks throughout the experiment. Total time-on-task was approximately 60 min. After the ERP experiment, participants performed a computerized Reading Span test (Just & Carpenter, 1992), which tests the ability to retain sentence-final words in memory as participants read aloud sets of unrelated sentences (pseudo-randomized sets of two to six sentences). Participants read a total of 100 sentences, and Reading Span score was computed as the total number of words that were correctly recalled. For full description of this Reading Span task, see van den Noort, Bosch, Haverkort, and Hugdahl (2008) and Nieuwland and Van Berkum (2006).

EEG Recording and Data Processing

The EEG was recorded at a 512-Hz sampling rate using a BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands) with 64 EEG electrodes, two mastoid electrodes and four EOG electrodes, active electrode reference (common mode sense), and passive electrode ground. The EEG was re-referenced offline to the average of the left and right mastoid.

For the N400 analysis, data were filtered (0.05–30 Hz), segmented into epochs from -200 to 1000 msec, corrected for eye movements and blinks using independent component analysis, baseline-corrected using 100 msec preceding word onset, and automatically screened for remaining artifacts (maximal/minimal allowed amplitude within an epoch at $75/-75$ μV). Participants were excluded from analysis if more than 1/3 of trials were rejected due to artifacts or condition-inconsistent responses (verification-instruction) or due to artifacts only (no verification), which left 57 participants for the analysis (29 with verification instruction, 28 without; average number of trials, verification: true-before, $M = 25.3$, $SD = 0.62$; false-before, $M = 25.1$, $SD = 0.60$; true-after, $M = 24.0$, $SD = 0.61$; false-after, $M = 23.8$, $SD = 0.58$; no verification, true-before, $M = 26.7$, $SD = 0.63$; false-before, $M = 26.9$, $SD = 0.61$; true-after, $M = 26.5$, $SD = 0.62$; false-after, $M = 27.1$, $SD = 0.59$). In the verification instruction, more true/false before-trials ended up being included than true/false-after-trials (true-before minus true-after, $M = 1.24$, $SD = 0.61$, $p < .05$; false-before minus false-after, $M = 1.38$, $SD = 0.52$, $p < .05$), whereas no differences were found for the no-verification instruction (all $Fs < 1$, ns).

For the sentence analysis, data were filtered (0.019–5 Hz band-width filter), segmented into sentence epochs lasting from -300 to 4500 msec relative to onset of "before/after" (thus lasting until the onset of the critical words), corrected for eye movements and blinks, baseline-corrected using the 300 msec preceding word onset, and then automatically screened for remaining artifacts (maximal/minimal allowed amplitude within an epoch at $150/-150$ μV). Participants were excluded from analysis if more than 1/3 of trials were rejected due to artifacts, which left 50 participants for the sentence analysis (26 participants with verification instruction, 24 without). The number of included before- and after-trials did not differ (before-trials, $M = 50.8$, $SD = 0.83$, after-trials, $M = 51$,

$SD = 0.87$) and was the same in the two instruction conditions (both $Fs < 1$, ns).

Statistical Analysis

To test the impact of sentence truth value in before- and after-sentences, the average N400 amplitude per condition was computed in the 300–500 msec time window (Kutas & Hillyard, 1980, 1984). Positive ERP effects in the post-N400 time window were tested in the 600–800 msec time window (e.g., Van Petten & Luka, 2012; Van Herten et al., 2006). A distributional analysis was used that involved all 64 electrodes employed using electrode grouping into ROIs, identical to the grouping used in Nieuwland (2014). A graphical representation of the ROI electrode grouping is provided in Figure 1, and full description is given in Nieuwland (2014). This grouping was used to separate the analysis of medially located electrodes (LMFC/RMFC, LMCP/RMCP) where N400 modulations are usually stronger (Nieuwland & Martin, 2012; Kutas & Federmeier, 2011; Nieuwland & Kuperberg, 2008) from the laterally located electrodes (LAF/RAF, LLFC/RLFC, LLCP/RLCP, LPO/RPO), with both groupings allowing tests fo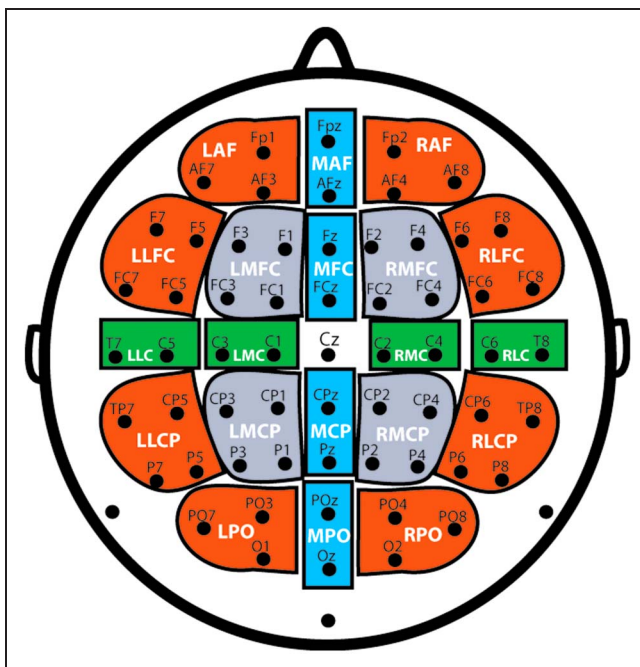r hemispheric differences and for anterior–posterior differences. Additional clusters were formed for the midline ROIs (MAF/MFC/MCP/MPO) and crossline ROIs (LLC/LMC/RMC/RLC).

Repeated-measures ANOVAs followed the 2 (Time: before, after) × 2 (Truth value: true, false) × 2 (Instruction: verification, no verification) design, with separate distributional factors for each ROI grouping. The medial, lateral, and crossline analysis each included a two-level factor (Hemisphere: left, right). The medial analysis included a two-level factor (Anteriority: frontal-central, central-parietal), whereas the lateral and midline analysis each included a four-level factor (Anteriority: anterior-frontal, frontal-central, central-parietal, parietal-occipital). Instruction was the only between-subject variable in all analyses. Where appropriate, Greenhouse–Geisser corrections and corrected F values are reported. Only statistical results with $p < .05$ are reported.

To test for the impact of temporal preposition on sentence processing, average amplitude per condition was computed in the 500–4500 msec time window, thus measuring brain activity up to the presentation of the critical words that rendered the sentence true or false. This time window collapses the two time windows used by Münte et al. (1998), given that before–after effects were found in both windows in the Münte et al. study. Repeated-measures ANOVAs followed the 2 (Time: before, after) × 2 (Instruction: verification, no verification) design, with the same distributional factors as listed above.

RESULTS

N400 Effects of Truth Value in Before-sentences and After-sentences

As shown in Figure 2, critical words in all conditions elicited a positive P2 component followed by a negative N400 component, with similar ERP waveforms for participants who explicitly verified the sentences and participants who did not. No effects were observed of task instruction in any of the N400 analyses.

The N400 medial analysis revealed a main effect of Truth value ($F(1, 55) = 5.3, p = .025, \eta_p^2 = .09$) and a Time × Truth value × Anteriority three-way interaction value ($F(1, 55) = 4.7, p = .034, \eta_p^2 = .08$), which was resolved by testing the Time × Truth value interaction at anterior electrodes and posterior electrodes separately. At anterior electrodes, no robust effects were observed. At posterior electrodes, however, a robust Time × Truth value interaction effect was found ($F(1, 54) = 4.9, p = .03, \eta_p^2 = .08$). Pairwise follow-up tests revealed that whereas a robust Truth value effect was observed for after-sentences (false-after minus true-after, $M = -1.2, SD = 0.40, p = .006, \eta_p^2 = .13$), no such effect was observed in before-sentences (false-before minus true-before, $M = -0.64, SD = 0.43, p = .14, \eta_p^2 = .04$). The N400s for false-after-sentences were marginally more negative than false-before-sentences ($M = -0.7, SD = 0.40, p = .08$,

Figure 1. Electrode configuration (black letters) and the ROI clusters that were used for statistical analysis (white letters). The last one or two letters refer to the anterior/posterior dimension: AF = anterior frontal; FC = frontocentral; C = central; CP = centroparietal; PO = parieto-occipital. The first letter of three-letter cluster names and the first two letters of four-letter cluster names refer to left–right dimension: L/R = left/right; LL/RL = left/right lateral; LM/RM = left/right medial. Medial ROIs are colored gray, lateral ROIs are colored red, midline ROIs are colored blue, and crossline ROIs are colored green.
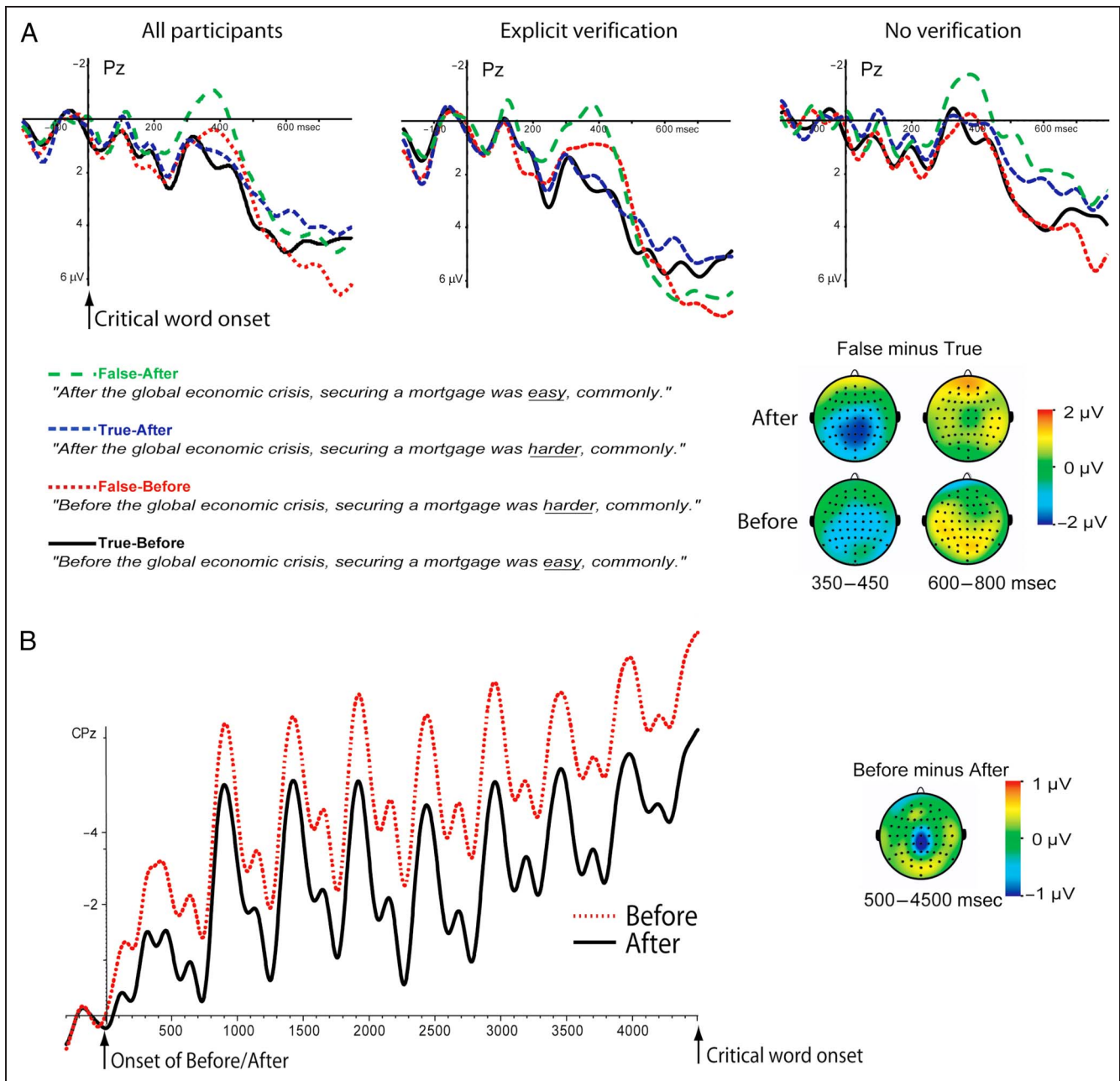
Figure 2. The graphs in (A) show the grand-averaged ERP waveforms elicited by critical words (CWs; underlined) in all four conditions at electrode locations Pz. The waveforms are filtered at 10 Hz for presentation purpose, and negative voltage is plotted upwards. Results are presented separately for all participants (left), participants who performed explicit verification (middle), and participants who did not perform explicit verification (right). Example stimuli are provided above the graphs. Scalp distributions of the relevant mean difference effects (false minus true sentences) in the 350–450 msec analysis window are given below the graphs. The graphs in (B) show the grand-averaged ERP waveforms elicited by the sentence-initial prepositions "Before" and "After," lasting throughout the sentence and ending at onset of the critical word. Scalp distributions of the relevant mean difference effect (before- minus after-sentences) in the 500–4500 msec analysis window is given on the right side of the graph.

$\eta_p^2 = .069$), whereas N400s for true-after-sentences did not substantially differ from those for true-before-sentences ($M = -0.18$, $SD = 0.63$, $p = .63$, $\eta_p^2 = .004$). Of note, because small differences in postsentence agree/disagree responses were found only in the true sentences, the difference in N400s for false-before- and false-after-sentences cannot be ascribed to differences in the strength with which participants disagreed with those sentences.

Out of parsimony, the results from the other ROI cluster analyses are not reported here, but they are available from the author. Whereas the crossline analysis only revealed a significant effect of Truth value, both the lateral analysis and the midline analysis revealed the same interaction effects as the medial analysis, with larger effects of Truth value in after-sentences than in before-sentences, observable at posterior channels where N400 effects are usually maximal. As in the medial analysis, the interaction

effects were driven by the differences between the false sentences.

Post-N400 Positive ERP Effects of Truth Value in Before-sentences and After-sentences

The medial analysis only revealed marginally significant effects, which are not reported here. The lateral analysis revealed a significant effect of Truth value ($F(1, 55) = 8.3$, $p < .01$, $\eta_p^2 = .132$) and a significant Time $\times$ Truth value $\times$ Anteriority three-way interaction effect ($F(3, 165) = 4.3$, $p < .05$, $\eta_p^2 = .072$). Follow-up tests revealed that, at the LAF/RAF ROI, false-after-sentences elicited more positive ERPs than true-after-sentences ($M = 1.1$, $SD = 0.50$, $p < .05$, $\eta_p^2 = .081$), whereas no differential effect occurred for before-sentences. In contrast, false-before-sentences elicited more positive ERPs than true-before-sentences at the LLCP/RLCP ROI ($M = 0.99$, $SD = 0.29$, $p = .001$, $\eta_p^2 = .179$) and the LPO/RPO ROI ($M = 1.1$, $SD = 0.34$, $p < .005$, $\eta_p^2 = .151$) while those ROIs showed no significant effects for after-sentences.

Sentence ERP Effects of the Temporal Prepositions

Sentence-initial "Before" elicited a sustained negative shift throughout the sentences compared to "after" at midline electrodes (Figure 2), whereas at some electrodes a positive effect was found (a figure with ERPs at all electrode locations is available from the author). No significant effects were observed in the medial, lateral, and crossline analysis. The midline analysis revealed a robust Time $\times$ Anteriority effect ($F(3.144) = 4.0$, $p < .05$, $\eta_p^2 = .077$), with follow-up tests showing that before-sentences elicited more negative ERPs at the MCP ROI ($M = -1.4$, $SD = 0.67$, $p < .05$, $\eta_p^2 = .083$). Inclusion of Reading Span score as covariate did not reveal robust effects.

DISCUSSION

This study investigated the impact of temporal terms like "before" and "after" on the online recruitment of real-world event knowledge by examining electrical brain activity (N400 ERPs) evoked by critical words that render a sentence starting with "Before" or "After" either true or false (e.g., "Before/After the global economic crisis, securing a mortgage was easy/harder"). False sentences elicited larger N400s than true sentences, reflecting the early semantic processing costs associated with false sentences, even when no explicit verification is required (Nieuwland, 2013, in press; Nieuwland & Martin, 2012; Nieuwland & Kuperberg, 2008; for behavioral findings, see Isberner & Richter, 2014; Singer, 2013; Rapp, 2008). Crucially, "before" was associated with a reduced N400 truth value effect compared to "after," which resulted from false-before-sentences eliciting smaller N400s than false-after-sentences, whereas true-before-

and true-after-sentences elicited similarly reduced N400s. An additional result was observed in the post-N400 time window, where false-before-sentences elicited an enhanced positive ERP effect compared to true-before-sentences, whereas no such effect was observed in after-sentences. In an examination of the sentence length ERPs elicited by the temporal terms "before" and "after" themselves, following Münte et al. (1998), "before" elicited more negative ERPs at central electrode channels than those elicited by "after."

Automatic Activation of Event Knowledge and Sentence Truth Value N400 Effects

The N400 results suggest that semantic retrieval of false words was facilitated in before-sentences but not in after-sentences. The automatic activation of event knowledge can thus impede the incremental semantic processes required to understand sentences starting with "before." Importantly, several variables known to impact N400 amplitude can be ruled out as confounding factors. Critical words in true sentences were equally predictable, and the critical words in false sentences were equally unpredictable. In addition, false-before-sentences were not simply considered "less false" than false-after-sentences, as evident in the same truth value judgments in the pre-rating test as well as the verification responses in the ERP experiment. Moreover, an interpretation in terms of simple lexical priming from context words also does not explain the current findings. Lower-level variables (e.g., lexical association or semantic relatedness) are known to have a stronger impact in incongruent sentences than in congruent sentences (e.g., Camblin, Gordon, & Swaab, 2007). Here, the critical word pairs were matched on semantic relatedness based on latent-semantic analysis.

The results of the current study testify to an asymmetric impact of event outcome representations on comprehension of sentences with "before" and "after." Irrespective of ultimate sentence truth value judgments, semantic retrieval of concepts is momentarily facilitated when they are consistent with the known event outcome compared to when they are not. However, this inappropriate facilitation incurs later processing costs as reflected in the subsequent positive ERP deflections. At face value, the results can be taken to mean that false-before-sentences are momentarily considered to be true, with detection of falsehood following thereafter. This conclusion is consistent with previous literature on "hard-to-detect" anomalies (Sanford et al., 2011; Nieuwland & Van Berkum, 2005), which are typically highly related to the previous context, and participants take longer to detect such anomalies than unrelated anomalies. Hard-to-detect anomalies are often overseen but, when detected, elicit attenuated N400s followed by positive ERP effects (Sanford et al., 2011). The attenuated N400s reflect the facilitated semantic retrieval of anomalous words, whereas the subsequent positive ERP effects are taken to reflect enhanced

monitoring processes that follow an erroneous initial interpretation (e.g., Van Petten & Luka, 2012). In the current study, similarly, the combination of the reduced N400 for false-before-sentences and the concomitant enhanced late positive ERP therefore suggest that participants had more difficulty to falsify before-sentences than after-sentences.

The current results are inconsistent with a strong and fully incremental version of situation model theory (e.g., Zwaan & Radvansky, 1998), in which readers construct a temporally accurate and fully updated representation of described events (e.g., Rinck et al., 2001). If a situation model is updated fully incrementally, concepts that are inappropriate with regard to the temporal structure of the described event should not be activated. The results therefore reflect a limit on full and incremental semantic processing of the temporal preposition "before," in as far as the meaning of "before" is not used as effectively to reduce activation of representations of the event outcome. One way to conceptualize this process is in terms of competition for activation between event outcome and initial state representations (Kukona, Altmann, & Kamide, 2014; Altmann, 2013; Hindy et al., 2012, 2013). Altmann and colleagues argue that whenever a narrated event involves different instantiation of the same object because of event changes (i.e., before and after an event), these different representations compete for activation. Applying this line of thinking to the current results, outcome representations may have a certain advantage in this competition process over initial state representations.

The inappropriate activation of knowledge is consistent with memory-based language processing theories (e.g., Cook & O'Brien, 2014; Cook, 2005; Gerrig & O'Brien, 2005). Memory-based theories posit that words initially activate prestored world knowledge and earlier concepts from the text and that the contents of active memory are subsequently integrated into the discourse context by inhibiting contextually irrelevant concepts. Importantly, because the initial stage is blind to contextual relevance or propositional truth value, world knowledge could hinder ongoing comprehension. Of note, the activation of event outcome knowledge is also broadly consistent with situation model theory (e.g., Zwaan & Radvansky, 1998). In the current materials, there was always a causal relation between the event described in the first clause and the outcome as described in the second clause. The online construction of a mental representation of the sentence is thus influenced by knowledge of the narrated event, in particular knowledge of the changes that the event caused. The result suggests that knowledge about current states of affairs is represented differently and is perhaps more prominent than knowledge of past states of affairs, which are no longer true. This aligns with previous research showing increased availability of concepts when relevant to a current event (e.g., "lunch" when someone is packing lunch) compared to an event that was completed earlier (e.g., someone had finished packing lunch earlier; e.g., Becker et al., 2013; Baggio, Van Lambalgen, & Hagoort, 2008; Ferretti et al., 2007).

A potential parallel can be drawn to the comprehension of counterfactual "what if" sentences, which require people to balance their factual knowledge about the world with their readiness to engage in suspension of disbelief (e.g., Searle, 1975). Research on online counterfactual comprehension suggests that readers maintain access to both counterfactual and factual interpretations, which can reduce processing sensitivity to anomalies in counterfactual sentences compared to factual sentences (e.g., Ferguson, 2012). Mental representations of what happened before an event can also be said to be counterfactual, if the event caused a change of state. Mental representations of the after-event situation can be said to be factual if that situation still holds true. The impact of automatically activated event knowledge on comprehension of before-sentences may be equivalent to the reported impact of factual knowledge on comprehension of counterfactual sentences.

Immediate and Sustained ERP Effects of "Before" and "After"

The current results do not suggest that before-sentences are generally more difficult than after-sentences. No N400 difference was obtained in true-before- and true-after-sentences. Moreover, there was no unambiguous indication that the initial parts of before-sentences were more cognitive demanding than after-sentences. The observed ERP effect elicited by "before," a sustained and central-posterior negativity compared to "after," had a different and much more limited scalp distribution than the effect reported by Münte et al. (1998). Moreover, the current ERP difference for the temporal terms was not modulated by working memory span. Hence, no direct evidence that "before" incurred immediate and long-lasting comprehension costs was found. There is no strong a priori reason to take the observed negative shift for "before" compared to "after" as evidence that one condition is more costly for processing than the other, also because the effect may have simply arisen from comparing two different lexical items. The discrepancy with the Münte et al. results is intriguing and could reflect the different syntactic uses of "before" (as preposition or as conjunction). Alternatively, "before" perhaps only incurs immediate costs when participants track temporal relations (Hoeks et al., 2004) or when it triggers a nonveridical interpretation (Baggio et al., 2012, 2015; Beaver & Condoravdi, 2003; Lascarides & Oberlander, 1993). In light of all the important differences between this study and the Münte et al. study, the current findings are not necessarily a failure to replicate the Münte et al. results but could be addressing a different linguistic phenomenon altogether.

Conclusion

The productive and combinatorial nature of human language enables us to talk and reason about events in the past, present, and future. Terms like "before" and "after" are immensely useful for expressing how a particular event changes one state of affairs into another. To establish that a given proposition correctly refers to before or after an event, we must compare that proposition with our real-world knowledge about the event. But does our knowledge about the outcome of the event impact our comprehension of a proposition referring to before the event? This brain potential study addressed this question by examining N400 effects of truth value elicited by sentences that started with "before" or "after" and contained a critical word that rendered each sentence true or false (e.g., "Before/After the global economic crisis, securing a mortgage was easy/harder"). Regardless of whether participants explicitly verified the sentences or not, false-after-sentences elicited larger N400s than true-after-sentences, consistent with the well-established finding that semantic retrieval of concepts is facilitated when they are consistent with real-world knowledge. However, although the truth judgments did not differ between before- and after-sentences, no such sentence N400 truth value effect occurred in before-sentences, whereas false-before-sentences elicited enhanced subsequent positive ERPs. Thus, irrespective of ultimate sentence truth value judgments, semantic retrieval of concepts is momentarily facilitated when they are consistent with the known event outcome compared to when they are not. However, this inappropriate facilitation incurs later processing costs as reflected in the subsequent positive ERP deflections. The results suggest that automatic activation of event knowledge can impede the incremental semantic processes required to establish that a sentence is true or false.

REFERENCES

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583–609.

Altmann, G. T. M. (2013). Anticipating the garden path: The horse raced past the barn ate the cake. In M. Sanz, I. Laka, & M. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis for linguistic structure* (pp. 111–130). Oxford: Oxford University Press.

Anscombe, G. E. M. (1964). Before and after. *Philosophical Review, 74,* 3–24.

Baggio, G., Van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language, 59,* 36–53.

Baggio, G., Van Lambalgen, M., & Hagoort, P. (2012). Language, linguistics and cognition. *Handbook of the Philosophy of Science, 14,* 325–355.

Baggio, G., Van Lambalgen, M., & Hagoort, P. (2015). Logic as Marr's computational level: Four case studies. *Topics in Cognitive Science, 7,* 287–298.

Beaver, D., & Condoravdi, C. (2003). A uniform analysis of 'before' and 'after'. In R. Young & Y. Zhou (Eds.), *Proceedings of SALT XIII* (pp. 37–54). CLC Publications, Cornell.

Becker, R. B., Ferretti, T. R., & Madden-Lombardi, C. J. (2013). Grammatical aspect, lexical aspect, and event duration constrain the availability of events in narratives. *Cognition, 129,* 212–220.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63,* 489–505.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446,* 127–143.

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,* 103–128.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior, 10,* 266–275.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472–517.

Claus, B., & Kelter, S. (2006). Comprehending narratives containing flashbacks: Evidence for temporally organized representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1031.

Coker, P. L. (1978). Syntactic and semantic factors in the acquisition of before and after. *Journal of Child Language, 5,* 261–277.

Cook, A. E. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes, 39,* 265–278.

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes, 51,* 26–49.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy, 9,* 37–61.

Evans, V. (2013). *Language and time: A cognitive linguistics approach.* New York: Cambridge University Press.

Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *Quarterly Journal of Experimental Psychology, 65,* 939–961.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 182.

Ferstl, E., Rinck, M., & Cramon, D. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience, 17,* 724–739.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 9,* 400–409.

Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes, 39,* 225–242.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163–189.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–441.

Hagoort, P., & Van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 362,* 801–811.

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object-state changes on event processing: Do objects compete with themselves? *Journal of Neuroscience, 32,* 5795–5803.

Hindy, N. C., Solomon, S. H., Altmann, G. T., & Thompson-Schill, S. L. (2015). A cortical network for the encoding of object change. *Cerebral Cortex, 25,* 884–894.

Hoeks, J. C., Stowe, L. A., & Wunderlink, C. (2004). Time is of the essence: Processing temporal connectives during reading. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 613–618). Mahwah, NJ: Erlbaum.

Isberner, M. B., & Richter, T. (2014). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes, 51,* 7–25.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General, 121,* 459.

Kukona, A., Altmann, G. T., & Kamide, Y. (2014). Knowing what, where, and when: Event comprehension in language processing. *Cognition, 133,* 25–31.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Lascarides, A., & Oberlander, J. (1993). Temporal connectives in a discourse context. In *Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics* (pp. 260–268). Utrecht: Association for Computational Linguistics.

Mandler, J. M. (1986). On the comprehension of temporal order. *Language and Cognitive Processes, 1,* 309–320.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass, 3,* 1417–1429.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66,* 545–567.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395,* 71–73.

Nieuwland, M. S. (2013). "If a lion could speak …": Online sensitivity to propositional truth value of unrealistic counterfactual sentences. *Journal of Memory and Language, 68,* 54–67.

Nieuwland, M. S. (2014). "*Who is he?*" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language, 76,* 1–28.

Nieuwland, M. S. (in press). Quantification, prediction and the online impact of sentence truth value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

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,* 1213–1218.

Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth value in counterfactual sentence comprehension. *Cognition, 122,* 102–109.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118,* 155–167.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary change deafness in discourse comprehension. *Cognitive Brain Research, 24,* 691–701.

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition, 36,* 688–701.

Rinck, M., Gámez, E., Díaz, J. M., & De Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory & Cognition, 31,* 77–86.

Rinck, M., Hähnel, A., & Becker, G. (2001). Using temporal information to construct, update, and retrieve situation models of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 67.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience, 23,* 514–523.

Searle, J. R. (1975). The logical status of fictional discourse. *New Literary History, 6,* 319–332.

Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science, 22,* 361–366.

Therriault, D. J., & Raney, G. E. (2007). Processing and representing temporal information in narrative text. *Discourse Processes, 43,* 173–200.

Trosborg, A. (1982). Children's comprehension of "before" and "after" reinvestigated. *Journal of Child Language, 9,* 381–402.

Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language, 63,* 158–179.

Van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke, UK: Palgrave Macmillan.

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the Reading Span test in different languages. *European Journal of Psychological Assessment, 24,* 35–42.

Van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18,* 1181–1197.

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

Ye, Z., Kutas, M., St George, M., Sereno, M. I., Ling, F., & Münte, T. F. (2012). Rearranging the world: Neural network supporting the processing of temporal connectives. *Neuroimage, 59,* 3662–3667.

Ye, Z., Milenkova, M., Mohammadi, B., Kollewe, K., Schrader, C., Dengler, R., et al. (2012). Impaired comprehension of temporal connectives in Parkinson's disease—A neuroimaging study. *Neuropsychologia, 50,* 1794–1800.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21.

Zwaan, R. A., Madden, C. J., & Stanfield, R. A. (2001). Time in narrative comprehension. In D. H. Schram & G. J. Steen (Eds.), *Psychology and Sociology of Literature* (pp. 71–86). Amsterdam: John Benjamins.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162.