Brain Responses to World Knowledge Violations: A Comparison of Stimulus- and Fixation-triggered Event-related Potentials and Neural Oscillations

Paul Metzner[1], Titus von der Malsburg[2], Shravan Vasishth[1], and Frank Rösler[3]

Abstract

■ Recent research has shown that brain potentials time-locked to fixations in natural reading can be similar to brain potentials recorded during rapid serial visual presentation (RSVP). We attempted two replications of Hagoort, Hald, Bastiaansen, and Petersson [Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–441, 2004] to determine whether this correspondence also holds for oscillatory brain responses. Hagoort et al. reported an N400 effect and synchronization in the theta and gamma range following world knowledge violations. Our first experiment ($n = 32$) used RSVP and replicated both the N400 effect in the ERPs and the power increase in the theta range in the time–frequency domain. In the second experiment ($n = 49$), participants read the same materials freely while their eye movements and their EEG were monitored. First fixation durations, gaze durations, and regression rates were increased, and the ERP showed an N400 effect. An analysis of time–frequency representations showed synchronization in the delta range (1–3 Hz) and desynchronization in the upper alpha range (11–13 Hz) but no theta or gamma effects. The results suggest that oscillatory EEG changes elicited by world knowledge violations are different in natural reading and RSVP. This may reflect differences in how representations are constructed and retrieved from memory in the two presentation modes. ■

INTRODUCTION

The assessment of eye movements and EEG has helped to advance psycholinguistic research substantially, but both methods have limitations. In eye tracking experiments, participants can read freely and adapt to the characteristics of the text. EEG measures have an excellent temporal resolution and provide information about processes in behaviorally mute epochs. However, for technical reasons, most EEG studies have participants read sentences in a rather unnatural, word-by-word fashion (rapid serial visual presentation: RSVP). This allows researchers to avoid a number of problems related to natural reading in EEG, the most prominent of which are artifacts induced by eye movements. Despite this obstacle, combining both methods appears to be an obvious way to overcome their respective weaknesses. Indeed, recent work using ERPs in natural reading has shown that the technical problems can be handled and that effects observed in RSVP are replicated in natural reading studies (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Kretzschmar, Bornkessel-Schlesewsky, & Schlesewsky, 2009; Hutzler et al., 2007). The current data show that, despite the general feasibility of the approach,

RSVP and natural reading elicit qualitatively different results when it comes to oscillatory brain dynamics.

Most language-related EEG research is based on ERPs. Because single-trial EEG is dominated by noise, participants are exposed to many instances of the same kind of stimulus. Subsequently, the data are averaged across trials for each condition. The underlying assumption is that the brain response to the stimulus is present in each trial whereas unrelated signals will be canceled out in the averaging process. The resulting effects can be categorized by polarity (negative-going or positive-going), onset, offset, peak latency, and distribution on the scalp. The most widely reported ERP signature in psycholinguistic research is the N400 effect (Kutas & Hillyard, 1980), a negative deflection at centroparietal electrodes, ranging approximately from 300 to 500 msec. It is sensitive to the effort required to process a word and is typically seen in response to violations involving word meaning. For instance, *socks* in *He spread the warm bread with socks* elicits a larger negativity than *butter* in the same position. The N400 effect is not restricted to semantic violations but can also be observed following statements that do not match common world knowledge (Hald, Steenbeek-Planting, & Hagoort, 2007; Hagoort, Hald, Bastiaansen, & Petersson, 2004). The amplitude of the effect also varies as a function of word frequency (Dambacher, Kliegl, Hofmann,

[1]Universität Potsdam, [2]University of California, San Diego, [3]Universität Hamburg

& Jacobs, 2006; Van Petten & Kutas, 1990) and predictability (Dimigen et al., 2011; Dambacher et al., 2006).

The N400 effect has been replicated successfully in natural reading situations. Kretzschmar et al. (2009) used graded antonyms like *The opposite of black is white/yellow/nice* and found an N400 effect with a peak around 300 msec for both unpredicted completions (*yellow* and *nice*). Dimigen et al. (2011) found an N400 effect with a centroparietal distribution and a peak at 384 msec (at electrode Pz) for low-predictability words. Dimigen et al. also reported earlier effects with a similar topography, but those were not statistically significant.

Time–frequency representations (TFRs) provide an alternative framework for EEG analysis. Like any signal that varies over time, the EEG can be decomposed into a number of oscillations of different wavelengths and phase shifts. This can be done with a fast Fourier transform (FFT; Cooley & Tukey, 1965), wavelet analysis (Schiff, Aldroubi, Unser, & Sato, 1994), or multitapering (Mitra & Pesaran, 1999). The power of particular frequency ranges in the resulting spectrum is informative about the underlying cognitive processes. For instance, increased theta waves in the range from 4 to 7 Hz and decreased alpha waves in the range from 8 to 13 Hz are often observed in the context of memory processes (Klimesch, 1999).

A number of studies have investigated the spectral signature of the conditions that evoke an N400 effect. Frisch and Schlesewsky (2001) investigated the interaction of grammaticality and animacy using sentences like *Paul fragt sich, welchen Angler der Jäger gelobt hat* (Paul asks himself which angler [ACC] the hunter [NOM] praised has) and *Paul fragt sich, welcher Angler der Jäger gelobt hat* (Paul asks himself which angler [NOM] the hunter [NOM] praised has), where two nouns are case-marked as subject, rendering the sentence ungrammatical; *Zweig* (twig) replaced *Jäger* for the inanimate conditions. They observed an N400 following grammaticality violations only if both noun phrases were animate. Roehm, Schlesewsky, Bornkessel, Frisch, and Haider (2004) reanalyzed the data (*n* = 16) from Frisch and Schlesewsky (2001) and reported an N400 effect that was not described in the original paper. It appeared when an inanimate noun phrase with subject case followed an animate noun phrase with object case. Roehm et al. investigated oscillations in the delta and theta band (1–7.5 Hz) for the two N400 effects that looked similar in the time domain analysis and reported increased power in the upper theta band (6–7.5 Hz) for inanimate versus animate conditions and increased power in the lower theta band (3.5–5 Hz) for ungrammatical versus grammatical conditions. Both ungrammatical conditions showed increased power in the delta range (1–3.5 Hz) in comparison to the animate grammatical condition, but the inanimate grammatical condition did not. Roehm et al.'s results thus show that a manipulation that elicits similar effects in the time domain can have entirely different responses in the frequency domain.

Hagoort et al. (2004) used ERP and TFR analyses to investigate the access to semantic knowledge and world knowledge in language processing. They compared the response to an adjective in semantically correct sentences (*The Dutch trains are yellow and very crowded*), in semantically ill-formed sentences (*The Dutch trains are sour and very crowded*), and in sentences that were semantically valid but incongruent with world knowledge (*The Dutch trains are white and very crowded*). The ERPs for the different violation types were virtually identical, but the traces in the oscillatory response were distinct. Both violations elicited an increase in theta band power, although this effect was stronger in semantic violations, whereas only world knowledge violations led to a marked increase in gamma power. Thus, the superficial similarity of the ERPs across conditions does not necessarily entail identical responses in the TFR.

Hald, Bastiaansen, and Hagoort (2006) conducted another experiment with semantic violations as in Hagoort et al. (2004). They found a stronger power increase at bilateral temporal electrodes in the theta range (3–7 Hz) and an increase in gamma power (around 40 Hz) at right frontal electrodes in semantically ill-formed sentences. In a previous study by Bastiaansen, van der Linden, ter Keurs, Dijkstra, and Hagoort (2005), theta activity at temporal sites has been linked to lexico-semantic access, which is consistent with Hald et al.'s (2006) results.

A number of other studies have investigated the use of world knowledge in online language processing. They cannot be reviewed here because of space restrictions but attest to how reliably these kinds of violations elicit behavioral and electrophysiological effects (see, e.g., Menenti, Petersson, Scheeringa, & Hagoort, 2009; Hald et al., 2007; Warren & McConnell, 2007; Chwilla & Kolk, 2005; Rayner, Warren, Juhasz, & Liversedge, 2004).

Using RSVP, Hagoort et al. (2004) and Roehm et al. (2004) reported varying spectral results in the context of superficially similar ERP effects. This raises the question whether the close correspondence of ERP effects in RSVP and natural reading translates into an analogous relation of oscillatory dynamics in the different presentation modes. We conducted two experiments where participants read world knowledge violations as in Hagoort et al. (2004). In Experiment 1, sentences were presented in a word-by-word fashion. In Experiment 2, participants read the same sentences in a natural reading setting while their eye movements and their EEG were recorded. If the spectral compositions of the EEG in serial presentation and natural reading are comparable, effects in the theta and gamma range as in Hagoort et al. (2004) should emerge in both experiments.

EXPERIMENT 1

Methods

Participants

We collected data from 32 self-reportedly right-handed members of the University of Potsdam student population

(24 women, 8 men). They had normal eyesight or wore corrective lenses and were between 19 and 49 years old ($M = 26$). They were not told what the study was about. Written consent was collected from all participants, and they were compensated with course credit or money.

Materials

The experimental material comprised 120 minimal pairs of German sentences with a control condition and a version that was incongruent with common world knowledge. For instance, *Paris is the capital of France* would be a control sentence that is consistent with common world knowledge, whereas *Rome is the capital of France* is not. The critical word *France* was held constant across conditions and an earlier part of the sentence was manipulated to render the sentence incorrect (*Rome* instead of *Paris*). The sentences had an average length of 7.6 words ($SE = 0.2$), and the critical word had an average length of 8.1 characters ($SE = 0.2$). The items were split into two lists in a latin-square design and pseudorandomized. Each participant thus saw only one version of each item. Another 180 items from an unrelated experiment on sentence processing were interleaved with the material for the current study. They were on average 18.1 words ($SE = 0.04$) long.

Procedure

Participants signed a consent form at the beginning of the session and were seated in a shielded booth approximately 60 cm from the stimulus display. After the electrode cap was prepared, participants read five practice sentences to familiarize themselves with the procedure. When they had finished the practice trials, they proceeded with the experiment. For Experiment 1, we adopted the presentation procedure described in Hagoort (2003). Sentences were presented word by word in the center of a screen with a resolution of 1680×1050 pixels. Each word was presented for 300 msec in 28-point Arial, followed by a blank display that lasted 300 msec. Next, a blank display with a variable duration between 1 and 2 sec preceded a comprehension question that was answered with the press of a button. Another 1150 msec intervened between the response and the onset of the next trial. Every 20 sentences, participants received feedback about their performance in the comprehension questions of the last block. After 90, 180, and 270 sentences, they took a short break. An experimental session lasted for approximately 2 hr, including preparation and debriefing.

Recording and Analysis

Recording. The EEG was recorded from 32 Ag/AgCl electrodes mounted in a 10-20 design (Jasper, 1958) in a shielded electrode cap (Advanced Neuro Technology, Enschede, Netherlands). Eye movements and blinks were

monitored with additional bipolar electrodes on the left and right outer canthus and the infraorbital ridges of the right eye. Both EEG and EOG were recorded with a low-pass filter with a cutoff at 138.24 Hz and digitized at a sampling rate of 512 Hz. Recordings were initially referenced to the left mastoid and later converted a common average reference. Impedances were kept between 5 and 10 kΩ at all times.

Preprocessing. The EEG data were preprocessed in BrainVision Analyzer (Brain Products, Munich, Germany) where the signal was first resampled to 500 Hz and filtered (0.3–100 Hz band-pass, 50 Hz notch). Eye movements were corrected with an Independent Components Analysis (ICA; Jung et al., 2000), the specifics of which are described in the methods section of Experiment 2. The corrected signal was segmented from −1000 to 2000 msec relative to stimulus onset. Segments with muscle artifacts or slow drifts were discarded, which led to the loss of 70 trials (1.9%). All further processing steps and analyses were performed in *R* (R Core Team, 2013), where the data were baseline-corrected relative to a 100-msec interval preceding the stimulus for the ERP analyses.

For the spectral analysis, Fourier transformations were performed for overlapping windows of 1000-msec length in 10-msec steps from fixation onset until 2000 msec following the fixation. This resulted in 100 frequency spectra for each trial and electrode that were related to a prestimulus baseline window of 1000 msec length by calculating the change in power for each frequency in decibels. To avoid spectral leakage at the window edges, each window was filtered by means of a Tukey window (Tukey, 1967) with symmetric Hann functions in the rising and falling part, comprising 10% of the window length. The FFT does not produce power estimates for single time points but for a time span so latencies and durations for spectral effects are reported relative to these windows. The onset of the first time window in which an effect appears is treated as its onset. The duration of an effect is the time from its onset until the onset of the last time window wherein the effect is present. Thus, an effect from 200 to 400 msec is an effect that begins in the time window from 200 to 1200 msec and ends in the time window from 400 to 1400 msec.

We analyzed the EEG with a cluster-based random permutation procedure (Maris & Oostenveld, 2007). Although we anticipated an ERP effect with a centroparietal distribution and a peak around 400 msec, prior research has shown that the timing and topography of an effect can differ slightly between natural reading and serial presentation (Dimigen et al., 2011; Kretzschmar et al., 2009). The cluster-permutation test offers an elegant approach to finding effects without distributional assumptions while controlling the multiple-comparison problem. The test was implemented as follows. First, for each electrode and time point, a pairwise *t* test of the two conditions (congruent vs. incongruent) was performed with one data point per

participant and condition. Next, spatiotemporal clusters of sample statistics above a certain threshold were formed with connected components labeling (Samet & Tamminen, 1988).[1] The t values of all samples in a cluster were then summed up to yield a cluster statistic. To assess a cluster's significance, conditions were repeatedly and randomly swapped within participants and the cluster-permutation procedure was then performed on the resulting data set. From each of 5000 iterations, the maximal cluster statistic was used to create a distribution representing the null hypothesis. We considered a cluster significant if its cluster statistic fell in the lower 2.5th or upper 97.5th percentile of this distribution.

For power spectra, separate cluster-permutation tests for discrete frequency ranges were performed that were defined as follows: delta (1–3 Hz), theta (4–7 Hz), lower alpha (8–10 Hz), upper alpha (11–13 Hz), beta (14–30 Hz), and gamma (31–70 Hz). Within each frequency range, the average power change for each participant, electrode, time window, and condition entered into the analysis as described above for the ERP. Although it is common practice to relate frequency ranges to each participant's individual alpha frequency, this was not done in the current study. Because Hagoort et al. (2004) used fixed frequency bands, such a step would have introduced another difference between the studies, further complicating any comparison.

Results

Behavioral Data

Participants fared well on the comprehension questions, answering correctly in 92.9% of congruent and 94.7% of incongruent trials. The difference between conditions was not significant in a paired t test on subject averages ($t(31) = 1.83, p > .08$). All trials were used for the analysis, regardless of whether or not the response was correct.

ERP

The cluster-permutation test showed an increased negativity with a centroparietal distribution from 238 to 810 msec (peak at 420 msec) in incongruent trials. Both timing and topography of this cluster are indicative of the anticipated N400 effect. Another cluster in the same time window indicated a positivity at frontal electrodes (322–532 msec, peak at 386 msec) and appeared to be the result of the common average reference. Following this cluster, there were four successive positivities at centroparietal to parietal electrodes (576–742 msec, peak at 702 msec; 752–942 msec, peak at 906 msec; 1142–1204 msec, peak at 1182 msec; 1322–1380 msec, peak at 1350 msec). Lastly, there were two short-lived late negativities at frontopolar and frontal electrodes (1150–1224 msec and 1316–1398 msec, peaks at 1190 and 1362 msec; Figure 1).

TFR

There was one marginally significant cluster in the TFR ($p < .06$). It indicated stronger synchronization in incongruent trials than in congruent trials in the theta range from 0 to 460 msec (peak at 260 msec) at frontal to frontocentral electrodes and was right lateralized. There were no other significant or marginally significant clusters in the TFR (Figure 2).

Figure 1. (A) Grand-averaged ERP for electrodes Cz (N400) and CP5 (late positivity) in Experiment 1, low-pass-filtered at 30 Hz. Solid lines show congruent amplitudes; dashed lines show incongruent amplitudes. (B) Topographic map of amplitude differences (incongruent minus congruent) in the N400 time window in Experiment 1 (300–500 msec). (C) Topographic map of amplitude differences (incongruent minus congruent) in the time window of the late positivities in Experiment 1 (550–1200 msec). The electrode from the respective waveform plot is underlined in the topographic map.
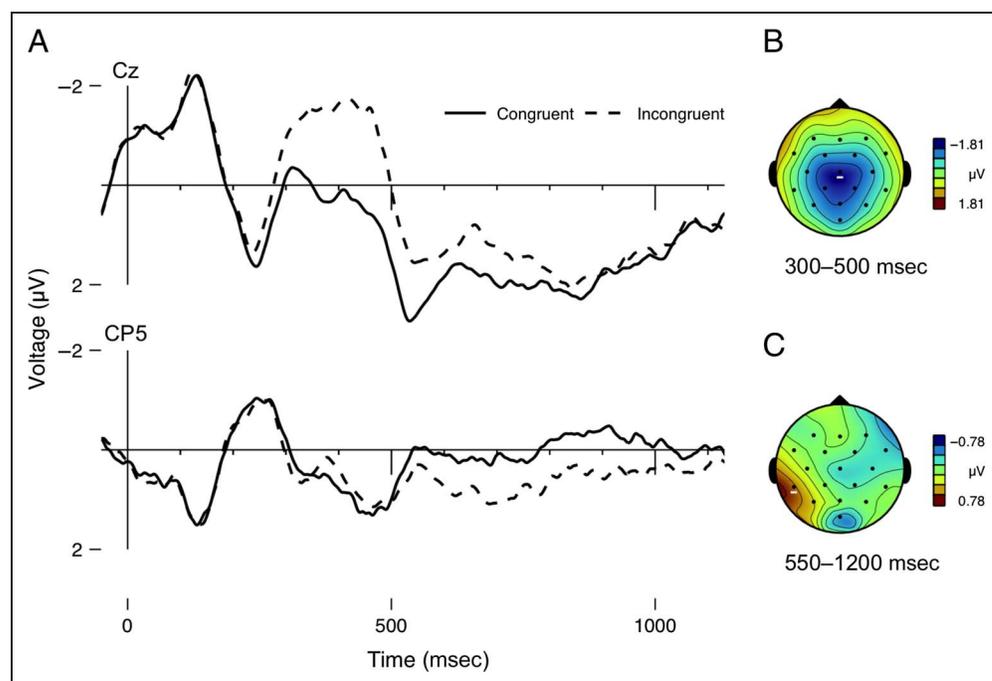
Figure 2. (A) Time–frequency plot of power changes in congruent and incongruent trials at electrode Fz in Experiment 1 (frequencies up to 10 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the theta range (4–7 Hz) in the time window of the theta effect in Experiment 1 (0–460 msec). The electrode from the time–frequency plot is underlined in the topographic map.
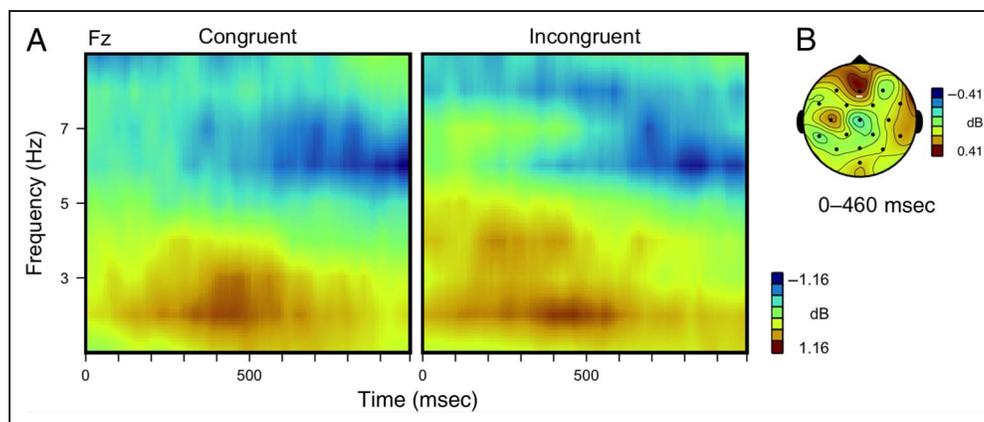
Discussion

In the ERP, Experiment 1 successfully replicated the N400 effect from Hagoort et al. (2004): There was a relatively larger negativity following world knowledge violations at centroparietal electrodes with a peak around 420 msec. In the TFR, a power increase in the theta range at frontal electrodes was larger in incongruent than in congruent trials. Although this effect was statistically only marginally significant, both the frequency range and direction of the effect are consistent with Hagoort et al. (2004). These results provide a validation of our materials and experimental setup and set the stage for Experiment 2.

EXPERIMENT 2

Experiment 1 provided a validation of the German materials and the general setup by partially replicating the N400 effect and the theta power increase in Hagoort et al. (2004). In Experiment 2, we tested the same sentences in a natural reading setting with concurrent eye movement and EEG recordings to answer the question outlined in the Introduction.

Methods

Participants

Fifty-two participants (37 women, 15 men) were recruited from the University of Potsdam student population (19–34 years, $M = 25$). All participants were right-handed by self-report, native speakers of German, and had normal or corrected-to-normal vision. They had not participated in Experiment 1 and were naive with regard to the aims of the study. They gave written informed consent to the procedure and received either course credit or money for compensation. Because of recording errors, we discarded all data from one participant and the EEG data from another two participants. This left eye-tracking data from 51 and EEG data from 49 participants. To obtain a consistent data set for comparing eye movements and EEG, we used only the data from the 49 participants with

a complete set of observations (i.e., eye movements and EEG).

Materials

The materials were the same as in Experiment 1. The items were presented together with an equal number of items for an unrelated experiment on sentence processing. Because of the nature of the other experiment, those sentences were more complex and longer than the sentences for the present experiment. Every sentence consisted of a main clause and a subordinate clause and had an average length of 22.4 words ($SE = 0.2$).

Procedure

Aside from a few details, the procedure was identical to the procedure in Experiment 1. After the electrode cap was prepared, the eye tracker was calibrated. Sentences were not presented word by word in the center of the display but left-justified, vertically centered on a single line in 26-point Arial; the display had a resolution of 1680×1050 pixels. To finish a sentence, participants fixated the bottom right corner of the display. We chose this method over a button press to prevent anticipatory eye movements to the beginning of the sentence. That way, regressions from the end of the sentence could be safely linked to the processing of that region. The sessions in Experiment 2 were slightly longer than in Experiment 1 at around 2.5 hr.

Recording and Analysis

Eye movements. Fixational eye movements were recorded with a desktop-mounted EyeLink 1000 (SR Research, Mississauga, ON, Canada) in remote mode. This allowed participants to sit comfortably without a chin rest, which reduced myogenic artifacts. Gaze position was sampled at 500 Hz from the right eye with a spatial resolution

Metzner et al. 1021

of 0.01° and an average accuracy of 0.54° in the vertical center of the screen.

We excluded fixations shorter than 20 msec and longer than 1200 msec from analysis, which led to the loss of 2.03% of all fixations. The remaining fixations were aggregated into the standard fixation measures: first fixation duration, gaze duration, and regression probability. First fixation duration is the duration of the first progressive fixation on a word (i.e., coming from the left). Gaze duration is the duration of all fixations on a word from the first progressive saccade until the eye leaves the word again. Regression probability denotes the probability to make a regressive saccade from a word immediately after entering it with a progressive saccade (i.e., before leaving it to the right).

Fixation measures were analyzed with linear mixed-effects models using the package *lme4* (Bates, Maechler, & Dai, 2009) in *R*. The critical word (*France*) and the word preceding it (*of*) were analyzed if and only if they received a progressive fixation. Trials without a progressive fixation on any of these words were discarded. Before analysis, it was determined whether a variable had to be transformed to afford normality of the residuals and which transformation would stabilize variance (Box & Cox, 1964). Following this procedure, all eye movement analyses were performed on log-transformed duration variables. Binary responses (regression: yes or no) were analyzed using the logit link in a generalized linear mixed-effects model. All models were fit with varying intercepts and slopes for each fixed factor, including a correlation term, unless there was a failure to converge or the correlation estimated implied a degenerate variance covariance matrix; in such cases, the model was simplified until the model converged and had no degeneracy.

EEG. The EEG was recorded with the same setup as in Experiment 1, but was measured to the common average reference already online. Like in Experiment 1, an ICA was used to identify brain activity related to horizontal and vertical eye movements. The ICA used a biased variant of the Infomax algorithm and was trained on the filler sentences. By means of this, it was ensured that eye movements from sentence reading featured in the training data but effects that were time-locked to fixational eye movements in the critical sentences were not systematically removed. A classic PCA was used for preparatory sphering. All channels were included except for the mastoid electrodes. Components with a frontopolar or bipolar frontal distribution were identified and removed from the signal because those components were assumed to represent vertical and horizontal eye movements, respectively. The components were related to the HEOG and VEOG by ascertaining that activity in those channels was minimized through the procedure. As another post hoc validation, we computed the ERPs and TFRs of the signal before correction and submitted them to the same analyses as the corrected signal. We found neither N400 effects nor theta or gamma increases in these data. The success of the correction procedure is illustrated in Figure 3. After correction, trials with severe artifacts were removed from the data, which led to the loss of 154 trials (3.1%). In an offline procedure, the first progressive fixation on a word was identified from the eye-tracking data, and its time stamp was aligned with the EEG data with the help of synchronization markers at the beginning and end of each trial. In 18.5% of the experimental trials, the target word did not receive a progressive fixation. All following segmentation and preprocessing was the same as in Experiment 1.

Results

Behavioral Data

Participants scored high on the comprehension questions with 92.7% accuracy in congruent and 95.0% in incongruent
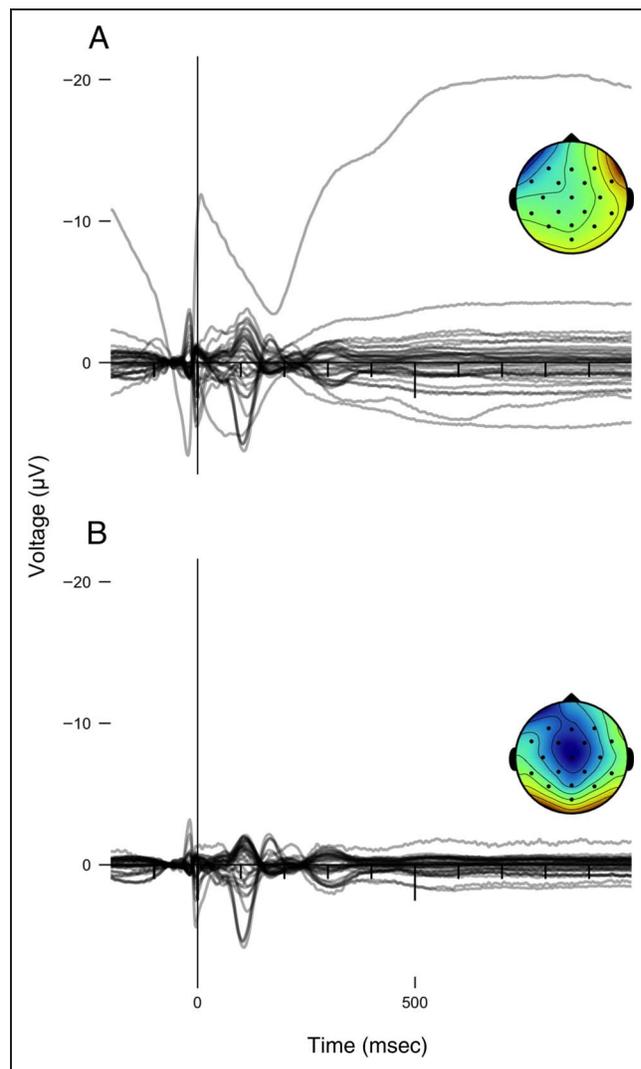
Figure 3. Grand-averaged ERP for randomly selected fixations from Experiment 2 before (A) and after (B) artifact correction. Topographic maps show mean amplitude in the first 500 msec following the fixation.

trials. The difference in accuracy between conditions was significant in a paired *t* test (*p* < .05), with lower accuracy in congruent trials.

Eye Movements

The eye movement record for the critical word exhibited clear effects of world knowledge incongruence (see Figure 4). Participants had longer first fixation durations (265 vs. 278 msec, Est. = 0.03, *SE* = 0.01, *t* = 4.75), gaze durations (351 vs. 379 msec, Est. = 0.05, *SE* = 0.01, *t* = 5.02), and regression probabilities (35.8 vs. 39.0%, Est. = 0.10, *SE* = 0.04, *z* = 2.58) in incongruent trials.

Kretzschmar et al. (2009) reported no effect in the eye movement record for the precritical word. To check whether this was also the case in this study, eye movements on the precritical word were also analyzed. As Figure 4 shows, there were no effects on the precritical word in the fixation record and the fixation measures diverged only after the critical word.

ERP

The ERP time-locked to the first fixation on the critical word contained a significant cluster from 222 to 514 msec, peaking at 378 msec (Figure 5). It indicated a relative negativity in incongruent trials with a right lateralized, occipitoparietal distribution. As in Experiment 1, a positivity at frontal to frontocentral accompanied the N400 effect. It ranged from 318 to 626 msec (peak at 476 msec). A second positivity started at 692 msec and lasted until the end of the analysis window at 1400 msec (peak at 1382 msec, centered around CP1).

As in Kretzschmar et al. (2009), there were effects in the ERP on the precritical word. Similar to the ERP on the critical word, there was a centroparietal negativity from 334 to 826 msec and a simultaneous positivity at frontal electrodes. Both effects occurred slightly later than on the critical word with peak latencies of 608 and 658 msec, respectively. However, parafoveal preview was not controlled (e.g., via a boundary paradigm) and the EEG from precritical and critical word may have overlapped because of the short time interval between the fixation on the precritical and the critical word (295 msec on average). In other words, the effects observed on the precritical word may actually have been in response to the critical word, not the precritical word. Thus, any interpretation of the effects on the precritical word would be purely speculative.

TFR

There were significant TFR effects in the delta range and in the upper alpha range. No other frequency range showed significant differences.

Delta range (1–3 Hz). Power in the delta range increased at central sites in time windows from 0 to 990 msec
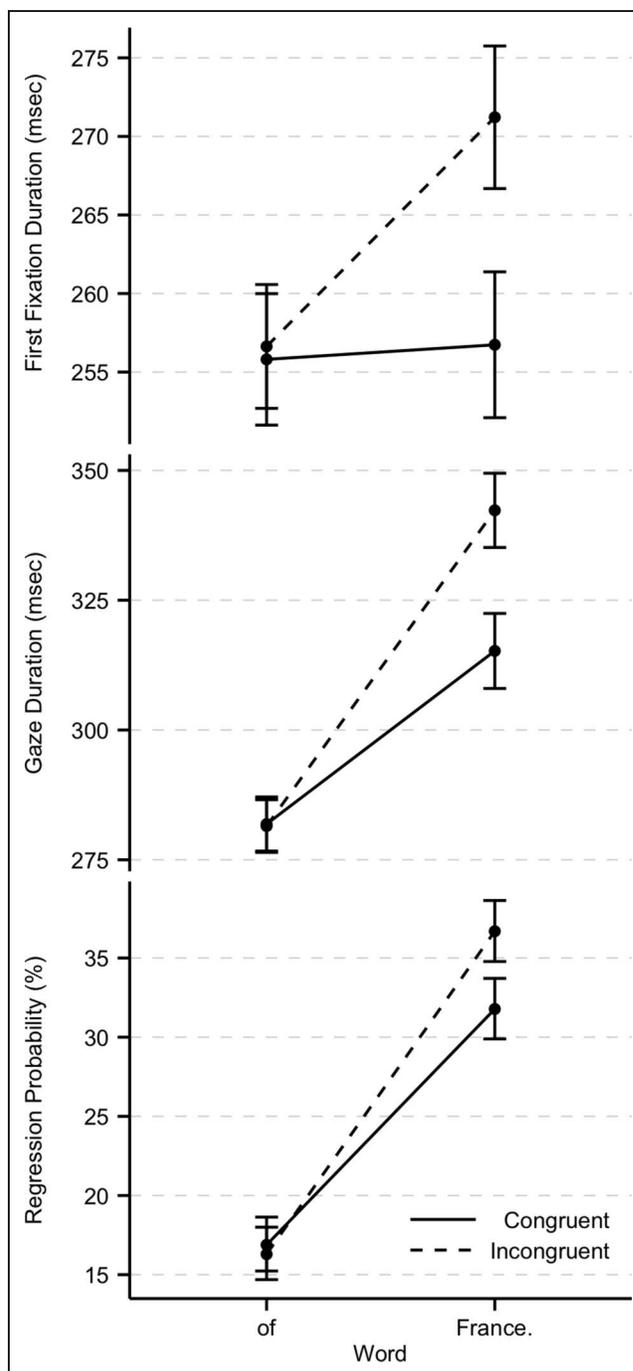
Figure 4. First fixation duration, gaze duration, and regression probability on the precritical (*of*) and critical word (*France*) in Experiment 2 as a function of world knowledge congruence: solid lines for congruent and dashed lines for incongruent words. Error bars are 95% confidence intervals; the plotted data are partial effects from linear mixed-effects models (after removing effects due to different participants and items).

(peak at 500 msec) following the first fixation on the critical word (Figure 6). This synchronization was significantly larger in incongruent trials than in congruent trials. The center of this spatiotemporal cluster was at electrode Cz.
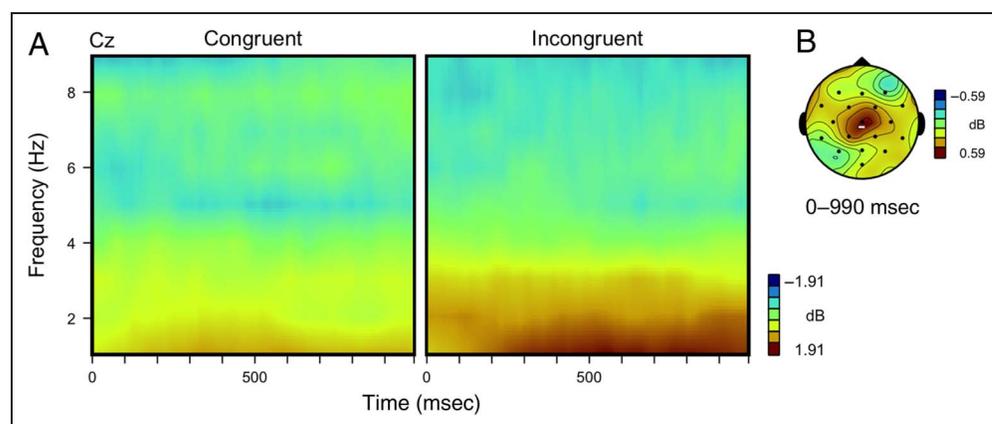
Figure 5. (A) Grand-averaged ERP for electrodes Pz and CP1 in Experiment 2, low-pass-filtered at 30 Hz. Solid lines show congruent amplitudes; dashed lines show incongruent amplitudes. (B) Topographic map of amplitude differences (incongruent minus congruent) in the time window of the occipitoparietal negativity in Experiment 2 (200–500 msec). (C) Topographic map of amplitude differences (incongruent minus congruent) in the time window of the late positivity in Experiment 2 (700–1200 msec). The electrode from the respective waveform plot is underlined in the topographic map.
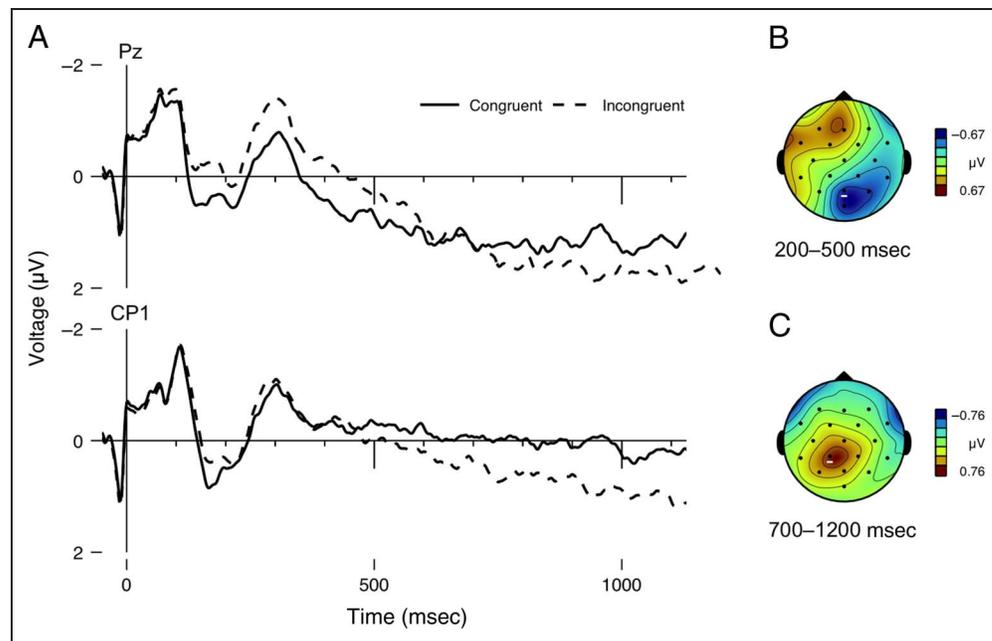
Upper alpha range (11–13 Hz). A second cluster indicated desynchronization in the upper alpha range at occipitoparietal electrodes around POz (Figure 7). It started immediately after the fixation on the critical word, lasted until 760 msec postfixation (peak at 520 msec), and was larger for incongruent than for congruent trials.

Single-frequency analysis. The specific choice of 1–3 Hz for the delta and 3–7 Hz for the theta band may have been responsible for the presence of a delta effect and the lack of a theta effect. To rule out this possibility, cluster-permutation tests were performed for single frequencies in the range from 1 to 7 Hz. This yielded clusters at 1, 2, and 5 Hz. The effects in the delta range had a central to frontocentral distribution; the effect at 5 Hz ranged from frontal to centroparietal electrodes. It is due to a small power decrease in congruent trials and an equivalent power increase in incongruent trials. There were no other significant single-frequency clusters.

Discussion

Experiment 2 replicated the N400 effect in the time domain from Experiment 1 and Hagoort et al. (2004). In addition, we found a late positivity at centroparietal electrodes that resembled the late positive shift reported by Hald (2003).[2] In the TFR, we found effects in the delta and upper alpha range but no effects in the theta or gamma range, which is at odds with prior results. We discuss potential explanations for the different results in Experiment 1 and 2 below.

GENERAL DISCUSSION

We conducted two experiments to compare ERPs and the oscillatory dynamics of the EEG in word-by-word presentation and natural reading. In both experiments, participants read single sentences that were either congruent (e.g., *Paris is the capital of France*) or incongruent with

Figure 6. (A) Time–frequency plot of power changes in congruent and incongruent trials at electrode Cz in Experiment 2 (frequencies up to 8 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the delta range (1–3 Hz) in the time window of the delta effect in Experiment 2 (0–990 msec). The electrode from the time–frequency plot is underlined in the topographic map.

Figure 7. (A) Time–frequency plot of power changes in congruent and incongruent trials at electrode POz in Experiment 2 (frequencies between 6 and 14 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the upper alpha range (11–13 Hz) in the time window of the upper alpha effect in Experiment 2 (0–760 msec). The electrode from the time–frequency plot is underlined in the topographic map.
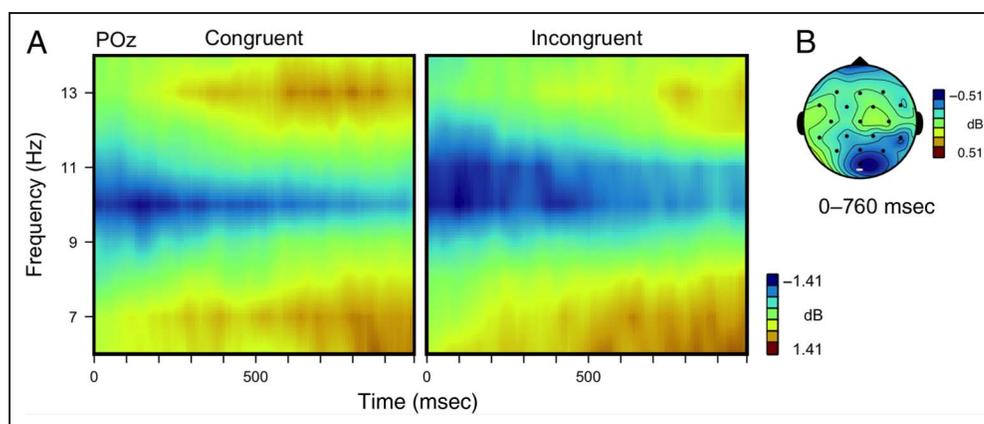
common world knowledge (*Rome is the capital of France*). Such a manipulation elicited an N400 effect and increased activity in the theta and gamma range in prior studies using RSVP (Hald et al., 2007; Hagoort et al., 2004).

Experiment 1 used a word-by-word presentation paradigm and successfully replicated the N400 effect in the ERP found by Hagoort et al. (2004). In Experiment 2, the analyses of eye movements and ERP also revealed the anticipated patterns. The critical word in incongruent trials led to increased first fixation durations, gaze durations, and regression rates. This is consistent with the results from Dambacher and Kliegl (2007), who observed the same pattern in fixation durations in an experiment on word frequency and predictability with eye movements and EEG obtained from different participants. Additionally, in line with Hagoort et al. (2004) and Hald et al. (2007), there was an increased negativity at parietal electrodes with a peak around 400 msec following incongruent words in both Experiment 1 and Experiment 2. This lends further support to the assumption that ERPs from RSVP and natural reading yield similar results.

Experiment 1 confirmed the power increase in the theta range from Hagoort et al. (2004). In Experiment 2, the TFR analysis revealed two different effects. Compared with a prefixation baseline, power in the delta range increased and power in the upper alpha range decreased following the first fixation on the critical word. Both effects were significantly larger in incongruent trials. This is at odds with prior results and requires an explanation.

The effect in the delta range is difficult to interpret because, with few exceptions, synchronization in the delta range has not yet been reported in language-related studies. As noted above, Roehm et al. (2004) described increases in evoked and whole delta power at electrode Pz in grammaticality violations. Roehm, Bornkessel-Schlesewsky, and Schlesewsky (2007) found increased delta power in three experiments. Using graded antonyms like Kretzschmar et al. (2009), they reported an N400 and a P600 in the ERP. In the TFR, they found increased delta and theta power in the N400 time window and increased delta power in the P600 time window. In the sec-

ond experiment, word pairs were presented without a sentence context. The second word (with the first word being *black*) could be a valid antonym (*white*), a related word (*yellow*), an unrelated word (*nice*), or a pseudoword. In comparison to valid antonyms, all word pairs elicited an N400 but no P600. In the TFR, unrelated words elicited more power in the lower theta range than valid antonyms. Pseudowords induced larger delta power in comparison with both valid antonyms and unrelated words. In the third experiment, Roehm et al. found increased delta power in object-initial versus subject-initial clauses and the opposite pattern in the theta range. Although the manipulations in these four experiments are different from the world knowledge violations in this study, the results show that delta power is sensitive to higher cognitive processes such as language processing.

In cognitive domains outside of language processing, the delta range has received somewhat more attention (for a review, see Harmony, 2013). Harmony, Alba, Marroquín, and González-Frankenberger (2009) reported increased delta power at frontal electrodes in the no-go condition of a go/no-go task and linked increased delta activity with the inhibition of movement. Moreover, Knyazev (2007) concluded that power changes in the delta and alpha range are inversely related but ascribed inhibitory processes to the alpha range. A simultaneous increase in delta power and decrease in alpha power is the very pattern found in Experiment 2 where readers may have inhibited a planned eye movement and engaged in memory retrieval upon encountering an unexpected noun. Thus, there is evidence for an involvement of the delta range in sensorimotor control as well as higher cognitive processes such as language processing.

In contrast to synchronization in the delta range, desynchronization in the alpha range and synchronization in the theta range are well-described oscillatory brain responses (Klimesch, 2012). Both have been observed in the context of memory-demanding tasks (Klimesch, 1999; Klimesch, Schimke, & Schwaiger, 1994) and appear to play a role in language processing (e.g., Roehm, Klimesch, Haider, & Doppelmayr, 2001). Roehm et al. (2001)

investigated the involvement of the two frequency ranges in language processing while participants read sentences in four chunks. The crucial manipulation was whether or not they had to name the superordinate concept for a probe word in the penultimate chunk (e.g., "bird" for "sparrow"). In trials with that additional task, Roehm et al. observed a stronger power decrease in the upper alpha band at occipital and frontal electrodes; the theta band did not show qualitative differences between the two conditions. Roehm et al. concluded that theta oscillations reflect domain-general working memory processes whereas the upper alpha band is sensitive to linguistic processes. This may also be the reason why activity in the theta band was more prominent in Experiment 1 where sentences were presented word by word. This presentation modality may have imposed a higher load on the working memory system than natural reading.

Additional evidence for an involvement of theta and alpha oscillations in language processing comes from a study on open- and closed-class words (Bastiaansen et al., 2005). In prior studies using ERPs, open-class words elicited a larger N400 than closed-class words (King & Kutas, 1995; Van Petten & Kutas, 1991) and a frontal negative shift (Brown, Hagoort, & ter Keurs, 1999). Bastiaansen et al. (2005) reported a stimulus-evoked power decrease in the alpha and beta band as well as a power increase in the theta band that were larger following open-class words. The theta increase was strongest at left temporal and occipital electrodes and the alpha desynchronization at right occipito-temporal electrodes. The topography of the alpha effect is particularly interesting because it matches with the distribution of the upper alpha effect in the current study. Because the alpha decrease in Bastiaansen et al. (2005) did not vary as a function of word class, they ascribed it to general sensory processing of incoming information. However, their figures show that the alpha decrease was more widespread and slightly larger at occipito-temporal sites following open-class words. The reason for this quantitative difference may be that the search for a specific entry is actually more effortful in the considerably larger set of open-class words in the mental lexicon. It also matches the current results and those of Roehm et al. (2001) where more effortful processing led to decreased alpha activity at occipital electrodes.

Hagoort et al. (2004) reported synchronization in the theta range around 5 Hz and in the gamma range between 35 and 40 Hz but no effects in the alpha band. Whereas their theta effect was markedly stronger in semantic violations but also present in world knowledge violations, the gamma effect was restricted to the latter. From this divergence, Hagoort et al. concluded that the two violation types are treated differently on a neuronal level. The theta effect is likely to reflect relatively more effortful memory access in semantically incongruent trials. The nature of the gamma effect, however, is less clear. Hagoort et al. note that gamma oscillations have been ascribed to integrative processes in both local and distributed neural networks but do not explain why this would be more important in world knowledge violations than in semantic violations.

The experimental designs of Hagoort et al. (2004) and this study are almost identical, so it is not clear why we did not find an increase in the gamma range. Note that the current materials and Hagoort et al.'s (2004) differ in at least three aspects. First, the current studies were run in German. Second, the critical word was the same in congruent and incongruent sentences of the current study, whereas an earlier word in the sentence was varied. In contrast, Hagoort et al. had varied the critical word itself. Finally, the critical word was sentence-final in 90 of 120 items in the current study whereas Hagoort et al. (2004) made sure that this was never the case. Wrap-up effects at the end of the sentence may have introduced a higher noise level, which may have attenuated effects in the gamma range. However, because Experiment 1 replicated the N400 effect and the power increase in the theta range despite all these differences, it is not clear why they would lead to a selective attenuation or deletion of the gamma effect.

Another potential source of the differences lies in the way to compute power changes in the EEG. In the current study, we obtained power spectra with windowed FFTs. The choice of an FFT over wavelet analysis is not likely to have a substantial influence on the results because both approaches are formally equivalent (Bruns, 2004). Because of the requirements of the FFT, however, we used the same baseline window from 1000 msec preceding fixation onset to fixation onset for all frequencies. Hagoort et al. (2004) used a wavelet transform and analyzed power changes relative to a baseline window from 150 msec preceding stimulus onset to stimulus onset. For wavelets with slower core frequencies that did not fit into that window, it was extended to the left and right.[3] To check whether the different baselines may be responsible for the diverging results, we repeated all analyses with a baseline window from −575 to 425 msec. This resulted in slight changes in the timing of the TFR effects but had no qualitative impact on the results. Thus, the difference in baselines is apparently not responsible for the lack of a theta or gamma increase in Experiment 2.

The specific choice of 1–3 Hz for the delta range and 4–7 Hz for the theta band may also have obscured effects around the boundary between the two ranges. To exclude this possibility, we performed a post hoc analysis of Experiment 2 where we submitted power changes for single frequencies to the clustering algorithm instead of averages for predefined frequency bands. We found single-frequency effects at 1, 2, and 5 Hz. Crucially, there were no effects around the boundary between the delta and theta range (i.e., at 3 and 4 Hz). The effect at 5 Hz indicated desynchronization in congruent trials and synchronization in incongruent trials. This is not consistent with Hagoort et al. (2004), who observed theta power increases with varying magnitude in all conditions. Thus,

the choice of frequency bands was apparently not the reason for the lack of an effect in the theta range, either.

Given that Experiment 1 replicated the N400 effect and the theta increase in Hagoort et al. (2004), it appears more plausible that the different results in Experiment 2 are because of the distinct reading situations. In auto-paced word-by-word presentation, participants cannot skip words or make regressive saccades. By contrast, in natural reading situations, short and highly predictable words are skipped and regressive saccades occur frequently when readers experience processing difficulties (Rayner, 1998). A possible consequence of the different demands in natural reading and RSVP is that the sentence comprehension system adapts resource allocation as well as encoding and retrieval strategies to the reading situation. A speculative account of the different oscillatory responses in RSVP and natural reading goes as follows. In experimental settings with word-by-word presentation, the sentence parser most likely builds rich representations of the currently processed material, which leads to relatively easy retrieval from working memory in the event of unanticipated or mismatching input. This memory access is possibly reflected by a power increase in the theta range. For a natural reading situation, it can be assumed that the encoding is less elaborate because earlier parts of the sentence can be reread if processing difficulties arise. The increased delta and decreased alpha activity might reflect the inhibition of progressive eye movements and increased attention (see Harmony, 2013).

The increased theta power in RSVP may also be a task effect rather than a feature of language processing. An implication is that listening studies should show similar results as they share at least two features with word-by-word presentation: Earlier material cannot be reheard, and the pace is not under the listener's control. In contrast to RSVP, listeners have to segment a continuous auditory stream. In an MEG study, Wang et al. (2012) reported an N400m as well as power decreases in the alpha and beta range in response to semantic violations. These results do not support the view that the theta increases in RSVP are solely because of readers' inability to control the presentation speed and revisit earlier material. In Kretzschmar et al. (2013), participants read longer stretches of text on different media naturally while their eye movements and brain potentials were recorded. Their mean cumulative fixation duration on a page and the absolute power in the theta range were positively correlated. Thus, theta power increases are not limited to word-by-word presentation but also occur in natural reading, which further challenges the view of theta power increases as task effects.

In summary, the current results confirm the general feasibility of recording EEG in natural reading by replicating Hagoort et al.'s (2004) N400 effect. The time–frequency analysis shows, however, that oscillatory brain dynamics are qualitatively different in natural reading and serial presentation. This may reflect differences in how representations are constructed and retrieved from mem-ory in the two presentation modes. Further experiments are necessary to delineate the differences and similarities of stimulus- and fixation-triggered brain responses.

Notes

1. The choice of this threshold does not directly affect the significance testing because it is identical for the original clusters and the bootstrapping procedure. We used a standard alpha level of .05.
2. Hagoort et al. (2004) analyzed the EEG data from Hald (2003), who analyzed a longer time window and reported a biphasic response comprising an N400 and a late positivity. Hagoort et al. focused their analysis on a shorter time window and therefore report only a monophasic N400.
3. According to correspondence with one of the authors.

REFERENCES

Bastiaansen, M. C. M., van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience, 17,* 530–541.

Bates, D. M., Maechler, M., & Dai, B. (2009). *lme4: Linear mixed-effect models using S4 classes*. R package.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, Methodological, 26,* 211–252.

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience, 11,* 261–281.

Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods, 137,* 321–332.

Chwilla, D. J., & Kolk, H. H. J. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research, 25,* 589–606.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, 19,* 297–301.

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research, 1155,* 147–162.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084,* 89–103.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 4,* 552–572.

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport, 12,* 3391.

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

Hald, L. (2003). *The integration of semantic versus world knowledge during on-line sentence comprehension.* Doctoral dissertation, University of Arizona, 2002. Dissertations Abstracts International, DAI-B 63/12, 6114.

Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language, 96,* 90–105.

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research, 1146,* 210–218.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience, 7,* 1–10.

Harmony, T., Alba, A., Marroquín, J. L., & González-Frankenberger, B. (2009). Time–frequency-topographic analysis of induced power and synchrony of EEG signals during a go/no-go task. *International Journal of Psychophysiology, 71,* 9–16.

Hutzler, F., Braun, M., Võ, M. L.-H., Engl, V., Hofmann, M., Dambacher, M., et al. (2007). Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research, 1172,* 124–129.

Jasper, H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology, 111,* 1745–1758.

King, J. W., & 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.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29,* 169–195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences, 16,* 606–617.

Klimesch, W., Schimke, H., & Schwaiger, J. (1994). Episodic and semantic memory: An analysis in the EEG theta and alpha band. *Electroencephalography and Clinical Neurophysiology, 91,* 428–441.

Knyazev, G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience & Biobehavioral Reviews, 31,* 377–395.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport, 20,* 1613–1618.

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: Concurrent EEG-eyetracking evidence from the reading of books and digital media. *PloS One, 8,* e56178.

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

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190.

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience, 21,* 2358–2368.

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal, 76,* 691–708.

R Core Team. (2013). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1290.

Roehm, D., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2007). The internal structure of the N400: Frequency characteristics of a language related ERP component. *Chaos and Complexity Letters, 2,* 365–395.

Roehm, D., Klimesch, W., Haider, H., & Doppelmayr, M. (2001). The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. *Neuroscience Letters, 310,* 137–140.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *NeuroReport, 15,* 409–412.

Samet, H., & Tamminen, M. (1988). Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10,* 579–586.

Schiff, S. J., Aldroubi, A., Unser, M., & Sato, S. (1994). Fast wavelet transformation of EEG. *Electroencephalography and Clinical Neurophysiology, 91,* 442–455.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In B. Harris (Ed.), *Spectral analysis of time series* (pp. 25–46). New York: Wiley.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18,* 380–393.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19,* 95–112.

Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., et al. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping, 33,* 2898–2912.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14,* 770–775.