Brain and Cognition 135 (2019) 103569

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension

Francesca Delogu*, Harm Brouwer, Matthew W. Crocker

Language Science and Technology, Campus C7, Saarland University, 66123 Saarbrücken, Germany

ARTICLE INFO

Keywords:
Discourse comprehension
Event-Related Potentials (ERPs)
N400
P600
Semantic integration

ABSTRACT

The functional interpretation of two salient language-sensitive ERP components – the N400 and the P600 – remains a matter of debate. Prominent alternative accounts link the N400 to processes related to lexical retrieval, semantic integration, or both, while the P600 has been associated with syntactic reanalysis or, alternatively, to semantic integration. The often overlapping predictions of these competing accounts in extant experimental designs, however, has meant that previous findings have failed to clearly decide among them. Here, we present an experiment that directly tests the competing hypotheses using a design that clearly teases apart the retrieval versus integration view of the N400, while also dissociating a syntactic reanalysis/reprocessing account of the P600 from semantic integration. Our findings provide support for an integrated functional interpretation according to which the N400 reflects context-sensitive lexical retrieval – but not integration – processes. While the observed P600 effects were not predicted by any account, we argue that they can be reconciled with the integration view, if spatio-temporal overlap of ERP components is taken into consideration.

1. Introduction

The N400 and the P600 are the two most salient components of the event-related potential (ERP) signal that are differentially modulated during incremental language comprehension. Yet, their functional interpretation is still a matter of debate. The tremendous number of ERP studies conducted in recent decades show a variety of findings, including monophasic N400 and P600 effects, but also bi-phasic N400/P600 patterns. Any viable theory or model of the electrophysiology of language understanding should be able to account for this full spectrum of data, while ideally providing an internally consistent functional interpretation of the two ERP components.

The N400 is a negative deflection peaking around 400 ms post stimulus-onset, with a centro-parietal distribution. It was first discovered in response to semantically anomalous words (Kutas & Hillyard, 1980), but its amplitude is sensitive to a variety of other semantic factors, including the degree to which a word is expected given the preceding context (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Lau, Phillips, & Poeppel, 2008), independent of whether this context is a single word, a sentence fragment, or a discourse (for a review, see Kutas & Federmeier, 2000, 2011).

The P600 is a positive deflection starting at about 500 ms post stimulus-onset and lasting several hundred milliseconds, with centro-parietal and sometimes frontal distributions. Increased P600 amplitudes have been observed in response to syntactic violations or syntactically unexpected words, as well as syntactically complex sentences (e.g. Hagoort, Brown, & Groothusen, 1993; Kaan, Harris, Gibson, & Holcomb, 2000; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994; Osterhout & Mobley, 1995). P600 effects have also been observed with sentences involving implausible thematic role assignments (e.g., Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005), introduction of new discourse referents (e.g., Burkhardt, 2006, 2007), irony (Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Henst, & Noveck, 2013), or other pragmatic factors (e.g., Delogu, Drenhaus, & Crocker, 2018; Hoeks, Stowe, Hendriks, & Brouwer, 2013, see also Hoeks & Brouwer, 2014 for a review).[1]

While there is wide consensus that these two components differentially index relevant aspects of cognitive processing, theories disagree with regard to the interpretation of their precise functional role in language comprehension. The N400 has been discussed under three main accounts: (1) the access/retrieval account (Brouwer, Fitz, & Hoeks, 2012; Kutas & Federmeier, 2000, 2011; Lau, Almeida, Hines, & Poeppel, 2009; Lau et al., 2008; van Berkum, 2009, 2010), (2) the

* Corresponding author.
E-mail address: delogu@coli.uni-saarland.de (F. Delogu).

[1] While positivities following the N400 time window have been labeled Post-N400 Positivity (PNP), Late Positive Complex (LPC), Syntactic Positive Shift (SPS), and P600, we will use the theory-neutral label "P600" to collectively refer to such effects.

https://doi.org/10.1016/j.bandc.2019.05.007
Received 31 October 2018; Received in revised form 20 May 2019; Accepted 21 May 2019

integration account (Brown & Hagoort, 1993, 2000; Hagoort, Hald, Bastiaansen, & Petersson, 2004), and (3) the "hybrid" account (Baggio & Hagoort, 2011; Lau, Namyst, Fogel, & Delgado, 2016; Nieuwland et al., in press).

According to the access/retrieval account, the N400 amplitude reflects the effort involved in retrieving from long-term memory conceptual knowledge associated with the eliciting word, which is influenced by the extent to which this knowledge is cued (or primed) by the preceding context. More specifically, lexical retrieval can be conceptualized as a function that maps the eliciting *word form* into a representation of *word meaning*, while taking into account the context in which it occurs (this stage of word processing has been called "pre-lexical" or "lexical" by Kutas & Federmeier, 2011). On this view, reduced N400 amplitudes reflect facilitated access of lexical information.

In contrast, the integration account takes the N400 to index the effort involved in integrating the *word meaning* of the eliciting word form with the preceding context, to produce an updated *utterance interpretation* (this has been termed "post-lexical" integration processing by Kutas & Federmeier, 2011), and "unification" by Hagoort, Baggio, & Willems, 2009). On this view, increased N400 amplitudes reflect increased integration difficulty.

Finally, the "hybrid" view takes the N400 to reflect aspects of both access and integration processes. More specifically, the N400 amplitude reflects both the effort involved in retrieving *word meaning* from *word form* and integrating it into the *utterance interpretation* (Baggio & Hagoort, 2011, refer to these operations as "pre-activation" & "unification", respectively).

The P600, on the other hand, has been initially associated with the cost of revising, repairing, or reanalysing an existing (morpho-)syntactic structure (e.g., Friederici, 1995; Hagoort, Brown, & Osterhout, 1999; Osterhout & Holcomb, 1992), or as an index of syntactic integration difficulty (e.g., Kaan et al., 2000; Kaan & Swaab, 2003). As discussed in more detail below, more recent findings showing P600 effects in response to semantic or pragmatic factors have put into question such purely syntactic interpretations of the P600, giving rise to alternative accounts in which the P600 is argued to index conflict monitoring/resolution (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Osterhout, 2005; Kos, Vosse, van den Brink, & Hagoort, 2010; Kuperberg, 2007; van Herten, Kolk, & Chwilla, 2005) or semantic integration processes (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer et al., 2012).

The ongoing uncertainty regarding the precise functional role of the N400 and P600 components can be attributed to often overlapping predictions of competing hypotheses. The retrieval versus integration and hybrid accounts of the N400 make similar predictions for many of the extant experimental designs: contextually unexpected words will typically be predicted to be both more difficult to retrieve and integrate. With regard to the P600, syntactic reanalysis/reprocessing accounts are subsumed by a semantic integration account, as any manipulation that impedes syntactic analysis will necessarily impede the construction of an utterance interpretation. For example, agreement violations such as "The spoilt child throw(s)…" (Hagoort et al., 1993) are not only syntactic violations but also induce difficulty in establishing a coherent utterance interpretation, as it is not clear whether the speaker's intended meaning involves reference to a single child or multiple children (see Brouwer et al., 2012, for further discussion).

In order to decide between the different hypotheses – particularly the linking of integration processes with both the N400 and the P600 – we should consider experimental designs in which competing accounts of the two components make diverging predictions in a single study. While several studies have sought to disentangle the integration from the retrieval view of the N400 (e.g., Brown & Hagoort, 1993; Lau et al., 2009), or a syntactic from a non-syntactic account of the P600 (e.g., van Herten et al., 2005), very few have investigated the functional interpretation of both components simultaneously (see Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011, for an example).

Studies investigating the so-called "Semantic P600"-effect, as elicited by semantic illusion constructions, go some way towards realizing such a design (see for a review Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer et al., 2012; Kuperberg, 2007). Here, words that are difficult to integrate do not modulate the N400 component, but rather elicit a P600 effect relative to their non-anomalous controls. For example, Hoeks et al. (2004) examined syntactically well-formed Dutch sentences in which the verb arguments appeared in a semantically anomalous order (e.g., "The javelin has the athletes thrown"). Relative to a control sentence (e.g., "The javelin was by the athletes thrown"), the final verb of the anomalous sentence elicited a P600 effect, but no N400 effect. Similar results have been found in several other related studies (e.g., Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nieuwland & van Berkum, 2005; van Herten, Chwilla, & Kolk, 2006; van Herten et al., 2005, see also Sanford, Leuthold, Bohan, & Sanford, 2011).

The "Semantic P600" findings fundamentally informed the Retrieval-Integration hypothesis (Brouwer et al., 2017, 2012). In this account, comprehension proceeds in biphasic N400/P600 cycles indexing the word-by-word retrieval–integration processes that are operative during incremental language comprehension. This model assumes that N400 amplitude is modulated by processes related to lexical retrieval, and explains the absence of N400 effects in semantic illusion sentences as resulting from contextually-cued retrieval mechanisms (the words "javelin" and "athletes" are semantically associated with the word "thrown", leading to facilitated retrieval of the lexical features of that word). If the N400 reflects contextually-driven, but non-combinatorial retrieval processes, then integrative processes are to be found in another ERP component, which, according to the authors, is the P600. More specifically, the P600 is assumed to be a family of late positivities – varying in amplitude, scalp distribution, latency, and duration – reflecting the different sub-processes that underlie "the word-by-word construction, reorganization, or updating of a mental representation of what is being communicated" (Brouwer et al., 2012). Examples of such integrative sub-processes include referent accommodation, thematic role revision, and pragmatic inferences. Brouwer and Hoeks (2013) further speculate that these different sub-processes might have different, potentially overlapping, subparts of the core generator of the P600 component family (relatedly, Gouvea, Phillips, Kazanina, & Poeppel, 2010, speculate that different properties of the P600 reflect different sub-processes underlying syntactic operations).

Others, however, have sought to reconcile the "Semantic P600" findings with the integration view of the N400 and a reanalysis/reprocessing perspective on the P600, using multi-stream processing architectures (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Osterhout, 2005; Kos et al., 2010; Kuperberg, 2007; van Herten et al., 2005). In these models, the N400 indexes a plausibility-driven, syntax-independent integration stream. The P600, in contrast, indexes reanalysis or reprocessing of the input that results from disagreement between this plausibility-driven stream and a syntax-driven algorithmic stream. These multi-stream models explain the absence of an N400 effect in semantic illusion sentences as resulting from the ease with which the syntax-independent stream can construct a plausible interpretation of the sentence on the basis of the meaning of its content words, indicating that comprehenders do not immediately notice the presence of a semantic violation (in the previous example, this plausible interpretation would result from combining the meanings of "javelin", "athletes" and "throw" into a meaning representation in which the athletes have thrown the javelin). This interpretation conflicts with the one produced by the syntax-driven stream, in which it is the javelin that throws the athletes, thereby triggering the re-processing of the input. Re-processing, in this case, is argued to involve additional effort in (re-) analysing the syntactic structure of the sentence in order to accommodate the alternative more plausible thematic structure. It is this process which is argued to result in the observed P600 effect (Kim &

Osterhout, 2005). Thus, crucially, in contrast to the Retrieval-Integration account, such multi-stream explanations of the "Semantic P600" findings typically hinge on the availability of a semantically attractive, but syntactically unlicensed interpretation (see Brouwer et al., 2012, for a review).

In order to test retrieval versus integration accounts of the N400, and semantic integration versus syntax-based accounts of the P600, we used stimuli in which the target words were presented in syntactically unambiguous and semantically plausible target sentences. Retrieval and integration difficulty were manipulated on the basis of discourse-driven world knowledge violations. That is, unlike most studies in which semantic illusion phenomena have been tested in one-sentence stimuli, our stimuli consisted of mini-discourses in which the first sentence set up the context and the second contained a target word where both retrieval and integration processes were measured (see Schumacher, 2011, 2014, for a similar design). Since all experimental conditions varied the context but used exactly the same target sentence, the local sentential context of the target word could have no influence on retrieval and integration processes.

This manipulation allowed us to examine retrieval and integration processes using discourse-driven implausibility rather than more severe anomalies involving implausible thematic role structures, often combined with animacy violations (see, e.g., "For breakfast the eggs would eat", Kuperberg et al., 2003, or "The doctor asks his assistant again who had called that early. The assistant responds that the hepatitis had called that early.", Schumacher, 2011). Secondly, no conflict between a plausibility- and a syntax-driven stream could be produced within either the context or the target sentence, as neither of their syntactic structures could be revised to accommodate a more plausible interpretation of the discourse. Our stimuli therefore eschewed both the relevance of multi-stream explanations of the N400, and any possible reanalysis/conflict based explanation of the P600. More generally, this design allows us to directly (1) tease apart effects of lexical retrieval from semantic integration in the N400, and (2) identify effects of semantic integration in the P600 above and beyond syntactic reanalysis/reprocessing effects.

In the critical condition (1b), we tested target words (e.g., "menu") that were contextually cued, thereby facilitating lexical retrieval, but difficult to integrate with the wider discourse. This critical condition was compared to a baseline condition (1a) and a further control condition (1c):

(1) a. *Baseline*
John entered the restaurant. Before long, he opened the menu and…
b. *Event related violation condition*
John left the restaurant. Before long, he opened the menu and…
c. *Event unrelated violation condition*
John entered the apartment. Before long, he opened the menu and…

In the baseline condition (1a), the context sentence described a character entering a location or starting an activity. The second sentence contained a target noun that was semantically associated with the location/activity introduced in the context (e.g., "restaurant"/"menu") and, together with the verb, described a plausible subsequent event. In the event related violation condition (1b), the verb of the context sentence was replaced by verbs such as "left" or "finished", so that that the target noun was still cued by the location/activity mentioned in the context, but the event described was now implausible in that context as it violated relevant world knowledge.

We also included a further control condition (1c) – the event unrelated violation condition – in which both retrieval and integration of the target word should be difficult. Specifically, the location/activity introduced in the context was altered in such a way that the target noun was no longer cued by the context ("menu" is weakly associated with "apartment"), but the event described was implausible in that context.

If the N400 elicited by the target word in the event related condition (1b) patterns with the baseline condition (1a), this would suggest that the N400 is modulated by recently retrieved conceptual knowledge associated with the eliciting word *restaurant* (recall the absence of N400 effects in "Semantic P600" studies discussed above). Furthermore, the absence of an N400 effect for event related targets relative to baseline would provide strong evidence against the integration and hybrid accounts, as it would indicate that the N400 is insensitive to the difficulty of integrating the target word into the implausible utterance interpretation. If, however, the N400 to event related targets (1b) patterns with the event unrelated condition (1c), this would indicate that the N400 is sensitive to integration difficulty, consistent with both the integration and the hybrid accounts of the N400. Conversely, such an outcome would challenge the access/retrieval accounts: While both *utterance interpretation* prior to the target word and recently retrieved *word meanings* of the preceding context are known to modulate retrieval difficulty (Chow, Lau, Wang, & Phillips, 2018), the recently retrieved conceptual knowledge associated with *restaurant* would predict at least some attenuation of the N400 to the target relative to the event unrelated condition.

To test whether or not the P600 reflects semantic integration (Brouwer et al., 2012) above and beyond syntactic reanalysis processes (e.g., Hagoort et al., 1993; Osterhout & Holcomb, 1992), our stimuli were syntactically well-formed and unambiguous such that no alternative analysis – grammatical or otherwise – of the critical condition was available that could make the discourse plausible. If the P600 exclusively reflects syntactic reanalysis processes, no P600 effect should be observed for (1b) compared to (1a). If, however, the P600 indexes more general semantic integration processes, the target noun in the event related violation condition (1b) should produce a P600 effect compared to the baseline condition (1a), indicating that its meaning was more difficult to integrate into the wider discourse. Moreover, just as in condition (1b), the semantic integration account of the P600 predicts a P600 effect for (1c) compared to (1a), whereas the syntactic reanalysis account predicts no effect.

2. Material and methods

2.1. Participants

Twenty-six students from the University of Saarland took part in the experiment. All were right-handed, native speakers of German, and had a normal or corrected-to-normal vision. All participants gave written informed consent and were paid for taking part in the experiment.

2.2. Materials

We constructed ninety German sentence-pairs in three conditions as illustrated in (1) (see Appendix A for the original versions in German). To ensure that the target nouns were lexically associated with the locations/activities mentioned in the context in the baseline as well as in the event related violation condition (1a-b), while weakly associated in the event unrelated violation condition (1c), we asked 20 participants to rate the semantic relatedness between the primes (the locations/activities mentioned in the context sentence) and the target nouns appearing in the final sentence on a 1 (not at all related) to 7 (strongly related) scale. Mean ratings for the prime-target pairs taken from conditions (1a-b) was 6.32 (SD = 0.53), while for those taken from condition (1c) was 1.56 (SD = 0.46).

In a plausibility judgment task with three counterbalanced lists, 30 independent participants rated the plausibility of each item in each condition on a 1–7 scale. Items were presented up to the target noun, to avoid plausibility ratings to be affected by sentential materials appearing after it. The baseline condition was judged to be more plausible (M = 6.28, SD = 0.53) than both the event related (M = 2.42, SD = 0.80) and the event unrelated violation conditions (M = 1.93, SD = 0.82). Pairwise comparisons with Bonferroni correction showed

significant differences between all conditions (all ps < .01).

We also estimated the cloze probability of the target word in all three conditions. Three counterbalanced lists were created with the sentence pairs presented up to and including the determiner preceding the target word (e.g., "Peter betrat das Restaurant. Wenig später öffnete er die …"). We collected responses from 10 participants per list. Notice that by including the determiner, the number of possible completions in all conditions was narrowed (as determiners in German are marked for gender), thereby increasing the chances of observing the target noun in the baseline but also in the event related condition. The cloze probability of the target word was .38 (SD = .33) in the baseline condition, .13 (SD = .19) in the event related condition, and .008 (SD = .04) in the event unrelated condition. Pairwise comparisons with Bonferroni correction showed a significant difference between all conditions (all ps < .001). The increased cloze probability of the event related targets compared to the event unrelated ones might partly be due to the fact that the target word in the event related condition was sometimes accommodated within a phrase that completed the fragment in a plausible way. For instance, one participant completed the German version of "Tom ended a campfire. Right away he piled up the …", with the phrase "verbliebene Holz" (remaining wood), where the target word "Holz" appears in a plausible continuation. The fact that the target words were related to the scenario described in the context and matched in gender with the determiner provided in the stem might have increased the likelihood of being produced. Indeed, in a second cloze test study, where target cloze probabilities were estimated for the discourse fragments excluding the determiner, we found that the cloze probability of the target word in the event related violation condition was considerably lower (Baseline: .24 (SD = .26), Event related violation: .05 (SD = .14); Event unrelated condition: .0 (SD = .0). Pairwise comparisons with Bonferroni correction showed significant differences between the baseline and the two violating conditions (ps < .0001) but no significant difference between the event related and the event unrelated conditions (p = .15). Importantly, in both cloze studies the probability of the target word in the baseline condition was significantly higher than in the event related violation condition. Therefore, if no N400 effect is observed between these conditions (as we hypothesize based on the N400-as-retrieval hypothesis), this will not be explainable in terms of word predictability as quantified by cloze probability, but only in terms of semantic association with the context.

Three counterbalanced lists were created so that each item appeared in each list in a different condition. The experimental items were intermixed with 90 filler passages created to counterbalance the number of primed and plausible items. The fillers included 30 plausible passages in which the verb argument in the final sentence was unrelated to the scenario introduced in the context (as judged by the authors) and the verb in the context was of the "leaving"/"finishing" type (so that participants could not use the verb in the context to predict the plausibility of the item), 30 plausible items in which the verb argument was semantically associated with the context, and 30 implausible items in which the verb argument was unrelated to the context.

2.3. Procedure

Participants were seated in a sound-proof, electro-magnetically shielded chamber. Stimuli were presented with the E-prime software (Psychology Software Tools, Inc.) in white font on a black background. After a short training session, all items were presented in pseudo-randomized order in 4 blocks, with breaks after each block. Each trial started with a screen prompting participants to press a button to start reading the passages. The context sentence appeared in its entirety until participants pressed a button. Then a fixation cross appeared for 750 ms, after which the target sentence was presented word-by-word in the center of the screen, for 350 ms plus 150 ms inter-stimulus interval. After each trial, participants judged the plausibility of the passages by pressing one of two buttons (yes - no) on a response box.

2.3.1. Electrophysiological recording and processing

The EEG was recorded by means of 26 active scalp electrodes placed according to the 10–20 system. The signal was referenced and digitized at a sampling rate of 500 Hz. Data were recorded using FCz as reference and AFz as ground. The horizontal electro-oculogram (EOG) was monitored with two electrodes placed at the outer canthi of each eye and the vertical EOG with two electrodes above and below the right eye. Electrode impedance was kept below 5 KΩ for all scalp electrode sites, and below 10 KΩ for the EOG electrodes. During recording no online filters were used.

2.3.2. Analyses

The EEG signal was band-pass filtered offline at a 0.015–30 Hz and re-referenced to the average of the left and right mastoid electrodes. Segments time-locked to the target nouns were extracted with an interval of 200 ms preceding and 1000 ms following the onset of the stimulus, and semi-automatically screened for ocular and muscular artifacts. This led to discarding five participants showing excessive artifacts. Baseline correction used the 200 ms interval preceding the onset of the stimulus.

In accordance with previous literature (Hoeks et al., 2004; Kim & Osterhout, 2005; Kutas & Federmeier, 2011; Nieuwland & van Berkum, 2005) and visual inspection of the data, we computed mean amplitudes for each condition and electrode in the 300–500 ms (N400) and in the 600–1000 ms (P600) time windows. In order to investigate the topographic distribution of the relevant effects, data from midline and lateral electrodes were treated separately. Data from midline sites included three electrodes (Fz, Cz, Pz). Data from lateral sites were grouped into four regions of interest (ROIs): left anterior (F3, FC1, FC5), right anterior (F4, FC2, FC6) left posterior (P3, CP1, CP5) right posterior (P4, CP2, CP6). Within each time window, ANOVAs were carried out first with Condition (baseline, event related violation, event unrelated violation) and electrode sites (15 levels) as repeated measure factors. For the topographic distribution, in addition to condition, the ANOVAs on midline sites included anterior-posterior (AP) distribution (anterior, central, posterior), while the ANOVAs over later sites included AP distribution (anterior, posterior) and Hemisphere (left, right) as within-subject factors. The Greenhouse-Geisser correction was applied to all ANOVAs with greater than one degree of freedom in the numerator. In such cases, the corrected p-value is reported. Generalized *eta*-squared (η^2G) is reported as a measure of effect size.

3. Results

3.1. Plausibility judgments

Participants judged the stimuli to be plausible at the following rates: Baseline, 95% (SD = 22); Event related violation, 32% (SD = 46); Event unrelated violation, 8% (SD = 28). Pairwise comparisons with Bonferroni correction revealed a significant difference between all conditions (all ps < .01). These results qualitatively mirror the plausibility ratings on a 1–7 Likert scale observed in the norming study. The event related violation condition was judged significantly more implausible than the baseline condition, but more plausible than the event unrelated violation condition. In particular, the relatively high percentage of items in the event related violation condition being judged as plausible may result from the binary nature of the task. We will return to this point in Section 4.

3.2. ERPs

Grand-average waveforms to target nouns in all three conditions are shown in Fig. 1. Visual inspection of the waveforms shows a difference between conditions in the N400 time window (300–500 ms) and in a later time window starting at around 700 ms and lasting until the end of the epoch.

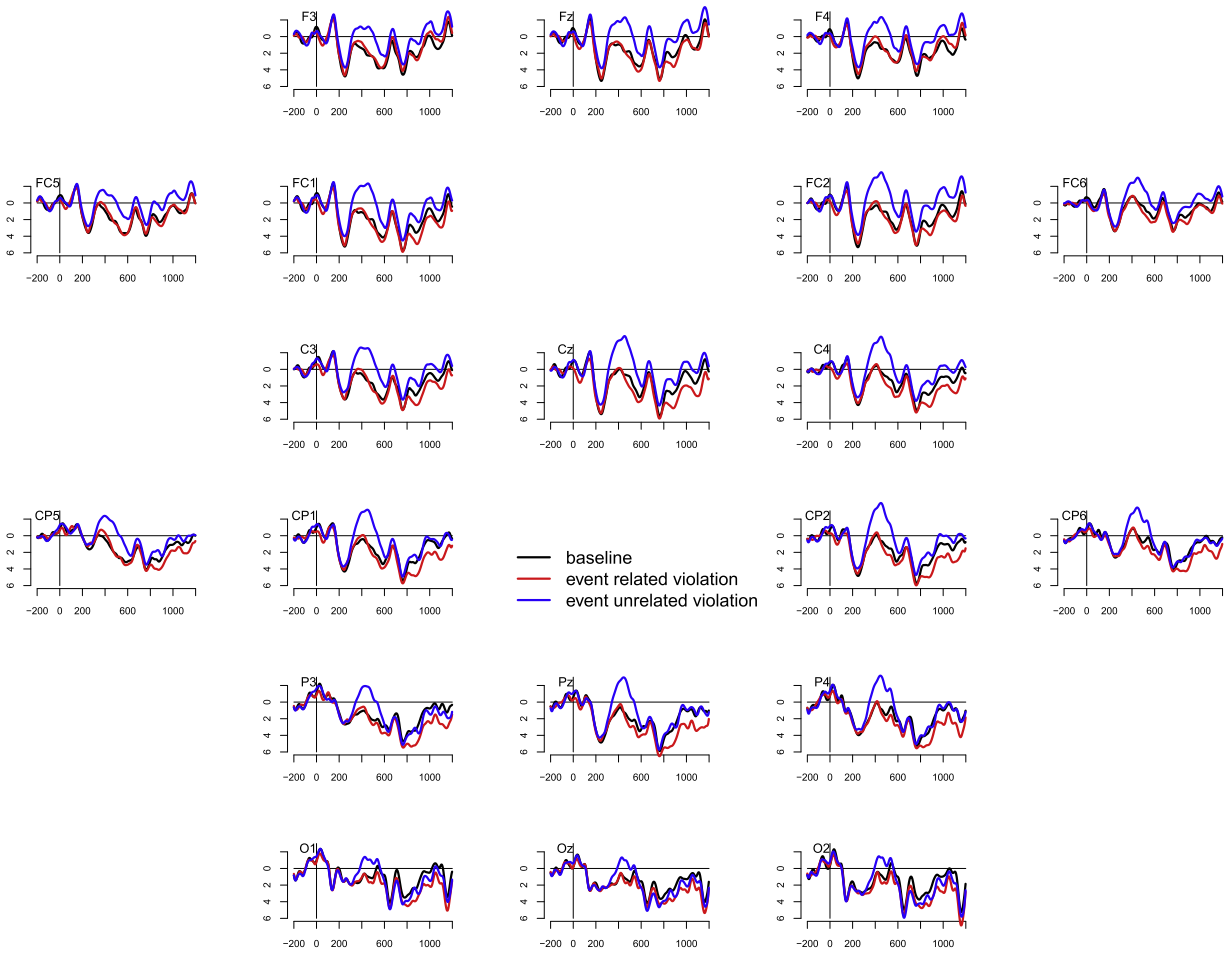

Fig. 1. Grand-average ERP waveforms time-locked to the target nouns in the baseline (black line), event related violation condition (red line) and event unrelated violation condition (blue line), for a subset of electrodes. Negative voltages are plotted upwards. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.1. N400 time window (300–500 ms)

The overall ANOVA yielded a significant effect of Condition, F(2, 40) = 22.32, p = < .0001, η^2G = .16, and a Condition-by-Electrode interaction, F(28, 560) = 2.39, p = .03, η^2G = .007. The ANOVA on midline electrodes revealed a significant effect of Condition, F(2, 40) = 22.8, p = < .0001, η^2G = .18, and no interaction with AP distribution (F < 1). Pairwise comparisons (see Table 1 and the topographic maps in Fig. 2) showed no differences between the event related violation condition (M = 1.17, SD = 3.24) and the baseline (M = 1.12, SD = 2.92), but a larger negativity elicited by the target nouns in the event unrelated violation condition (M = −1.91, SD = 2.66) compared to the baseline.

The ANOVA over lateral sites revealed an effect of Condition, F (2,40) = 21.91, p < .0001, η^2G = .17, and no interaction with AP distribution or Hemisphere (Fs < 1). Pairwise comparisons showed again no significant effects between the event related violation condition (M = 0.88, SD = 2.75) and the baseline (M = 0.99, SD = 2.39), but a larger negativity elicited by the target nouns in the event unrelated violation condition (M = −1.58, SD = 2.26) compared to the baseline.

To summarize, we found clear evidence that the N400 was only sensitive to contextual cues, with weakly associated targets in the event unrelated violation condition eliciting larger N400 amplitudes compared to strongly associated targets in the baseline condition. No N400 effect was observed for difficult to integrate (but contextually cued) target words in the event related violation condition compared to the baseline condition.

Table 1
ANOVAs on ERPs to target nouns across the N400 time window and the P600 time window.

			300–500 ms			600–1000 ms		
		df	*F*	*p*	η^2G	*F*	*p*	η^2G
ERV vs. Baseline								
Midline	Cond	(1, 20)	< 1	.92	< .001	3.90	.06	.03
	Cond × AP	(2, 40)	< 1	.80	< .001	1.01	.35	.004
ROIs	Cond	(1, 20)	< 1	.81	< .001	1.52	.23	.01
	Cond × AP	(1, 20)	< 1	.62	< .001	4.81	.04	.006
	Cond × H	(1, 20)	< 1	.33	< .001	< 1	.90	< .001
EUV vs. Baseline								
Midline	Cond	(1, 20)	49.1	< .001	.21	3.03	.10	.02
	Cond × AP	(2, 40)	< 1	.37	< .01	3.90	.04	.009
ROIs	Cond	(1, 20)	48.6	< .001	.21	2.93	.10	.02
	Cond × AP	(1, 20)	< 1	.98	< .001	8.36	< .01	.004
	Cond × H	(1, 20)	< 1	.96	< .001	< 1	.99	.001

Notes. ERV = Event Related Violation condition; EUV = Event Unrelated Violation condition; Cond × AP = Condition × Anterior–Posterior distribution; Cond × H = Condition × Hemisphere.

3.2.2. P600 time window (600–1000 ms)

The overall ANOVA yielded a Condition-by-Electrode interaction, F (28, 560) = 2.18, p = .04, η^2G = .02. The ANOVA over midline sites revealed an effect of condition, F(2, 40) = 5.97, p = .007, η^2G = .07) and no interaction with AP distribution (p = .14). Pairwise comparisons

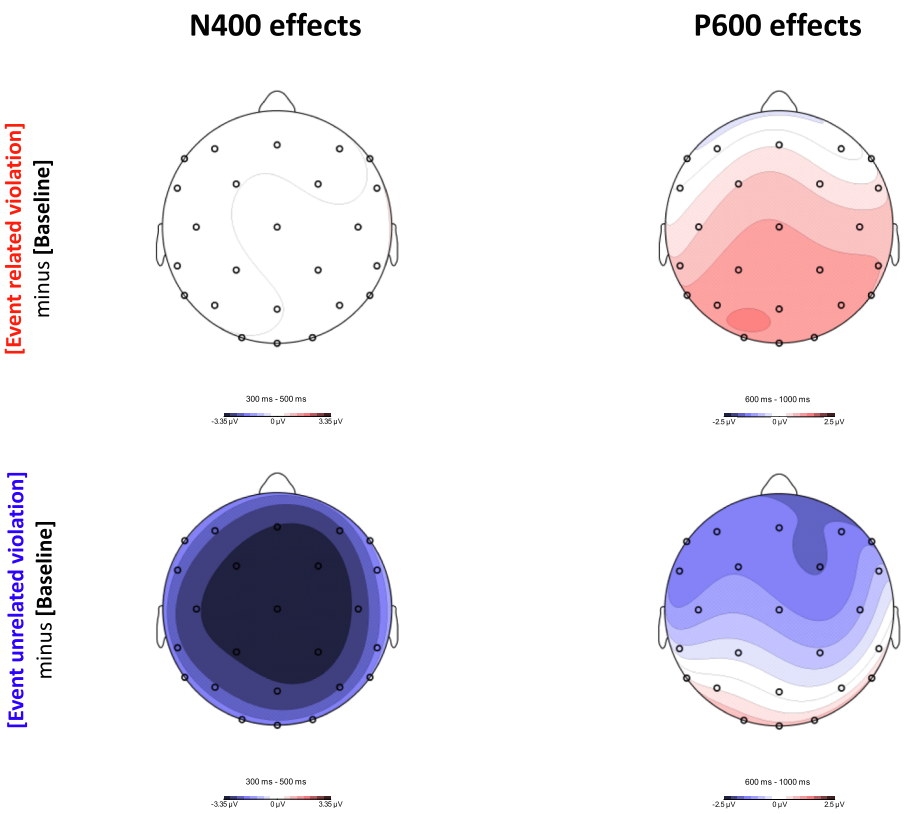

Fig. 2. Topographic maps of the effects in the N400 time window (300–500 ms, left column) and the P600 time window (600–1000 ms, right column). The upper panel shows the difference between the event related violation condition and the baseline; the lower panel shows the difference between the event unrelated violation condition and the baseline.

(see Table 1 and Fig. 2) showed that the event related condition was marginally more positive (M = 3.47, SD = 2.50) than the baseline condition (M = 2.55, SD = 2.53), while the event unrelated condition was more negative (M = 1.62, SD = 2.95) than the baseline, especially on anterior sites.

The ANOVA over lateral sites revealed an effect of Condition, F(2, 40) = 3.79, p = .03, η^2G = .04, and an interaction with AP distribution, F(2, 40) = 4.41, p = .02, η^2G = .008. Pairwise comparisons showed that the event related violation condition was more positive than the baseline over posterior sites (event related: M = 3.56, SD = 2.45; baseline: M = 2.60, SD = 2.30), while the event unrelated violation condition was more negative than the baseline over anterior sites (event unrelated: M = 0.78, SD = 2.75, baseline: M = 2.10, SD = 2.35) (see Fig. 2).

Since P600 effects to semantic (as well as syntactic) manipulations have been found to vary with respect to their latency, with some studies observing effects starting around 600 ms (e.g., Kim & Osterhout, 2005) while others in later time windows (e.g. Nieuwland & van Berkum, 2005), we performed pairwise comparisons in three 200 ms time windows at 100 ms onset intervals (600–800 ms, 700–900 ms, and 800–1000 ms), to assess where the effect was most pronounced.[2] The results are reported in Table 2 (see also the topographic maps in Fig. 3). The P600 effect for the event related violation condition emerged after 700 ms and was significant between 800 ms and 1000 ms. In this time window, pairwise comparisons in the midline electrodes showed that target nouns in the event related violation condition elicited a larger positivity (M = 3.40, SD = 2.79) than in the baseline condition (M = 2.01, SD = 2.54), while nouns in the event unrelated violation condition were marginally more negative (M = 0.99, SD = 3.11) than in the baseline, especially in frontal electrodes (see Fig. 3). On lateral sites, event related violating nouns elicited a larger positivity (M = 2.67, SD = 2.54), more pronounced over posterior electrodes (see Fig. 3), compared to the baseline (M = 1.88, SD = 2.08), while event unrelated violating nouns elicited a larger negativity (M = 1.12, SD = 2.56) than the baseline, more pronounced over anterior electrodes. This delayed latency of the P600 effect is consistent with previous findings on semantic illusions in discourse (Nieuwland & van Berkum, 2005), where the P600 effect was observed after 700 ms (although in the auditory modality).

To summarize, target nouns that were contextually cued but difficult to integrate into a plausible interpretation elicited a (late) positivity relative to the baseline condition, supporting the semantic integration view of the P600. No such late positivity was observed for the event unrelated violating target nouns, but rather an anterior

Table 2
ANOVAs on ERPs to target nouns across three overlapping time windows.

		600–800			700–900			800–1000		
		F	*p*	η^2G	*F*	*p*	η^2G	*F*	*p*	η^2G
ERV vs. Baseline										
Midline	Cond	< 1	.36	.007	3.76	.06	.03	8.54	.008	.05
	Cond × AP	< 1	.52	.002	< 1	.43	.002	1.23	.29	.004
ROIs	Cond	< 1	.68	.001	1.74	.20	.02	3.93	.06	.031
	Cond × AP	2.85	.11	.004	3.80	.06	.006	5.96	.02	.008
	Cond × H	< 1	.63	.00	< 1	.97	.00	< 1	.82	.00
EUV vs. Baseline										
Midline	Cond	2.33	.14	.02	2.85	.11	.02	3.02	.10	.03
	Cond × AP	3.70	.04	.007	2.90	.08	.05	3.43	.06	.01
ROIs	Cond	2.24	.15	.016	2.70	.12	.02	2.96	.10	.03
	Cond × AP	7.03	.02	.007	5.83	.02	.009	7.52	.01	.012
	Cond × H	< 1	.68	.00	< 1	.92	.00	< 1	.73	.00

Notes. ERV = Event Related Violation condition; EUV = Event Unrelated Violation condition; Cond × AP = Condition × Anterior-Posterior distribution. Cond × H = Condition × Hemisphere; numerator df = 1, denominator df = 20 for all F tests except for Cond × AP over the midline sites (numerator df = 2, denominator df = 40)

[2] In general, different onset latencies of P600 effects have been reported in the literature, from very early (Hoeks et al. 2013) to very late (Schacht et al. 2014).

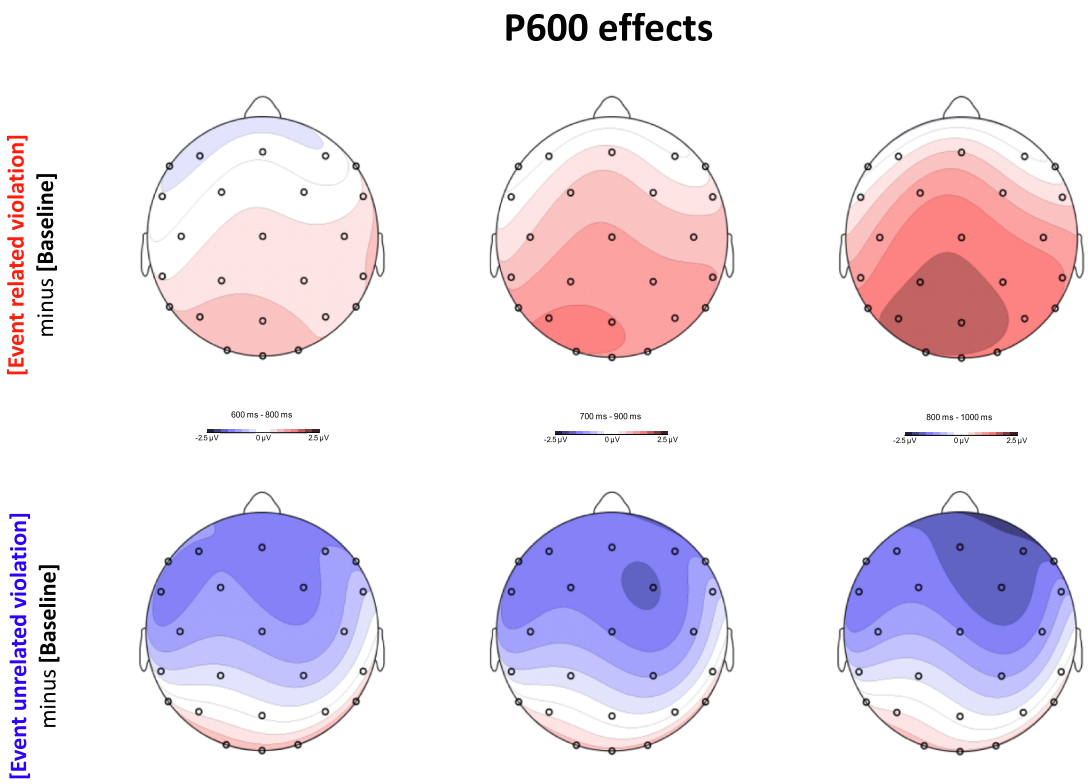

Fig. 3. Topographic maps of the P600 effect within three overlapping time windows (600–800 ms, left column; 700–900 ms, middle column; 800–1000 ms, right column). The upper panel shows the difference between the event related violation condition and the baseline; the lower panel shows the difference between the event unrelated violation condition and the baseline.

negativity, suggesting that the N400 effect elicited in the preceding time window was prolonged. This long-lasting effect may thus have obscured a potential P600 effect elicited in this condition, due to spatiotemporal overlap between the N400 and the P600 (Brouwer & Crocker, 2017; Hagoort, 2003). As can be seen from Fig. 1, in the P600 time window the event unrelated violation condition becomes gradually more positive over central and posterior sites. Fig. 2 and 3 show that this relative positive shift occurs in both violation conditions: in the P600 time window, both the event related and the event unrelated violation conditions become more positive than the baseline over posterior sites. The only difference is that the event unrelated violation condition is more negative than the baseline in the previous (N400) time window, while the event related violation condition shows no difference. To further investigate this pattern, we examined whether the trend for the event unrelated violation condition to become gradually more positive over posterior sites resulted in a significant effect over occipital electrodes (O1, Oz, O2), where the event unrelated violation condition appears to be more positive than the baseline (see Fig. 1). Pairwise comparisons in the 600–1000 time window confirmed the larger positivity for the event related violation condition (M = 3.40, SD = 2.06) compared to the baseline (M = 2.30, SD = 1.98), F(1, 20) = 5.27, p = .03, η^2G = .07. Crucially, the event unrelated violation condition was significantly more positive (M = 3.19, SD = 2.19) than the baseline, F(1, 20) = 5.17, p = .03, η^2G = .04, while it did not differ from the event related violation condition (F < 1). Thus, it is possible that a centro-parietal P600 effect in the event unrelated condition was masked by an overlapping, long-lasting negativity. We will return to this point below.

4. Discussion

In this paper, we set out to test competing theories of the two most salient ERP components – the N400 and the P600 – for language comprehension. Specifically, the N400 has been linked to processes related to lexical retrieval (e.g., Kutas & Federmeier, 2011), fully compositional semantic integration (e.g., Brown & Hagoort, 1993), or both (Baggio & Hagoort, 2011; Lau et al., 2016; Nieuwland et al., in press), while the P600 has been associated with syntactic reanalysis (e.g. Osterhout & Holcomb, 1992) or more general semantic integration processes (Brouwer et al., 2012).

To test these hypotheses, we measured ERP responses to critical words that, relative to a baseline condition (1a), were similarly easy to retrieve but more difficult to integrate (1b, the event related violation condition), or more difficult to both retrieve and integrate (1c, the event unrelated violation condition). The event related violation condition elicited a P600 effect compared to the baseline condition, but no N400 effect. The event unrelated violation condition, in turn, elicited a sustained N400 effect, but no P600 effect. These results allow for the following conclusions with regard to the functional interpretation of the N400 and the P600:

The N400 does not index semantic integration. The absence of an N400 effect for the event related violation condition compared to the baseline condition is inconsistent with both the semantic integration and the hybrid accounts of the N400 (Baggio & Hagoort, 2011; Brown & Hagoort, 1993, 2000; Hagoort et al., 2004). That is, the event related and the event unrelated violation conditions – both of which were judged as implausible compared to the baseline condition – would both be predicted to elicit increased N400 amplitude, indexing more effortful compositional semantic integration processes. Hybrid accounts, such as the one outlined in Nieuwland et al. (in press), according to which the earlier portion of the N400 reflects word meaning activation processes (as measured by word predictability, operationalized as cloze probability) while the later portion reflects integration processes (as measured by plausibility ratings), are also not supported by the present findings. Neither cloze probabilities nor plausibility ratings patterned with the observed N400 modulation, but only the semantic association

between the critical word and the context (see Nieuwland & van Berkum, 2005, Fischler, Bloom, Childers, Roucos, & Perry, 1983, for similar findings). Importantly, these findings can also not be reconciled with the integration view of the N400 assumed in multi-stream-based explanations developed in response to the semantic illusion phenomenon: there is no available semantically attractive (but syntactically unlicensed) alternative interpretation that renders the mini-discourses in the event related violation condition plausible. Hence, even a plausibility-driven stream would encounter integrative difficulty, predicting increased N400 amplitude. Further, one might be tempted to suggest that the absence of an N400 effect for the event related condition is due to some sort of shallow integrative processing (e.g., Ferreira, 2003; Rabovsky, Hansen, & McClelland, 2018), in which comprehenders only recall a restaurant-going event (and not that the event has ended), such that "opening the menu" seems coherent. Such an explanation, however, would fail to explain comprehenders' awareness of the violation, as manifest by the robust P600 effect for this condition (see Sanford et al., 2011).

The N400 indexes lexical retrieval processes. The N400 elicitation pattern for both the event related and the event unrelated violation conditions is consistent with the lexical retrieval account, in which the N400 amplitude is modulated by the degree to which contextual cues facilitate retrieval of conceptual knowledge associated with the eliciting word from long-term memory (e.g., Brouwer et al., 2012; Kutas & Federmeier, 2000; Lau et al., 2008; van Berkum, 2009). Specifically, an N400 effect is observed for target words judged as weakly associated with the context, as in the event unrelated condition (1c), but is not observed for strongly associated target words, as in the event related condition (1b).

The P600 does not reflect syntactic reanalysis/reprocessing alone. The observed P600 effect elicited by the event related violation condition relative to the baseline condition is inconsistent with an account of the P600 that is limited to syntactic reanalysis (e.g., Osterhout & Holcomb, 1992). The target sentences of our stimuli were globally and locally grammatical and unambiguous. Moreover, these findings cannot be reconciled with the view that semantic P600 effects arise from a conflict between a plausibility stream and a syntactically driven stream, as argued in multi-stream models (e.g., Kim & Osterhout, 2005; Kuperberg, 2007; van Herten et al., 2005). That is, unlike most semantic illusion studies that motivated multi-stream models, there was no thematic role re-assignment within either the context or the target sentence that could resolve the implausibility. Since any attempts to recover a plausible interpretation of the discourse would involve acting on the global (world knowledge-driven) interpretation of the discourse, and not on the syntactic analysis of the two sentences (which cannot be revised to accommodate any plausible re-interpretation of the discourse), no P600 effect is predicted.

Does the P600 reflect semantic integration processes? The observed P600 effect for event related violating targets is consistent with a semantic integration account of the P600 (Brouwer et al., 2017, 2012). Relative to the baseline condition, words that were semantically associated with the preceding context, but more difficult to integrate into it, elicited a positivity starting at around 700 ms. While the latency of this effect appears to be delayed relative to previous work on semantic P600s using one-sentence stimuli (e.g., Kim & Osterhout, 2005), it is consistent with later P600 effects observed for semantic illusions in discourse (Nieuwland & van Berkum, 2005). The delayed nature of the effect might therefore result from integration difficulty arising not at the local sentential context, but at the discourse level, that is, when the meaning of the critical word is integrated into the representation of the global situation, which is then checked against world knowledge.

The semantic integration account of the P600, however, predicts both violating conditions to produce a P600 effect, since both conditions described an implausible event. The event unrelated violation condition, however, showed a sustained N400 effect, but no P600 effect. It could be argued that the differential effects in the two violating conditions may depend on task-related factors, supporting the view that the P600 is a member of the P300 component family (Coulson, King & Kutas, 1998a, 1998b; Leckey & Federmeier, in press). The P300 is a domain-general response to unexpected, task-relevant events (see, e.g., Picton, 1992, for a review). Indeed, some studies have highlighted the sensitivity of the P600 to task manipulations (e.g., Hahne & Friederici, 2002; Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008; Schacht, Sommer, Shmuilovich, Martíenz, & Martín-Loeches, 2014), although the presence of a task is not a prerequisite for observing a P600 effect (e.g., Nieuwland & van Berkum, 2005). According to Coulson et al. (1998a), semantic anomalies (as well as syntactic anomalies) should elicit late positivities when they are easy to classify as unacceptable and are task-relevant (see also Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). This is not the case in our study, however, as we should have observed a stronger P600 in the event unrelated violation condition compared to the event related violation condition, as the event unrelated condition was easier to classify as implausible, as demonstrated by both the plausibility ratings in the norming study and the plausibility judgments in the ERP experiment. Conversely, it could be argued that P600 amplitude is enhanced with anomalies that are more difficult to classify as implausible, as might be argued to be the case for the event related violating targets in our study. It should be noticed, however, that increased task discrimination difficulty is associated with *lower* P300 amplitudes and increased P300 *latency* (Fabiani, Gratton, Karis, & Donchin, 1987; Magliero, Bashore, Coles, & Donchin, 1984; Polich, 1987; Sassenhagen et al., 2014). Thus, it is unlikely that the presence of a P600 effect in this condition is related to task difficulty, as we observed a higher P600 amplitude precisely when the task was presumably more difficult. However, to further examine this issue, we identified 20 experimental items that in the offline plausibility judgment task scored in the middle range (between 3 and 5 in a 7-point Likert scale). We assumed these items to be more difficult to classify in a yes-no plausibility judgment task. Statistical analyses showed that even when these more problematic items are removed, the event related condition elicits a reliably larger positivity compared to the baseline condition.[3] We also examined whether the P600 effect observed in the event related violation condition was more pronounced for participants who were less accurate in judging this condition as implausible in the online discrimination task (and who, arguably, might have found the task more difficult). We split participants into two groups, depending on their performance in the event related violation condition: one group included the 10 most accurate participants, while the other group included the 10 least accurate ones. As shown in Fig. 4, both groups showed an N400 effect for the event unrelated violation condition compared to baseline, but no N400 effect for the event related violation condition. In the 800–1000 ms time window, however, the analyses revealed that the most accurate group produced a reliable P600 effect relative to the baseline, while the least accurate group showed no significant effects. Thus, as predicted by the P600-as-integration hypothesis (see Brouwer et al., 2012), the P600 effect appears to be driven by whether or not participants detected the implausibility of the event related violation condition rather than by task difficulty. The results of the statistical analyses are reported in Appendix B.

It remains to be explained why the event unrelated violation condition did not produce a P600 effect. One explanation might be related to the extent to which people engage in integrative processes in the

[3] The comparison between the event related violation condition and the baseline in the 800–1000 ms time window produced a significant effect of Condition in the Midline analysis, $F(1,20) = 9.53$, $p = .006$, $\eta^2G = .09$, and in the ROIs analysis, $F(1,20) = 5.50$, $p = .03$, $\eta^2G = .06$. The ROI's analysis also showed a significant Condition x AP interaction, $F(1,20) = 4.51$, $p < .05$, $\eta^2G = .009$, indicating that the effect was more pronounced over posterior electrodes.

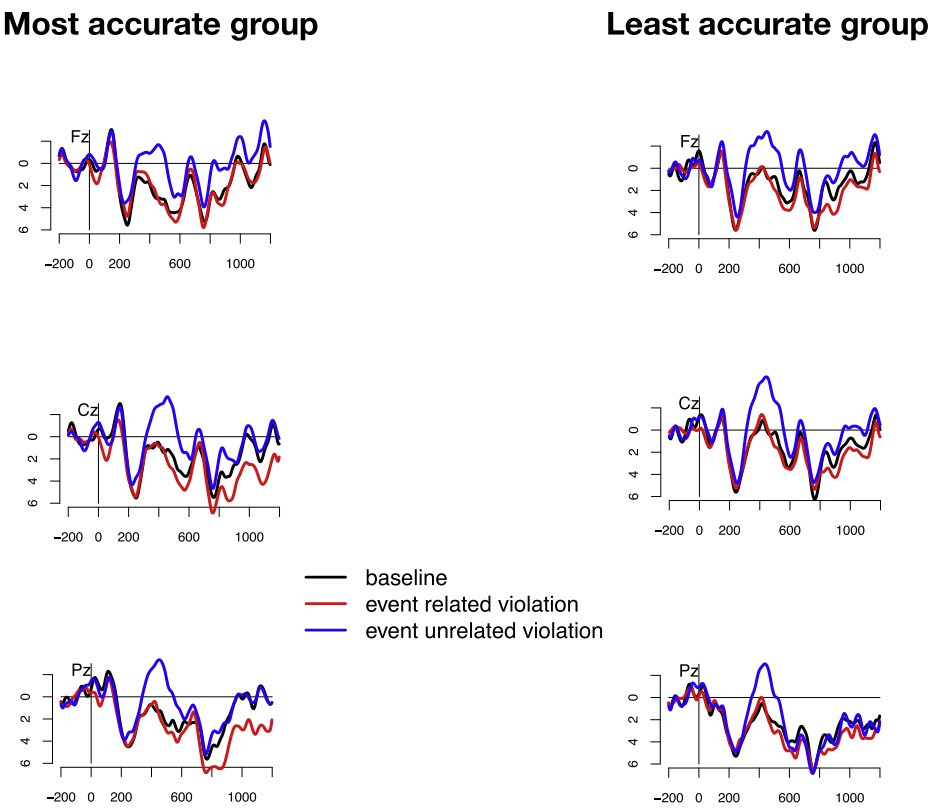

Fig. 4. Grand-average ERP waveforms time-locked to the target nouns (only midline electrodes) for two groups of participants (see main text, Section 4). The left panel shows waveforms for the 10 most accurate participants, the right panel shows the waveforms for the 10 least accurate participants. Negative voltages are plotted upwards.

presence of strong implausibilities. Previous work on joke comprehension has shown that when people don't understand a joke, the well-documented irony-related P600 effect (e.g., Regel et al., 2011) is no longer observed (Coulson & Kutas, 2001). This finding suggests that, under certain circumstances, people may not even try to make sense of implausible input. Since the event related violation condition was perceived as significantly less implausible than the event unrelated violation condition, participants may have tried to construct a plausible interpretation for the former (e.g., by inferring that John took the menu from the restaurant for some reason, in our example), but not for the latter (see, for example, item 13 in the Appendix: *Thorsten began to grill. First, he took the shampoo and…*), resulting in no P600 effect in this condition. This explanation, however, is not entirely satisfactory as P600 effects to strong implausibilities (and incongruities) are extensively documented in the literature (see, for example, van Petten & Luka, 2012, for an overview of observed P600 effects to semantic incongruities).

We therefore considered whether the observed pattern of effects may have a neurophysiological explanation grounded in component overlap (Brouwer & Crocker, 2017; Fabiani, Gratton, & Federmeier, 2007; Luck, 2005). At any given point, the ERP waveform merely shows the summation of the latent components contributing to the ERP signal at that time (Luck, 2005). Indeed, the processes underlying the N400 and the P600 may temporally overlap (Heikel, Sassenhagen, & Fiebach, 2018), and the ERP signal at any given point may reflect a combination of these components (Hagoort, 2003). Consequently, the presence of a large N400 effect might obscure a P600 effect. We noted earlier that the event unrelated condition clearly shows a positive-going deflection over centro-posterior sites, which becomes more positive than the baseline in occipital electrodes. This pattern might be due to the prolonged negativity starting in the N400 time window, which attenuates any positive deflection in the later time window, thereby obscuring an emerging P600 with a posterior distribution.

Importantly, such a component overlap explanation cannot be convincingly used to support the N400 as integration hypothesis. Such an account would imply that the observed P600 effect for the event related violation condition masked the preceding N400, explaining the absence of an N400 effect for this condition. However, an inspection of the waveform shows no evidence for the presence of an N400 effect in the early portion of the N400 or in more anterior electrodes, where a posteriorly distributed P600 should have very limited influence. Further, such an account implies a bi-phasic N400-P600 effect, which would then be expected in the event unrelated condition as well.

To summarise, our critical event related violation condition – in which target words were judged as being strongly associated with the context, but as resulting in implausible sentences – elicited no N400 effect, but a robust P600. An N400 was only elicited in the event unrelated violation condition – in which target words were judged as being weakly associated with the context, consistent with previous findings (e.g., Fischler et al., 1983; Hoeks et al., 2004; Kolk et al., 2003; Kuperberg et al., 2007; Nieuwland & van Berkum, 2005; Schumacher, 2011, 2014; van Herten et al., 2005, 2006). Crucially, on the retrieval account, association, predictability, and plausibility can influence retrieval of word meaning, thereby modulating N400 amplitude. On the integration (and hybrid) account, by contrast, implausibility necessarily predicts increased N400 amplitude regardless of lexical association. While the retrieval account can explain the elimination of the N400 we find for associated but implausible targets (as well as the influence of predictability and plausibility on the N400 found in other studies – e.g. Nieuwland et al., in press), integration and hybrid accounts must be compromised in order to explain the absence of integration difficulty for such implausible but associated targets.

The presence of a P600 effect for the event related violation is predicted by the implausibility of this condition. Two possible explanations for the absence of a predicted P600 effect in the event unrelated violation condition, both of which are consistent with the P600-as-integration hypothesis, were considered. Firstly, the stronger implausibility of the event unrelated violation condition might have

prevented participants from even attempting integration with the discourse context, but this is inconsistent with P600 effects reported for strong violations. Alternatively, we suggest that the P600 effect may have been obscured by the spatio-temporal overlap with a long-lasting negativity starting in the N400 time window. Importantly, our consideration of component overlap explanations is motivated to address the more general inconsistency in the pattern of findings on the P600, as evidenced by the van Petten and Luka (2012) meta-study on semantic incongruency, which is problematic for all accounts of the P600 (see Brouwer & Crocker, 2017, for discussion). Further, van Petten and Luka (2012) did not consider studies using a task – which would encourage (and allow assessment) of participants' engagement in by-item comprehension – possibly underestimating P600 effects in response to incongruency (Kolk et al., 2003). As the P600 for the event related condition cannot be explained by any syntactic ambiguity or violation in our study, this adds to the growing body of findings (e.g., Burkhardt, 2006, 2007; Delogu et al., 2018; Hoeks et al., 2013; Regel et al., 2011; Spotorno et al., 2013) suggesting that the P600 is a more general index of integration difficulty. Crucially, this account subsumes a syntactic reanalysis/reprocessing interpretation, as ungrammaticalities and structural revisions have direct consequences for semantic integration. The proposal that the P600 is an overarching index of compositional semantic integration processes – informed by syntax, semantics, and pragmatics – is consistent with the diversity of morphologically distinct P600s that have been reported in the literature, depending on the specific source of the integration difficulty (Brouwer & Hoeks, 2013). Collectively, we argue that the pattern of findings observed in the present study are most plausibly reconciled within the retrieval-integration account of the N400 and the P600 (Brouwer et al., 2012, 2017).

In order to exploit ERPs in the study of language, it is essential to clearly understand what neurocognitive processes they index. It is not uncommon to see studies in which very different conclusions about cognitive processes could be made depending on which interpretation of the relevant dependent measure is assumed. Take, for example, the well-documented insensitivity of the N400 to negation in sentences like "A robin is (not) a bird" (e.g., Fischler et al., 1983). From this finding, it is often argued that negation is not integrated online (e.g., Dudschig, Mackenzie, Maienborn, Kaup, & Leuthold, 2019). This conclusion rests on the assumption that the N400 indexes compositional semantic integration. The retrieval account of the N400, however, can straightforwardly explain this finding, as "robin" and "bird" are semantically associated. Thus, establishing the functional role of ERP components is a prerequisite to correctly interpreting experimental findings in the investigation of the language architecture. Until this issue is settled, care should be taken to draw conclusions concerning the assumed underlying cognitive processes.

In conclusion, while the observed pattern of N400 effects was predicted by the retrieval account alone, the pattern of findings in the P600 time window could not be directly explained by any of the considered accounts. In examining various means to reconcile this pattern of results with the extant accounts, we argued component overlap to be the most satisfactory. While spatio-temporal overlap must be invoked with care, it is a reality of electrophysiological measures which should be systematically addressed when interpreting data with respect to the predictions of relevant theories. To the extent that the component overlap explanation is correct, our findings suggest that the two most salient ERP components for language comprehension differentially index perhaps the two most fundamental operations of incremental language understanding, namely the retrieval of word meaning from long-term memory (N400) and the integration of this meaning into the utterance interpretation (P600). More specifically, we demonstrate that event knowledge established by the prior discourse influences these two operations differentially, such that association with the context is sufficient to ease lexical retrieval even when integration is difficult.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (SFB/CRC 1102 "Information density and linguistic encoding", Project A1).

Appendix A. Experimental stimuli in German

Each item consists of a context sentence and a target sentence, and is presented in three conditions: *baseline*, *event related violation*, and *event unrelated violation* (see example (1) in the Section 1. of the main article). Only context sentences vary per condition. The baseline and the event related violation conditions have different verbs, but share the NP of the verb phrase (first set of square brackets). The event unrelated violation condition has the same verb as in the baseline condition, but a different noun phrase (second set of square brackets). In the subsequent target sentence, the critical word is underlined.

1. Johann [*betrat/verließ* das Restaurant]/[betrat die Wohnung]. Wenig später öffnete er die Speisekarte und verschaffte sich einen Überblick.
2. Sabine [*betrat/verließ* das Kino]/[betrat die Schule]. Schnell ging sie zur Kasse und kaufte eine Karte.
3. Kevin [*betrat/verließ* den Bauernhof]/[betrat die Kirche]. Ohne zu zögern nahm er eine Mistgabel und fing an zu schaufeln.
4. Susi [*betrat/verließ* die Arztpraxis]/[betrat das Gehege]. Freundlich sprach sie mit der Arzthelferin über die Problematik.
5. Roman [*betrat/verließ* die Bücherei]/[betrat die Kneipe]. Kurz danach ging er zum Regal und suchte einen Krimi.
6. Lea [*begann/war fertig* damit Spaghetti zu kochen]/[begann die Wände zu streichen]. Vorsichtig öffnete sie die Nudelpackung mit einem Handgriff.
7. Peter [*erreichte/verließ* das Theater]/[erreichte den Supermarkt]. Wenig später ging er zur Loge und setzte sich.
8. Lisa [*betrat/verließ* den Bahnhof]/[betrat das Haus]. Wenig später war sie am Gleis und wartete auf den Zug.
9. Tim [*begann/hörte auf* im Fitnessstudio zu trainieren]/[begann seine Fahrstunde]. Umgehend war er auf dem Laufband und rann wie verrückt.
10. Maria [*erreichte/verließ* den Flughafen]/[erreichte das Einkaufszentrum]. Sofort ging sie zum Check-in um den Koffer aufzugeben.
11. Lukas [*kam zur/verließ die* Geburtstagsfeier]/[kam zum Unterricht]. Nach einer Weile überreichte er den Kuchen und gratulierte.
12. Marie [*ging zum/verließ den* Friseur]/[ging zur Uni]. Nach einem kurzen Moment fragte sie die Stylistin nach einem neuen Schnitt.
13. Thorsten [*begann/hörte auf* zu duschen]/[begann zu grillen]. Als erstes benutzte er das Shampoo und dann das Duschgel.
14. Claudia [*betrat/verließ* den Blumenladen]/[betrat das Tiergeschäft]. Schnell fragte sie nach einer Rose für eine Freundin.
15. Jonathan [*ging zum/verließ den* Friedhof]/[ging in die Küche]. Nachdenklich betrachtete er den Grabstein und sprach ein Gebet.
16. Lara [*ging in/verließ* den Garten]/[ging zum Herd]. Genau prüfte sie die Erde und goss dann die Blumen.
17. Roman [*ging ins/verließ das* Museum]/[ging in die Buchhandlung]. Akribisch betrachtete er die Skulpturen und war begeistert.
18. Annika [*ging zum/verließ den* Konditor]/[ging in die Boutique]. Schnell fragte sie nach einem Kuchen und bezahlte sofort.
19. Emil [*betrat/verließ* das Spa]/[betrat die Apotheke]. Nach einer Weile bekam er eine Massage und genoss sie.
20. Susanne [*betrat/verließ* die Reitschule]/[betrat das Bad]. Schnell setzte sie den Sattel auf und ritt los.

21. Martin [*kam zur/verließ die* Sicherheitskontrolle]/[kam zur Autowerkstatt]. Zügig öffnete er den Koffer und wurde durchgewunken.
22. Jutta [*ging zur/verließ die* Hotelrezeption]/[ging zur Post]. Sofort bekam sie ein Zimmer und dankte freundlich.
23. Jens [*ging zum/verließ den* Zahnarzt]/[ging zur Zollstelle]. Ängstlich ging er zum Röntgen und ließ sich behandeln.
24. Elena [*begann/hörte auf* Poker zu spielen]/[begann zu zeichnen]. Schnell setzte sie ihr Geld und gewann.
25. Bruno [*begann/beendete* seinen Urlaub]/[begann seinen Filmabend]. Endlich kam er am Meer an und genoss das Wetter.
26. Jana [*betrat/verließ* das Bürgeramt]/[betrat das Gartencenter]. Umgehend zeigte sie ihr Visum und stellte einen Antrag.
27. Nico [*fuhr auf/verließ* die Kirmes]/[fuhr zur Pizzeria]. Sofort ging er zum Auto-Scooter und fuhr eine Runde.
28. Beate [*ging/kam zurück vom* Wandern]/[ging schwimmen]. Nach einer Weile sah sie ein Edelweiß das sie gleich mitnahm.
29. Hubert [*nahm ein/verließ das* Taxi]/[nahm ein Flugzeug]. Freundlich begrüßte er den Fahrer und nannte sein Ziel.
30. Jenni [*begann/ hörte auf* zu baden]/[Jenni begann die Vorlesung]. Sofort verteilte sie den Schaum auf ihrem Körper.
31. Hans [*betrat/verließ* die Eishalle]/[betrat die Kletterhalle]. Flott schnürte er seine Schlittschuhe und ging auf die Bahn.
32. Clara [*hörte auf* zu bowlen]/[ging zum Schießstand]. Gekonnt erzielte sie einen Strike und gewann das Spiel.
33. Georg [*nahm teil an einem/beendete einen* Feuerwehreinsatz]/[nahm teil an Yoga]. Sofort benutzte er den Schlauch und löschte das Feuer.
34. Johanna [*begann/beendete* das Fußballspiel]/[begann die Jagd]. Nach kurzer Zeit machte sie ein Tor und jubelte laut.
35. Jürgen [*betrat/verließ* die Metzgerei]/[betrat die Bäckerei]. Sofort fragte er nach einem Steak und bezahlte.
36. Frauke [*betrat/verließ* den Zoo]/[ging ins Stadion]. Wenig später fotografierte sie ein Zebra beim Trinken.
37. Bernd [*betrat/verließ* den Waschsalon]/[betrat den Kiosk]. Schnell wählte er das Waschprogramm und lies es laufen.
38. Nicole [*ging in/verließ* den Zirkus]/[ging ins Büro]. Bald setzte sie sich auf die Tribüne und schaute sich das Spektakel an.
39. Viktor [*ging in/verließ* die Sauna]/[ging ins Labor]. Nach einer Weile machte er einen Aufguss und legte sich hin.
40. Gabi [*ging zum/verließ* den Tierarzt]/[/in die Umkleide]. Einen Moment später fragte sie die Arzthelferin nach Medikamenten.
41. Jakob [*betrat/verließ* den Tatort]/[betrat die Tankstelle]. Kurz darauf fotografierte er die Blutspritzer und war erstaunt.
42. Peter [*begann/hörte auf* Bäume zu fällen]/[begann die Kutsche zu fahren]. Nach einer Weile benutzte sie die Axt und traf gezielt den Stamm.
43. Tom [*begann ein/beendete das* Lagerfeuer]/[begann Frühstück zu machen]. Sofort stapelte er das Holz und zündete es an.
44. Kim [*ging zum/verließ den* Weihnachtsmarkt]/[ging ins Krankenhaus]. Kurz darauf erreichte er die Weihnachtskiosks und stöberte nach Geschenken.
45. Michael [*begann/war fertig* sein Mittagessen zu kochen]/[begann einen Kuchen zu backen]. Schnell schnitt er das Gemüse und warf es in den Topf.
46. Johann [*betrat/verließ* die Wohnung]/[betrat das Kino]. Sofort setzte er sich aufs Bett und dachte über den Tag nach.
47. Sabine [*betrat/verließ* die Schule]/[betrat den Bauernhof]. Schnell ging sie zum Klassensaal und grüßte die Schüler.
48. Kevin [*betrat/verließ* die Kirche]/[betrat die Arztpraxis]. Vorsichtig nahm er eine Kerze und zündete sie an.
49. Susi [*betrat/verließ* das Gehege]/[betrat die Bücherei]. Präzise begutachtete sie das Kalb und machte sich Notizen.
50. Thomas [*betrat/verließ* die Kneipe]/[begann Spaghetti zu kochen]. In Eile rief er den Kellner und bestellte ein Bier.
51. Lea [*begann/war fertig* die Wände zu streichen]/[erreichte das Theater]. Schnell wählte sie einen Pinsel und fing an zu streichen.
52. Peter [*erreichte/verließ* den Supermarkt]/[erreichte den Bahnhof]. Schnell kaufte er etwas Gemüse zum Kochen am Abend.
53. Lisa [*betrat/verließ* das Haus]/[begann im Fitnessstudio zu trainieren]. Kurz danach checkte sie den Kühlschrank und suchte etwas zu essen.
54. Tim [*begann/beendete* seine Fahrstunde]/[ging zum Flughafen]. Hektisch drückte er aufs Gaspedal und fuhr los.
55. Maria [*erreichte/verließ* das Einkaufszentrum]/[erreichte die Geburtstagsfeier]. Sofort begrüßte sie die Verkäuferin und suchte etwas aus.
56. Lukas [*begann/beendete* die Vorlesung]/[ging zum Friseur]. Eilig begrüßte er seine Studenten und fing an.
57. Marie [*ging zur/verließ die* Uni]/[ging duschen]. In Eile betrat sie den Hörsaal und öffnete ihr Buch.
58. Thorsten [*begann/hörte auf* zu grillen]/[betrat den Blumenladen]. Behutsam platzierte er den Grillanzünder und suchte die Zange.
59. Claudia [*betrat/verließ* das Tiergeschäft]/[betrat den Friedhof]. Eine Zeit lang stand sie am Käfig und dachte über den Kauf nach.
60. Jonathan [*ging in/verließ* die Küche]/[ging in den Garten]. Eine Weile säuberte er den Ofen und danach den Kühlschrank.
61. Lara [*ging zum/ging weg vom* Herd]/[ging ins Museum]. Schnell reinigte sie die Herdplatte und dann die Abzugshaube.
62. Roman [*ging in/verließ* die Buchhandlung]/[ging zum Konditor]. Schnell suchte er einen Roman für seine Frau.
63. Annika [*ging in/verließ* die Boutique]/[ging ins Spa]. Eine Zeit lang suchte sie nach Stiefeln und kaufte sie dann.
64. Emil [*betrat/verließ* die Apotheke]/[betrat die Reitschule]. Schnell kaufte er das Medikament und ging dann nach Hause.
65. Susanne [*betrat/verließ* das Bad]/[ging zur Sicherheitskontrolle]. Sofort entfernte sie ihr Makeup und putzte sich die Zähne.
66. Martin [*kam zur/verließ die* Autowerkstatt]/[kam zur Hotelrezeption]. Umgehend fragte er nach den Reifen für den Winter.
67. Jutta [*ging zur/verließ die* Post]/[ging zum Zahnarzt]. Umgehend bekam sie das Paket auf das sie gewartet hatte.
68. Jens [*ging zur/verließ die* Zollstelle]/[begann Poker zu spielen]. Schnell zeigte er sein Formular und schaute den Beamten an.
69. Elena [*begann/hörte auf* zu zeichnen]/[begann den Urlaub]. Umgehend nahm sie den Stift und zeichnete eine Blume.
70. Bruno [*begann/beendete* seinen Filmabend]/[betrat das Bürgeramt]. Schnell besorgte er das Popcorn aus der Küche.
71. Jana [*betrat/verließ* das Gartencenter]/[ging auf die Kirmes]. Nach einer Weile fragte sie nach einer Palme für den Balkon.
72. Nico [*fuhr zur/verließ die* Pizzeria]/[fuhr wandern]. Sofort griff er die Speisekarte und bestellte eine Margherita.
73. Beate [*ging/kam zurück vom* Schwimmen]/[fuhr mit dem Taxi]. Schnell prüfte sie das Becken und schwamm los.
74. Hubert [*nahm ein/verließ* das Flugzeug]/[begann zu baden]. Voller Begeisterung genoss er die Höhe denn er liebte das Fliegen.
75. Jenni [*begann den/hörte auf mit dem* Unterricht]/[betrat die Eishalle]. Sofort öffnete sie die Tafel und begann zu schreiben.
76. Hans [*betrat/verließ* die Kletterhalle]/[begann zu bowlen]. Wenig später griff er das Seil und spannte es ein.
77. Clara [*ging zum/verließ* den Schießstand]/[ging auf einen Feuerwehreinsatz]. Sofort bekam sie einen Revolver und begann auf das Ziel zu schießen.
78. Georg [*nahm teil an/kam zurück vom* Yoga]/[nahm teil am Fußballspiel]. Nach einer Weile setzte er sich auf seine Yogamatte und machte eine Übung.
79. Johanna [*begann die/kam zurück von der* Jagd]/[betrat die Metzgerei]. Nach einer Weile erblickte sie den Hochsitz und ging darauf zu.
80. Jürgen [*betrat/verließ* die Bäckerei]/[betrat den Zoo]. Wenig später zeigte er auf das Croissant und wählte aus.
81. Frauke [*ging ins/verließ das* Stadion]/[ging zum Waschsalon]. Schnell wählte sie eine Tribüne und setzte sich.
82. Bernd [*betrat/verließ* den Kiosk]/[betrat den Zirkus]. Sofort kaufte er eine Zeitung und ging wieder heim.

83. Nicole [*betrat/verließ* das Büro]/[betrat die Sauna]. Zuerst ging sie zum Schreibtisch und schaltete dann den Computer ein.
84. Viktor [*ging ins/verließ das* Labor]/[ging zum Tierarzt]. Schnell begann er das Experiment und machte sich Notizen.
85. Gabi [*ging in/verließ* die Umkleide]/[betrat den Tatort]. Schnell probierte sie die Kleidung an und ging zur Kasse.
86. Jakob [*betrat/verließ* die Tankstelle]/[begann Bäume zu fällen]. Sofort sagte er seine Nummer und bezahlte.
87. Hilde [*begann/war fertig* die Kutsche zu fahren]/[begann ein Lagerfeuer]. Schnell bestieg sie den Kutschbock und los ging es.
88. Tom [*begann/hörte auf* Frühstück zu machen]/[ging zum Weihnachtsmarkt]. Sofort öffnete er die Butter und machte sich ein Brot.
89. Kim [*ging ins/verließ das* Krankenhaus]/[fing an sein Mittagessen zu kochen]. Gleich ging sie zur Rezeption und fragte nach dem Weg.
90. Michael [*fing an/war fertig* damit einen Kuchen zu backen]/[betrat das Restaurant]. Als erstes nahm er einen Schneebesen und schlug Eiweiß steif.

Appendix B. Subsidiary analyses based on participants' performance on the online plausibility judgment

The following tables report the results of the ANOVAs for two groups of participants, identified on the basis of their performance in the online plausibility judgment task. Table 3 reports the results for the group including the 10 most accurate participants in judging the event related violation condition as implausible (range: 94–77%), Table 4 reports the results for the 10 least accurate participants (range: 77–24%) (see main text, Section 4).

Table 3
ANOVAs on ERPs to target nouns in the N400 and P600 time windows for the 10 most accurate participants.

		300–500 ms				800–1000 ms		
		df	*F*	*p*	η^2G	*F*	*p*	η^2G
ERV vs. Baseline								
Midline	Cond	(1, 9)	< 1	.97	< .001	9.70	.01	.06
	Cond × AP	(2, 18)	< 1	.58	.002	2.39	.14	.02
ROIs	Cond	(1, 9)	< 1	.87	< .001	3.53	.09	.04
	Cond × AP	(1, 9)	< 1	.40	.001	6.29	.03	.03
	Cond × H	(1, 9)	1.13	.31	< .001	< 1	.68	< .001
EUV vs. Baseline								
Midline	Cond	(1, 9)	18.4	.002	.32	1.96	.20	.03
	Cond × AP	(2, 18)	< 1	.87	< .001	1.46	.26	.01
ROIs	Cond	(1, 9)	16.6	.003	.29	1.50	.25	.03
	Cond × AP	(1, 9)	< 1	.89	< .001	4.96	.05	.02
	Cond × H	(1, 9)	< 1	.72	< .001	< 1	.88	< .001

Notes. ERV = Event Related Violation condition; EUV = Event Unrelated Violation condition; Cond × AP = Condition × Anterior-Posterior distribution. Cond × H = Condition × Hemisphere.

Table 4
ANOVAs on ERPs to target nouns in the N400 and P600 time windows for the 10 least accurate participants.

		300–500 ms				800–1000 ms		
		df	*F*	*p*	η^2G	*F*	*p*	η^2G
ERV vs. Baseline								
Midline	Cond	(1, 9)	< 1	.98	< .001	1.87	.20	.05
	Cond × AP	(2, 18)	< 1	.60	< .001	< 1	.82	< .001
ROIs	Cond	(1, 9)	< 1	.78	< .001	1.02	.34	.03
	Cond × AP	(1, 9)	< 1	.99	< .001	< 1	.38	.001
	Cond × H	(1, 9)	< 1	.80	< .001	< 1	.83	< .001
EUV vs. Baseline								
Midline	Cond	(1, 9)	66.2	< .001	.17	< 1	.49	.01
	Cond × AP	(2, 18)	1.55	.24	.002	3.08	.10	.01
ROIs	Cond	(1, 9)	98.9	< .001	.21	< 1	.38	.02
	Cond × AP	(1, 9)	< 1	.60	< .001	5.20	.05	.01
	Cond × H	(1, 9)	< 1	.66	< .001	< 1	.48	< .001

Notes. ERV = Event Related Violation condition; EUV = Event Unrelated Violation condition; Cond × AP = Condition × Anterior-Posterior distribution. Cond × H = Condition × Hemisphere.

References

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes, 26*, 1338–1367.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews, 59*(1), 55–73.

Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology, 8*, 1327.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science, 41*, 75–83.

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

Brouwer, H., & Hoeks, J. C. J. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience, 7*, 758.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: evidence from masked priming. *Journal of Cognitive Neuroscience, 5*, 34–44.

Brown, C. M., & Hagoort, P. (2000). On the electrophysiology of language comprehension: Implications for the human language system. In M. Crocker, P. Pickering, & C. Clifton (Eds.). *Architectures and mechanisms for language processing* (pp. 213–237). Cambridge University Press.

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

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport, 18*(17), 1851–1854.

Chow, W. Y., Lau, E. F., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience, 33*, 803–828.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21–58.

Coulson, S., King, J. W., & Kutas, M. (1998). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13*(6), 653–672.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters, 316*, 71–74.

Delogu, F., Drenhaus, H., & Crocker, M. W. (2018). On the predictability of event boundaries in discourse: An ERP investigation. *Memory and Cognition, 46*, 315–325.

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., & Leuthold, H. (2019). Negation and the N400: Investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Language, Cognition and Neuroscience, 34*, 309–319.

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.). *Handbook of psychophysiology* (pp. 85–119). New York, NY, US: Cambridge University Press.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advance in Psychology, 2*, 1–78.

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*, 469–495.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 574–579.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W., Jr. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400–409.

Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. *Neuropsychologia, 49*, 574–579.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neurological and neurophysiological data. *Brain and Language, 50*(3), 259–281.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*, 149–188.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Vol. Ed.), *The cognitive neuroscience: Vol. 4*, (pp. 819–836). Boston: MIT Press.

Hagoort, P., Brown, C., & Groothusen (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Hagoort, P., Brown, C., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. Brown, & P. Hagoort (Eds.). *The Neurocognition of Language* (pp. 273–316). Oxford, UK: Oxford University Press.

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

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*(3), 339–356.

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. F., & Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language, 59*, 54–96.

Heikel, E., Sassenhagen, J., & Fiebach, C. J. (2018). Time-generalized multivariate analysis of EEG responses reveals a cascading architecture of semantic mismatch processing. *Brain and Language, 184*, 43–53.

Hoeks, J. C. J., & Brouwer, H. (2014). Electrophysiological research on conversation and discourse processing. In T. Holtgraves (Ed.). *The OxfordHandbook of Language and Social Psychology* (pp. 365–386). Oxford University Press.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*(1), 59–73.

Hoeks, J. C. J., Stowe, L. A., Hendriks, P., & Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PLoS ONE, 8*, e7359.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98–110.

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.

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Stucture and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1–36.

Kos, M., Vosse, T., van den Brink, D., & Hagoort, P. (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology, 1*, 1–11.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain and Language, 100*(3), 223–237.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117–129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

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.

Lau, E. F., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language, 111*(3), 161–172.

Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabora:Psychology, 2*(1), 1–19.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.

Leckey, M., & Federmeier, K. D. (in press). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*. https://doi.org/10.1111/psyp.13351.

Luck, S. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 171–186.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., …, & Wolfsthurn, S.V.G.Z. (in press). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*. https://doi.org/10.1101/267815.

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.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 786–803.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 30*, 739–773.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*, 456–479.

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 68*, 311–320.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour, 2*, 693–705.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience, 23*(2), 277–293.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awarness: An ERP study. *Journal of Cognitive Neuroscience, 23*(3), 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.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., & Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE, 9*, e91226.

Schumacher, P. B. (2011). The hepatitis called…: Electrophysiological evidence for enriched composition. In J. Meibauer, & M. Steinbach (Eds.). *Experimental pragmatics/semantics* (pp. 199–219). Amsterdam: John Benjamins.

Schumacher, P. B. (2014). Content and context in incremental processing: "the ham

sandwich" revisited. *Philosophical Studies, 168*, 151–165.
Spotorno, N., Cheylus, A., Henst, J. B. V. D., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS ONE, 8*, e66839.
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). Palgrave Macmillan.
van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad – Any implications for neuropragmatics? *Italian Journal of Linguistics, 22*, 181–208.
van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*(7), 1181–1197.
van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241–255.
van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190.