

Anomalies at the Borderline of Awareness: An ERP Study

Anthony J. Sanford¹, Hartmut Leuthold¹, Jason Bohan¹,
and Alison J. S. Sanford²

Abstract

Behaviorally, some semantic anomalies, such as those used to demonstrate N400 effects in ERPs, are easy to detect. However, some, such as “after an air crash, where should the survivors be buried?” are difficult. The difference has to do with the extent to which the anomalous word fits the general context. We asked whether anomalies that are missed elicit an ERP that could be taken as indicating unconscious recognition, and whether both types elicit an N400 effect when they are detected. We

found that difficult anomalies having a good fit to general context did not produce an N400 effect, whereas control “easy-to-detect” anomalies did. For difficult anomalies, there was no evidence for unconscious detection occurring. The results support a qualitative distinction in the way the two types of anomalies are processed, and the idea that semantic information is simply not utilized (shallow processing) when difficult anomalies are missed. ■

INTRODUCTION

Certain types of semantic anomalies, related to the so-called Moses Illusion (Erickson & Matteson, 1981), demonstrate apparently shallow processing of semantic information during reading (Sanford & Sturt, 2002). The Moses Illusion emerges when readers are asked to answer the question “How many animals of each sort did Moses put on the Ark?” The answer given by many people is “two,” with them failing to notice that the Ark episode is attributed to Noah, not Moses. A similar example is Barton and Sanford’s (1993) case study of the anomaly “When an airplane crashes on a border with debris on both sides, where should the survivors be buried?” People frequently fail to notice that you do not bury survivors. It appears that what is happening during interpretation is that the intended message seems to be understood before a complete semantic analysis has been carried out (Sanford & Sturt, 2002; see also Ferreira, Ferraro, & Bailey, 2002).

A number of papers have explored such semantic anomalies, showing that they are hard to detect whether presented in an incidental detection paradigm with single examples (e.g., Hannon & Daneman, 2004; Daneman, Reingold, & Davidson, 1995; Barton & Sanford, 1993) or in an intentional monitoring setting with multiple examples (e.g., Bohan & Sanford, 2008; Hannon & Daneman, 2001; Reder & Kusbit, 1991). Furthermore, ease of detection has been shown to vary with linguistic factors, such as focus (e.g., Brédart & Modolo, 1988). Regarding an explanation of these anomalies, which we term anomalies at the borderline of awareness (*borderline anomalies* for short), it is apparent that they have a very good fit to the broader

context in which they appear. For example, in the survivors anomaly (Barton & Sanford, 1993), as a word, *survivors* fits an air crash context very well—it is relevant to the context and strongly associated with the situation of an air crash. However, at the local semantic level (*burying survivors*), the fit is poor, and at this level, the anomaly emerges. This is very different from intuitively easy-to-detect anomalies such as *John spread his bread with socks* (e.g., Kutas & Hillyard, 1980), where the fit of the anomaly to the global context appears to be poor (in a spreading stuff on bread context, *socks* is not a relevant item), and the critical word is also locally anomalous (*socks* is not, in fact, a spreadable substance).

There are two classes of explanation for missed anomalies in the hard-to-detect case, which we shall term the *shallow processing hypothesis* and the *reduced awareness hypothesis*. According to the shallow processing account, anomalies are not detected because the full meanings of the anomalous words are not retrieved and/or integrated with the representation of the discourse. Such a view is consistent with Sanford and Garrod’s (1998) Scenario Mapping and Focus Theory of text comprehension. Within this framework, new linguistic input is assumed to be used to identify a scenario, a mental representation of a specific situation that underlies the content of the linguistic input. For instance, such a situation may be an air disaster, in which a plane crashes, as presented above. The important feature of a scenario is that it is assumed to contain default information about actions, events, characters, and other props expected in the situation that the scenario denotes. So, for instance, one of the most common features of an air crash is that *there are few or no survivors: survivors* is therefore a default referent. Experimental work has shown that such highly predictable referents may be referred to at

¹University of Glasgow, Glasgow, UK, ²University of Strathclyde, Glasgow, UK

low cost by subsequent text (e.g., Garrod & Sanford, 1983). Extending this idea, Sanford and Garrod (1998, 2005) suggested that incoming content words are checked for association with the current scenario. This checking process does not use word meaning, simply being a check for whether the word is associated with the scenario in a strong statistical sense, that is, whether the word had a good or poor fit with the scenario.

We assume that words with a poor fit are afforded deeper analysis of their meaning because they are basically not predicted. This is supported by the finding that the *survivor* anomaly is much more easily detected in the context of a *bicycle crash* scenario than an air crash scenario (Barton & Sanford, 1993). Participants reported that *survivors* was a strange word to use in the context of a bicycle crash, but a good word to use in the context of an air crash. By contrast, words with a good fit to scenario receive *shallow* semantic processing (i.e., the core meaning is not necessarily fully retrieved) as indicated by failures to notice anomalies.

As an alternative to the shallow processing hypothesis, it is conceivable that the comprehension system retrieves the meaning of the anomalies and attempts to integrate the semantics of the word in question with the rest of the text. For example, in line with incremental processing accounts of language comprehension, one may assume that meaning is processed to the greatest depth possible given the information available (Carpenter, Miyake, & Just, 1995; MacDonald, Pearlmutter, & Seidenberg, 1994). However, for some reason, the fact of the anomaly may not reach awareness. This *reduced awareness hypothesis* is plausible in the light of work in the visual change detection literature. Failure to notice sometimes quite substantial changes in scenes is commonplace, and has been associated with failures to attend to parts of the scene (e.g., Simons & Levin, 1997). Despite a failure to notice the changes, however, information that a change has occurred has been shown to be sometimes available in recognition memory tasks (Hollingworth & Henderson, 2002). Furthermore, in reading, some anomalies affect eye tracking patterns, even when participants do not report their presence (e.g., incorrect homophones, like using *bair* when the correct word is *bare*; Daneman et al., 1995). Such results are more compatible with the reduced awareness idea than with the shallow processing idea, where semantic information is simply not available to the processor. At this stage, it is not clear why reduced awareness should occur in some circumstances but one might speculate that when the fit of a word to the global context is good, attention is given to other aspects of the text instead of to the processing of that word. For the moment, the reduced awareness hypothesis certainly seems like a possible alternative.

Understanding depth of processing phenomena is important for many reasons, but primarily because it relates more generally to the fundamental issue of the mechanisms underlying sentence comprehension. Failures to detect borderline anomalies have been taken as *prima facie*

evidence of shallow processing (e.g., Ferreira et al., 2002; Sanford & Sturt, 2002), but the exact way in which borderline anomalies are processed is not well explored. In fact, in only a handful of studies have on-line measurements been made to investigate some of the processes underlying borderline anomaly detection. Reder and Kusbit (1991) and Van Oostendorp and De Mul (1990) had participants read anomalous materials one word at a time using a self-paced procedure. When anomalies were missed, the general finding was that reading times were longer than when they were detected, suggesting that a failure to detect was not due to insufficient time for encoding. Unfortunately, from the reported data in these studies, it is not possible to draw conclusions about whether unconscious registration of an anomaly occurred. Daneman, Lennertz, and Hannon (2007) investigated the incidental detection of borderline anomalies while participants had their eye movements monitored. They provided evidence that detection occurred on-line during reading, without any substantial delay, although the effects obtained appeared in rerefixations rather than in first-pass reading times. They also showed that readers who detected the anomaly (detectors) spent longer rerefixating the anomaly than readers who did not and that this effect could not be attributed to detectors spending more time rerefixating in general. However, there were no data showing whether reading patterns in a non-detect state were any different from those obtained in a nonanomalous control condition. Bohan and Sanford (2008) used an intentional detection task with more materials, and found that several indices indicated disruption shortly after encountering anomalies, especially regressive eye movements. Significantly, however, there were no effects of disruption in cases where anomalies were undetected, compared to a nonanomalous baseline control. Taken at face value, this last finding offers support for the shallow processing hypothesis, rather than the reduced awareness hypothesis. However, it may still be the case that unconscious detection of unreported anomalous words occurs, but that this is simply not reflected in eye movements. Moreover, differences in how easy-to-detect and hard-to-detect borderline anomalies might be processed are not evident thus far in any analysis of eye movement behavior.

Thus, in the present article, we ask two questions. First, is the way in which borderline anomalies are processed any different from how standard, clearly detectable anomalies are processed? Secondly, is the processing of missed anomalies more consistent with the shallow processing hypothesis, or the reduced awareness hypothesis? We believe that ERPs are well suited to provide answers to these questions.

The first good reason for using ERPs is that the processes involved in language comprehension manifest themselves in qualitatively distinct ERP components (for recent reviews, see Nieuwland & Van Berkum, 2008; Kutas, van Petten, & Kluender, 2006). Thus, a centroparietally distributed negative-going deflection in the ERP with an

onset around 200 msec and a peak at about 400 msec (N400) is the brain's default response to content words. Importantly, the N400 is larger for words that are a poor fit than a good fit to sentential context (e.g., Kutas & Hillyard, 1980, 1984). It is also sensitive to semantic anomalies at the discourse level (Van Berkum, Hagoort, & Brown, 1999), to the predictability of a word within a given context (DeLong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Kutas & Hillyard, 1984), and to violations of world knowledge (Filik & Leuthold, 2008; Hagoort, Hald, Bastiaansen, & Petersson, 2004). Based on these reports, and because borderline anomalies are violations of real-world knowledge, one might expect that they should elicit a large N400.

The issue is not that simple, however. It is known that some types of anomaly elicit a late posterior positivity (LPP) in the ERP, often in the P600 range, rather than the classic N400 effect, and this finding, we shall argue, has implications for potential differences between the processing of easily detected anomalies and borderline anomalies. Although the P600 is traditionally more associated with syntactic processing difficulties (e.g., Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; for a review, see Hagoort, Brown, & Osterhout, 1999), it has been observed with violations of selection restrictions. Such an instance is exemplified by the sentence *At breakfast, the eggs ate the toast...*, where there is a late positivity in response to *ate* (e.g., Kim & Osterhout, 2005; Hoeks, Stowe, & Doedens, 2004; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; see Kuperberg, 2007, for a review). Here the selection restriction that is violated is animacy. Among the accounts of why the P600 effect occurs may be included selection restriction violations triggering conflict between syntactic parses and semantic representations (Kuperberg, 2007), and an explicit conflict monitoring between semantic and syntactic information (cf. especially Kolk et al., 2003, and Van de Meerendonk, Kolk, Chwilla, & Vissers, 2009, in this respect). Of particular interest to us, however, is that in many studies using this type of material, no N400 effect is observed, despite the obvious semantic character of the anomaly.

A similar failure to find an N400 effect occurred in a study by Nieuwland and Van Berkum (2005). For example, participants listened to a story where a passenger with a suitcase was at an airport check-in. When the story was built up so that the suitcase became a more and more salient part of the setting, having the check-in person speak to the suitcase rather than to the tourist (*Next the woman told the [suitcase]/[tourist]...*). This animacy violation elicited a posterior positivity between 600 and 1300 msec after the anomalous word. There was no N400 effect. However, when the sentence containing the anomaly was presented out of the fuller discourse context, there was an N400 effect.

The broad question of when an N400 effect and when a P600 effect is elicited by semantic anomalies is an issue of ongoing debate (see reviews by Van de Meerendonk et al., 2009; Bornkessel-Schlesewsky & Schlewsky,

2008; Kuperberg, 2007). However, we suggest that the lack of an N400 effect in the cases discussed above might be because the critical anomalous words have a good fit to global context, although they are locally anomalous. For instance, given the sentence *At breakfast, the eggs ate the toast...*, eggs and eating are both words that are used frequently in the context of breakfast—they have a good fit to global context. In contrast, classic anomalies that normally induce N400 effects have a poor fit to the general situation depicted by sentences in which they appear. Therefore, an initial and important goal of the present study is to determine the ERP correlates of borderline anomalies as compared to easy-to-detect semantic anomalies, which are known to trigger an N400 effect. Consider a typical example (1; see Methods section) in which the critical sentence may read as follows "...a 10-year sentence was given to the victim, but this was subsequently appealed." The anomaly is apparent on the word *victim*, the appropriate word being *accused*. Here the situational context plainly supports the use of the words *accused* and *victim*. Detection of the anomaly might occur because of a role violation with respect to the action of the judge, and this is based on situation-specific knowledge, rather than simple verb semantics—that *sentences* are not given to *victims*. Given the good fit to global context and role violation (situational role, in this case), one may expect borderline anomalies not to produce an N400 effect, but perhaps produce a P600-like effect.

Our second major goal is to investigate whether there is any evidence for detection of borderline anomalies in the brain, even when there is no awareness and no report. ERPs seem well suited to address this issue because a growing body of evidence suggests that they give access to implicit processes at various levels, from perceptual and semantic encoding to motor preparation processes (e.g., Fernandez-Duque, Grossi, Thornton, & Neville, 2003; Rolke, Heil, Streb, & Hennighausen, 2001; Leuthold & Kopp, 1998; Vogel, Luck, & Shapiro, 1998). Furthermore, there is some evidence that ERPs may be useful for the detection of brain activity in response to stimuli even in the absence of awareness and report of various events. Using an analysis of ERPs in the setting of visual change detection, Fernandez-Duque et al. (2003) showed that when changes occurred of which participants were unaware, a positive deflection between 240 and 300 msec occurred after the stimulus onset, relative to trials with no change (see also Kimura, Katayama, & Ohira, 2008). In the language domain, Vogel et al. (1998) employed the N400 in an attentional blink paradigm to reveal semantic processing of stimuli that participants fail to report. A context word preceded the start of a rapid serial visual presentation stream of distracters, into which a first digit target (T1) and a second word target (T2) were embedded, which both were to be identified and reported. Critically, even when participants failed to report T2, a larger N400 was triggered by the T2 word when it was unrelated rather than related to the context word. This result demonstrates the sensitivity of ERPs

to implicit semantic processes. Moreover, it is also worth mentioning that differences in language processing demands may be sometimes more sensitively indicated by N400 effects than by eye movement behavior. For example, using the same sentence materials in two separate ERP and eye tracking studies, differences in the processing of singular versus plural pronouns were revealed by N400 (Filik, Sanford, & Leuthold, 2008), but not by eye movement measures (Sanford, Filik, Emmott, & Morrow, 2008). Thus, should an N400 effect result with a borderline anomaly, it is, in principle, possible that it might occur even when the anomaly goes unreported. Of course, the same logic could apply to any other signature of language processing in the ERP. Such a finding would support the reduced awareness interpretation and run counter to the shallow processing account.

Whereas the primary aim of the present experiment was to investigate the ERP effects resulting from borderline anomalies, we also included materials that conform to the more classical, poor-fit anomalies. This was done primarily to guarantee that conventional semantic anomaly effects could be obtained in the same experimental run as was used for the good-fit materials.

METHODS

Participants

The 24 participants were right-handed students and research student volunteers at the University of Glasgow. The data of five participants were excluded from analysis because less than 10 artifact-free, missed anomaly trials were available for averaging.

Presentation and Apparatus

Experimental Run Time System software (BeriSoft Cooperation, 1987–2001, Frankfurt, Germany) running on a DOS computer controlled the presentation of stimuli on the computer monitor and the recording of behavioral responses. A fixation cross ($0.19^\circ \times 0.19^\circ$) was presented at the center of the monitor, in white on a black background. Spoken sentences were presented via a Sennheiser PX 100-headphone at an intensity comfortable for normal listening. Participants were seated in a dimly lit testing booth 80 cm away from the computer monitor. The space key of the computer keyboard was used to control sentence presentation, while a two-button response box was used to make responses. A microphone and loudspeaker in front of the participant permitted communication between the participant and the experimenter.

Materials

Two types of materials were constructed, one with a good fit to context, and one with the classical poor fit to context. We shall first consider the good-fit materials.

Good Fit to Context

A total of 135 materials with a good fit to context was devised, an example of which is:

- (1) Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year *{sentence/care order}* was given to the **victim**, but this was subsequently appealed.

The first sentence provides a setting for the story, that is, an introduction of the global context. The second sentence reinforces this global context, and provides two optional local context words, indicated in italics within brackets. It is these local context words that make the target word, shown in bold, either anomalous or nonanomalous. Thus, it is anomalous for a *sentence* to be given to a *victim*, but not for the victim to receive a *care order*.

To obtain a similar signal-to-noise ratio in average ERPs, it would be optimal if an equal number of trials would be available under detect and nondetect conditions, that is, if about 50% of the anomalous cases were detected. Earlier work (Bohan & Sanford, 2008) indicated that this level of detection would likely occur. With this in mind, the experiment was arranged such that, for any given participant, 90 of the good-fit materials would be presented in the anomalous condition, and 45 in the nonanomalous condition. This imbalance was used because if 50% of the anomalies were detected on average, this would place one third of participant responses into each of the three categories for analysis (anomaly detected, anomaly nondetect, non-anomalous). In order to ensure that each material occurred in both anomalous and nonanomalous conditions, it was necessary to rotate the materials over three presentation files. Within a given file, a particular material appeared in only one of the two conditions.

Classical, Poor Fit to Context Materials

Eighty poor-fit (classical anomaly) materials were constructed that could appear in either a nonanomalous condition or in an anomalous condition. The anomalies were created by using a word that was completely out of context, both locally and globally, corresponding to those known to produce N400 effects. An example is:

- (2) Leon was the manager of a struggling record shop. Yesterday, the owner told him that he would have to think of new ways to sell more *{letters/records}* if he wanted to keep his job. (*letters* = anomalous; *records* = nonanomalous control)

In any presentation file, half of these poor fit to context materials were anomalous, and half nonanomalous. By rotation over a pair of files, counterbalancing was possible. Finally, in order to merge the good fit to context and poor fit to context materials into presentation format, a total of six presentation files (3×2) was thus required.

The division of materials into good fit and poor fit to situation was crucial to the study. In a pretest, the relevance of the target words to the situations was assessed by means of a 7-point Likert scale (1 = does not fit and 7 = perfect fit). The average value for the borderline anomalies was 5.16, and for the easy anomalies it was 2.17, with $t(196.39) = 13.82, p < .0001$. The degree of freedom for t results from dropping the assumption of equal variance, as indicated by Levene's test for equality. These results show that the two types of materials successfully captured the desired goodness-of-fit of anomalous words to situations depicted.

The target words for the two anomaly-type manipulations did not differ from one another in terms of word frequency using the CELEX lexical database (Baayen, Piepenbroek, & Gulikers, 1995). Mean log frequencies were 1.30 per million for the good global fit targets, 1.30 for the poor global fit anomalies, and 1.36 for the global nonanomalous controls for the poor global fit anomalies ($F < 1$). There was no difference in mean word length either, at 6.5 letters for the good global fit targets, 6.1 for the poor global fit anomalies, and 6.0 for the poor global fit controls ($F = 2.0$).

All materials were recorded digitally by a female member of the Royal Scottish Academy of Music and Drama. This person had a trained voice, giving clear but natural enunciation of the materials. Trigger points were established for the onset of the critical word using Praat, a free-ware program for analysis and reconstruction of acoustic speech signals (Boersma & Weenink, 2005).

Procedure

Participants were asked to maintain fixation at the fixation cross while listening to the stories for normal comprehension. When they were ready to begin the participants pressed a "PROCEED" key on the response box and the first sentence was presented via headphones. When they had listened to the first sentence, they pressed the same key to proceed to the presentation of the second sentence. This began with a fixation cross presented for 1000 msec, which was for the first 500 msec displayed in red and then turned white. Participants were asked to press the response key labeled "NONSENSE" if they detected an anomaly. If they thought there was no anomaly, they pressed the "OK" key at the end of the sentence. There were no instructions to respond rapidly. The assignment of judgments to response keys was balanced across participants. After the auditory sentence presentation, the screen went blank for 1000 msec before a screen prompt asked the participants to verbally report their response to the experimenter. They did this for all trials. If there was an anomaly detected, they were also required to identify and explain it. The experimenter recorded what the participant said and whether or not they were correct. Participants completed an initial practice block of 8 stories, followed by 7 blocks of 35 stories and 1 final block of 14 stories. After the ERP recording, a multiple-choice ques-

tionnaire was delivered that checked participants understood all of the anomalies as being anomalous, which they did.

EEG Recording

A BioSemi Active-Two amplifier system was used for continuous recording of electroencephalographic (EEG) activity from 72 Ag/AgCl electrodes. EEG and EOG recordings were sampled at 256 Hz. The on-line reference electrode was the BioSemi Common Mode Sense (CMS) electrode (see www.biosemi.com/faq/cms&drl.htm for details). Off-line, all EEG channels were recalculated to an average mastoid reference and EEG activity was band-pass filtered (0.03–25 Hz, 6 dB/oct). Trials containing blinks were corrected using the adaptive artifact correction method of Brain Electromagnetic Source Analysis software (Ille, Berg, & Scherg, 2002). Automatic artifact detection software (Brain Electromagnetic Source Analysis) was run and trials with nonocular artifacts (drifts, channel blockings, EEG activity exceeding $\pm 75 \mu\text{V}$) were discarded, resulting in a loss of about 14.5% of the trials. The analysis epoch started 100 msec prior to the onset of the critical word, and lasted for a total duration of 1600 msec.

Data Analysis

For artifact-free trials, the signal at each electrode was averaged separately for each experimental condition time-locked to the onset of the critical word. The average number and range of trials included in each experimental condition is presented in Table 1. As can be seen in this table, the number of trials included in the missed-anomaly and the nonanomalous word condition of the good-fit anomalies was very similar to that of the poor-fit anomalous and nonanomalous, whereas more trials were included in the detected-anomaly word condition.

Average ERP waveforms were aligned to a 100-msec pre-onset baseline. Mean ERP amplitudes were measured in time intervals for N400 (300–600 msec) and for the late positivity (800–1100 msec) (e.g., Nieuwland & Van Berkum, 2005). Because some studies reported semantic

Table 1. Average Number of Trials and Range of Trials Included in the ERP Analysis as a Function of Experimental Condition

	<i>Condition</i>	<i>Average</i>	<i>Range</i>
Good global fit	Nonanomalous	33.3	20–41
	Anomaly detected	48.5	26–70
	Anomaly missed	27.5	13–47
Poor global fit	Nonanomalous	29.6	19–39
	Anomalous	29.8	15–38

anomaly effects on the auditory ERP waveform earlier than the N400 time interval (e.g., Hagoort & Brown, 2000), we also checked for possible anomaly effects in mean ERP amplitudes in the 200–300 and 300–400 msec time interval after onset of the critical word. ERP amplitudes at midline electrodes (Fz, FCz, Cz, CPz, Pz) were analyzed separately from data recorded over lateral electrodes, which were pooled to form regions of interest (ROIs) along a left–right dimension, an anterior-to-posterior dimension, and a dorsal–ventral dimension (cf. Filik et al., 2008). The six ROIs over the left hemisphere were: left anterior ventral (AF7, F7, FT7, F5, FC5), left anterior dorsal (AF3, F3, FC3, F1, FC1), left central ventral (TP7, T7, C5, CP5), left central dorsal (C3, CP3, C1, CP1), left posterior ventral (PO9', O9', P7, PO7, O1), and left posterior dorsal (P3, PO3, P1, P5); six homologue ROIs were defined for the right hemisphere.

Statistical analyses were performed by means of Huynh–Feldt corrected repeated measures analyses of variance (ANOVAs). Midline ERP amplitudes were analyzed by an ANOVA with variables condition (for poor fit: nonanomalous vs. anomalous; for good fit: nonanomalous vs. anomaly detected vs. anomaly missed), and electrode (Fz, FCz, Cz, CPz, Pz). ERP amplitudes over lateral ROIs were performed by an ANOVA with variables condition, hemisphere (left, right), ant–pos (anterior, central, posterior), and verticality (ventral, dorsal).

RESULTS

Detection Performance

For the poor-fit anomalies, detection rate was 95.2%. On average, participants correctly detected the borderline semantic anomalies at a rate of 63.9%, which provides an adequate split for the comparison of data from detected and missed anomaly trials.

Event-related Brain Potentials

Separate analyses were carried out for good global fit (borderline) anomalies and for poor fit (clear, classical).

Borderline (Good Global Fit) Anomalies

Figure 1 displays the grand-average ERP waveforms elicited by nonanomalous words, by well-fitting anomalous words that were either reported as being anomalous, or not reported as anomalous, as well as the topographic distribution of the detection effect. It is evident that anomalous words, either missed or detected, did not elicit a standard N400 effect in the 300–600 msec interval (all $F_s < 1$). Moreover, in none of the analyses of the earlier 200–300 and 300–400 msec time intervals were significant anomaly effects in mean ERP amplitudes obtained (all $F_s < 1.15$, $p_s > .34$).

Instead, an LPP was elicited after about 600 msec. Statistical analysis of midline LPP amplitude (800–1100 msec)

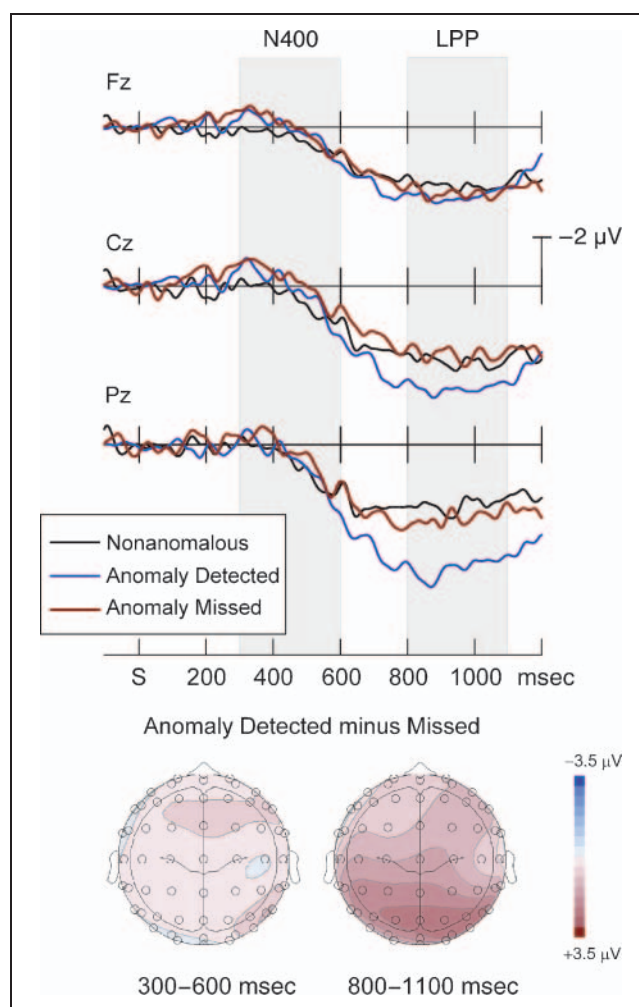


Figure 1. Grand-average ERP waveforms elicited at electrodes Fz, Cz, and Pz by well-fitting detected and missed anomalous words as well as nonanomalous words. The shaded areas indicate the time intervals for the analysis of N400 amplitude and LPP amplitude, respectively. Note that negativity is plotted upward. Bottom: Spline-interpolated topographic maps of the detection effect (anomaly detected minus missed). Isopotential line spacing is 0.5 μ V.

revealed a significant Condition \times Electrode interaction [$F(4, 72) = 4.0$, $p < .05$, $\eta_p^2 = .18$]. Planned comparisons indicated a larger LPP for detected anomalous words as compared to nondetected anomalous words [$F(4, 72) = 3.7$, $p < .05$, $\eta_p^2 = .17$], and compared to standard words [$F(4, 72) = 8.8$, $p < .001$, $\eta_p^2 = .33$]. LPP amplitude did not reliably differ for nondetected anomalous and standard words (all $F_s < 1.24$, $p_s > .30$). The significant Condition \times Hemisphere \times Ant–Pos interaction in the ROI analysis [$F(4, 72) = 5.5$, $p < .001$, $\eta_p^2 = .23$], indicated a larger LPP for detected than nondetected anomalous words over right posterior sites (cf. Figure 1).

Poor-fit Anomalies

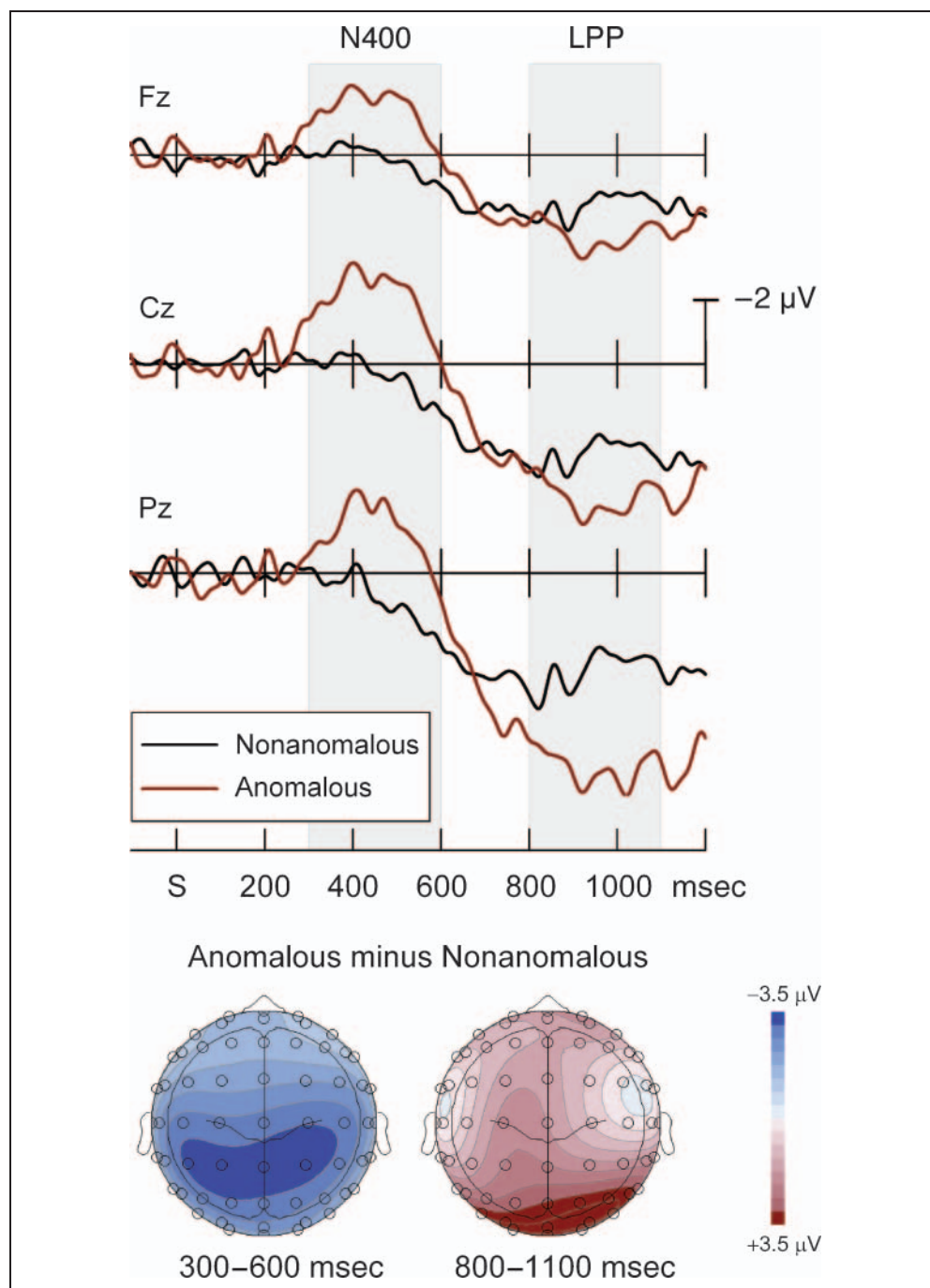
Figure 2 displays the grand-average ERP waveforms triggered by correctly identified anomalous versus nonanomalous

words, and the topographic distribution of the anomaly effect. It is clear that poor-fit anomalous words elicited a larger negativity than nonanomalous words in the 300–600 msec time interval [-1.9 vs. 0.7 μV ; $F(1, 18) = 8.0$, $p < .05$, $\eta_p^2 = .31$]. The influence of anomaly was strongest over posterior electrodes as indicated by the Congruency \times Electrode interaction [$F(4, 72) = 3.2$, $p < .05$, $\eta_p^2 = .15$], consistent with the N400 normally observed for this class of anomalies. The lateral ROI analysis confirmed the more negative ERP for anomalous than nonanomalous words [-2.2 vs. -0.2 μV ; $F(1, 18) = 6.2$, $p < .05$, $\eta_p^2 = .26$]. As can be seen in the topographic map,

this anomaly effect decreased from posterior to anterior electrodes [-2.3 and -1.8 μV vs. -1.2 μV ; $F(2, 36) = 4.2$, $p < .05$, $\eta_p^2 = .19$], and from dorsal to ventral electrodes [-2.4 vs. -1.2 μV ; $F(1, 18) = 6.2$, $p < .05$, $\eta_p^2 = .26$], again consistent with an N400 effect.

Figure 2 also indicates an emerging LPP that was larger to anomalous than nonanomalous words after about 800 msec. Statistical analysis of midline LPP amplitude (800–1100 msec) revealed a significant Congruency \times Electrode interaction [$F(4, 72) = 4.7$, $p < .01$, $\eta_p^2 = .21$]. This reflected a stronger anomaly effect over posterior electrodes, which was also evident in the ROI analysis [$F(2, 36) = 10.0$,

Figure 2. Top: Grand-average ERP waveforms at electrodes Fz, Cz, and Pz elicited by anomalous and nonanomalous words that are a poor fit to the context. The shaded areas indicate the time intervals for the analysis of N400 amplitude and LPP amplitude, respectively. Note that negativity is plotted upward. Bottom: Spline-interpolated topographic maps of the anomaly effect. Isopotential line spacing is 0.5 μV .



$p < .001$, $\eta_p^2 = .36$; cf. Figure 2]. In sum, the standard, ill-fitting anomalies showed first an N400 effect, and then an emerging LPP effect.

DISCUSSION

Although easy-to-detect anomalies typically have a poor fit to the global context of a passage, hard-to-detect anomalies typically have a good fit to the situation, but are locally anomalous. Behavioral work (Barton & Sanford, 1993) provided some direct evidence for this, using an analysis of a single case. From a processing viewpoint, hard-to-detect anomalies have been taken as evidence for shallow processing, in which the full meaning of a word is either not retrieved or is not integrated into the emerging discourse representation. Sanford and Garrod (1998) proposed that when a word has a good fit to the situation in which it is being used, then shallow processing will often result. The converse is that if a fit is poor, then fuller processing is the norm, which seems to be a good way of allocating processing resources. In the present study, we used ERPs as a potential means of uncovering the different types of processing afforded good fit and poor fit to global context words, and to check the validity of the shallow processing idea.

The results clearly showed that easy-to-detect anomalous words, with a poor fit to the sentential context, elicit a large, classic N400 effect. In contrast, the well-fitting borderline anomalies led to a situation where no N400 effect was observed at all. This lack of an N400 effect for borderline anomalies is consistent with our suggestion that the N400 effect indexes goodness-of-fit of a word to the global situation depicted in the stories. It is known that the N400 can be attenuated by lexico-semantic factors, such as a high degree of association between a word and other context words (see Kutas & Federmeier, 2000, for a review of this topic). This may be considered as an alternative explanation for the present data because the good-fit materials are essentially highly associated with the context situations.¹ Distinguishing these possibilities requires further research.

Additionally, as compared to nonanomalous sentences, a larger LPP between 800 and 1100 msec after the onset of the critical word was present in the cases of both well-fitting and poorly fitting anomalies. Therefore, this LPP would not appear to be a discriminating feature between the different types of anomalies studied here, in all likelihood, it reflects the use of an explicit judgment task (Kolk et al., 2003; but see Bornkessel-Schlesewsky & Schlewsky, 2008). The LPP that we observed cannot be attributed to animacy violations, for which P600 effects have been reported previously (cf. Kuperberg, 2007), but rather seems to be associated with a more general violation of the role a thing denoted by a word might play in a given situation. As suggested by Kuperberg (2007) and others, the LPP may simply reflect an attempt to reinterpret the ill-formed message resulting from an anomaly. Furthermore, the research of Kolk's group (e.g., Van de Meerendonk et al.,

2009; Vissers, Chwilla, & Kolk, 2007; Van Herten, Chwilla, & Kolk, 2006) suggests that the P600 effect is the result of explicitly monitoring the conflict between a well-formed syntactic interpretation of a sentence, and semantic analysis indicating an erroneous role mapping. This would be consistent with our finding of an LPP in the case of both good-fit and poor-fit anomalies. Finally, the LPP is known to be more prominent when an overt judgment of meaning violation is made, as in the present case (Kolk et al., 2003). However, even in the absence of an overt judgment task, LPPs still occur, as in the case of Nieuwland and Van Berkum (2005). It is, of course, possible that in the Nieuwland and Van Berkum case, people usually detected the anomalies, and because of that, effectively performed an implicit judgment operation. However, because there is no way of knowing how often detections took place in their experiment, there is no way to judge the situation. If, on some occasions, the anomalies were missed, and on others they were detected, then the magnitude of their LPP might be an average of these two states, resulting in a lower positivity than would be revealed should all anomalies be detected.

The second issue we examined concerned the nature of detected and undetected anomalies. We suggested that in the case of undetected anomalies, failure was caused by not recruiting the appropriate lexical semantic content during comprehension—the shallow processing account. This can be contrasted with a reduced awareness account, if the anomaly is detected at a system level, but the listener is unaware of this detection. We fractionated the anomaly trials into detect or nondetect, and compared them. There was no evidence of any differences in ERPs for undetected anomalies and nonanomalous controls. Rather, the key feature discriminating reported from missed borderline anomalies was the enlarged LPP in the former case. Thus, the absence of any measurable ERP effects for well-fitting anomalies that go unnoticed cannot be easily attributed to a general lack of sensitivity of ERPs to reveal borderline anomaly detection. The present ERP results therefore offer no evidence for the idea that anomalies might be detected within the comprehension system, but not reach awareness.

These results exactly parallel Bohan and Sanford's (2008) finding using an eye tracking methodology and materials similar to those in the present experiment. These authors found that there was an effect of anomaly on eye movements only on trials where the anomalies were detected and reported. When they were detected, the evidence clearly showed early effects on eye movement behavior, particularly with regressions. The failure to find any effect of anomaly in the absence of awareness lends further support to the shallow processing account rather than a reduction in access to awareness. Of course, because this result is effectively confirmation of a null hypothesis, it might be claimed that some other measure might possibly establish differences between missed anomalies and control nonanomalies. This is clearly a valid observation, but the present ERP results, together with the

eye movement behavior results of Bohan and Sanford (2008), address two of the most obvious approaches. Beyond this, an approach might be to analyze oscillatory brain responses (e.g., Hagoort et al., 2004) induced by missed anomalies and controls, for instance.

In sum, the claim that hard-to-detect anomalies of the “survivors” type and easy-to-detect anomalies are processed differently is supported, lending support to Sanford and Garrod’s (1998) claim, based on scenario mapping theory, that a good fit to global context serves to reduce the extent of local semantic processing. Such a process is arguably advantageous, because it is poor-fit words that, in general, will require full processing. Additionally, the data support the shallow processing hypothesis over an alternative resulting in mere reduced awareness.

Acknowledgments

We thank Jo Molle for assistance with the study. The project was funded by a grant from the Economic and Social Research Council (ESRC ES/G010757/1), UK, to A. J. S. and H. L., and, in part, by AHRC grant B/RG/AN8799/APN19525.

Reprint requests should be sent to Anthony J. Sanford, Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK, or via e-mail: a.sanford@psy.gla.ac.uk.

Note

1. We thank an anonymous reviewer for pointing out this possibility.

REFERENCES

- Baayen, R. H., Piepenbroek, R., & Gulikers, L. (1995). *The CELEX database*. [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Barton, S. B., & Sanford, A. J. (1993). A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477–487.
- Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer (Version 4.3.14) [Computer program]*. Retrieved from www.praat.org/ on 13 May 2005.
- Bohan, J., & Sanford, A. J. (2008). Anomaly detection at the borderline of consciousness: An eyetracking study. *Quarterly Journal of Experimental Psychology*, *61*, 232–239.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*, 55–73.
- Brédart, S., & Modolo, K. (1988). Moses strikes again: Focalisation effect on a semantic illusion. *Acta Psychologica*, *67*, 135–144.
- Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology*, *46*, 91–121.
- Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes*, *22*, 85–105.
- Daneman, M., Reingold, E. M., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 884–898.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.
- Erickson, T. D., & Matteson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Behaviour*, *20*, 540–551.
- Fernandez-Duque, D., Grossi, G., Thornton, I. M., & Neville, H. J. (2003). Representation of change: Separate electrophysiological markers of attention, awareness, and implicit processing. *Journal of Cognitive Neuroscience*, *15*, 491–507.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*, 554–558.
- Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1315–1326.
- Garrod, S. C., & Sanford, A. J. (1983). Topic dependent effects in language processing. In G. B. Flores d’Arcais & R. Jarvella (Eds.), *The process of language comprehension* (pp. 271–296). Chichester: Wiley.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*, 1518–1530.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). New York: 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.
- Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory & Cognition*, *29*, 449–461.
- Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, *37*, 187–204.
- 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*, 59–73.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended object and natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113–136.
- Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*, 113–124.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.
- Kimura, M., Katayama, J., & Ohira, H. (2008). Event-related brain potential evidence for implicit change detection: A replication of Fernandez-Duque et al. (2003). *Neuroscience Letters*, *448*, 236–239.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36.

- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.
- 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.
- Kutas, M., van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). Amsterdam: Academic Press.
- Leuthold, H., & Kopp, B. (1998). Mechanisms of priming by masked stimuli: Inferences from event-related potentials. *Psychological Science*, *9*, 263–269.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
- 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*, 691–701.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass*, *2*, 603–630.
- Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.
- Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses Illusion: Imperfect encoding, retrieval or match? *Journal of Memory and Language*, *30*, 385–406.
- Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology*, *38*, 165–174.
- Sanford, A. J., Filik, R., Emmott, C., & Morrow, I. (2008). They're digging up the road again: The processing cost of Institutional They. *Quarterly Journal of Experimental Psychology*, *61*, 372–380.
- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, *26*, 159–190.
- Sanford, A. J., & Garrod, S. C. (2005). Memory-based processing and beyond. *Discourse Processes*, *39*, 205–224.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 382–386.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*, 261–267.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. Th. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*, 1211–1224.
- 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*, 1181–1197.
- Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*, 35–46.
- Vissers, C. Th. W. M., Chwilla, D. J., & Kolk, H. H. J. (2007). The interplay of heuristic parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, *75*, 8–18.
- Vogel, E. K., Luck, S. J., & Shapiro, K. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1656–1674.