

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/54892>

Please be advised that this information was generated on 2021-09-20 and may be subject to change.

## OFFPRINT ORDERS FOR *Journal of Cognitive Neuroscience* 18:7

Prior to publication of your paper, offprints may be purchased at the prices listed below. All orders must be **prepaid**, and payment may be made **only** in the following forms: a check drawn on a U.S. bank, an international money order, VISA, MasterCard, or American Express (**no other credit cards can be accepted**). The MIT Press cannot guarantee that late orders will be placed; to ensure production of your offprints, this form and payment must be returned **within 14 days of receipt** to:

MIT Press Journals  
 Attn: Billing Department  
 238 Main St., Suite 500  
 Cambridge, MA 02142

*Please print or type, and fill out all items completely.*

- 1) Your name \_\_\_\_\_  
 Article name \_\_\_\_\_
  
- 2) How many copies do you want? (Number cannot be smaller than 100, and must be a multiple of 100.) \_\_\_\_\_
  
- 3) Total number of pages in article: \_\_\_\_\_. My article runs from pages \_\_\_\_\_ to \_\_\_\_\_.
  
- 4) Please refer to the price chart below to calculate the cost of your order. There are two sets of columns, one set for articles with color art and one for articles without color art; please make sure to use the correct columns.
  - *Articles with color:* for articles over 16 pages, please add the 16-page price to the price for the number of pages over 16 (i.e., for 20 pages, add the 16-page price and the 4-page price).
  - *Articles without color:* for articles over 32 pages, please add the 32-page price to the price for the number of pages over 32 (i.e., for 38 pages, add the 32-page price and the 6-page price).

pages in article	<i>WITH COLOR</i>		pages in article	<i>WITHOUT COLOR:</i>	
	first 100 offprints	each add'l 100 offprints		first 100 offprints	each add'l 100 offprints
1-4	\$ 49.00	\$ 16.00	1-4	\$ 37.00	\$ 13.00
5-8	\$ 91.00	\$ 30.00	5-8	\$ 57.00	\$ 22.00
9-16	\$ 169.50	\$57.45	9-12	\$ 78.00	\$ 34.00
			13-16	\$ 109.00	\$ 43.00
			17-20	\$ 129.00	\$ 55.00
			21-24	\$ 149.00	\$ 65.00
			25-28	\$ 173.00	\$ 76.00
			29-32	\$ 203.00	\$ 85.00

Total cost of offprints: \$ \_\_\_\_\_

Canadian orders please add 7% GST: \$ \_\_\_\_\_

TOTAL: \$ \_\_\_\_\_

PAYMENT (please print or type):

Amount of check enclosed: \_\_\_\_\_ made payable to The MIT Press.

- VISA # \_\_\_\_\_
- MasterCard # \_\_\_\_\_
- American Express # \_\_\_\_\_

Expiration date: \_\_\_\_\_ Signature: \_\_\_\_\_

**SHIP TO:**

Name \_\_\_\_\_  
 Address \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

# When Heuristics Clash with Parsing Routines: ERP Evidence for Conflict Monitoring in Sentence Perception

Marieke van Herten, Dorothee J. Chwilla, and Herman H. J. Kolk

## Abstract

■ Monitoring refers to a process of quality control designed to optimize behavioral outcome. Monitoring for action errors manifests itself in an error-related negativity in event-related potential (ERP) studies and in an increase in activity of the anterior cingulate in functional magnetic resonance imaging studies. Here we report evidence for a monitoring process in perception, in particular, language perception, manifesting itself in a late positivity in the ERP. This late positivity, the P600, appears to be triggered by a conflict between two interpretations, one delivered by the standard syntactic algorithm and one by a plausibility heuristic which combines individual word meanings in the most plausible way. To resolve this conflict, we propose that the brain reanalyzes the memory trace of the perceptual input to check for the possibility of a processing error. Thus, as in Experiment 1, when the reader is presented

with semantically anomalous sentences such as, “The fox that shot the poacher. . .,” full syntactic analysis indicates a semantic anomaly, whereas the word-based heuristic leads to a plausible interpretation, that of a poacher shooting a fox. That readers actually pursue such a word-based analysis is indicated by the fact that the usual ERP index of semantic anomaly, the so-called N400 effect, was absent in this case. A P600 effect appeared instead. In Experiment 2, we found that even when the word-based heuristic indicated that only part of the sentence was plausible (e.g., “...that the elephants pruned the trees”), a P600 effect was observed and the N400 effect of semantic anomaly was absent. It thus seems that the plausibility of part of the sentence (e.g., that of pruning trees) was sufficient to create a conflict with the implausible meaning of the sentence as a whole, giving rise to a monitoring response. ■

## INTRODUCTION

Monitoring refers to a process of cognitive control aimed at output optimization. The existence of such a process has been proposed for different domains. In the language domain, monitoring manifests itself in the phenomenon of “self-repair” in speech. In “overt” self-repairs, the speaker interrupts the utterance after an error has been made, retraces to the beginning of the word or phrase, and then produces the correct form (e.g., “I thought she. . .I thought he was looking at me”). Levelt (1983) argues that, in addition to overt repairs, there are also “covert” repairs, in which—due to a process of prearticulatory editing—an error is intercepted at the level of planning. Covert repairs manifest themselves by editing terms, word repetitions, or pauses (e.g., “I thought—er—I thought he was looking at me”). An important argument for the existence of prearticulatory editing is that overt repairs sometimes occur after just one phoneme has been produced. Such rapid interruptions presumably do not leave enough time for a process of overt error recognition. According to Levelt’s theory of error monitoring in speech,

speakers detect their errors in the same way as they detect errors in the speech of others: via the comprehension system (Levelt, 1983). Recently, Hartsuiker and Kolk (2001) have provided computational evidence for this theory.

There is by now extensive event-related potential (ERP) evidence for a monitoring process in the action domain. Errors in choice reaction time (RT) tasks elicit an “error-related negativity” (ERN), typically occurring around 100 msec after the error has been made (see Yeung, Botvinick, & Cohen, 2004, for a recent review). ERN activity is also observed if participants are told that an error has been made, whether this was true or not. This implies that an overt motor response is not required for the ERN to occur. Both functional magnetic resonance and ERP studies have provided support for the hypothesis that the anterior cingulate cortex (ACC) is involved in this error monitoring process.

Different answers have been given to the question as to what constitutes the trigger of the monitoring process. A first possibility is that it is triggered by a mismatch between the observed and the intended response. One disadvantage of this option is that it cannot account for the fact that error monitoring, as reflected in ACC activity, is not only present in erroneous trials but even

if no error is made. In particular, ACC activity has been observed when multiple responses compete for the control of action. Thus, a second possible trigger of the monitoring process—in addition to the error as such—is the presence of a “conflict.” Conflict in this context refers to the concurrent activation of incompatible responses. Situations in which there is a conflict will generally be situations in which many errors are made, but the theory says that it is the conflict that elicits a monitoring process, not the observation of the error as such. Nevertheless, participants are generally able to detect errors, and this ability must also be accounted for. Yeung et al. (2004) provide a conflict model of error monitoring by the ACC, and this model also includes a mechanism for error detection.

Errors in production, it seems, are extensively monitored for, both in the language and in the action domains. However, we also make occasional errors of perception (e.g., misreading a word) or comprehension (e.g., misunderstanding a speaker) and there is no a priori reason why such errors would not be monitored for as well. Observing an error of perception may, for instance, lead one to change one’s perceptual strategy and thereby improve perception. Nevertheless, monitoring perceptual errors has received very little attention, not only in the action domain but also in the language domain. However, how could the brain monitor for errors of perception? After all, there are no errors to observe, opposed to real errors in production. However, the conflict hypothesis developed for action monitoring reveals a possible mechanism for monitoring in perception. As explained above, in the action domain, the simultaneous activation of two incompatible responses is assumed to trigger a monitoring response. Similarly, if language perception leads to the activation of two incompatible interpretations, a conflict would arise, signaling the possibility of a processing error. Such a conflict could trigger a monitoring response to check for the possibility of such an error. One example would be the so-called garden path sentence. In garden path sentences (e.g., “The woman persuaded to answer the door. . .”), initially, one interpretation is chosen, but has to be replaced by a different interpretation later on. In the case of the example sentence, readers initially assume that the sentence is about a woman persuading someone, but after reading the sentence part following the verb, they realize that the sentence is about a woman being persuaded. Such sentences generally elicit a late positive ERP effect, occurring roughly between 500 and 800 msec after stimulus onset, the P600 effect (e.g., Osterhout & Holcomb, 1992).

Garden path sentences are one example of a situation in which different analyses of the same linguistic string elicit a conflict. Another type of conflict in the language domain is related to the existence of heuristics or “perceptual strategies.” Although sentence processing is assumed by many to involve algorithmic analysis of

syntactic structure, such heuristics have been proposed to play an important role from the beginning of the psycholinguistic enterprise (e.g., Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Bever, 1970). Heuristics can be regarded as “rules of thumb”: highly economical strategies that are usually but not invariably effective in extracting meaning. Although the number of possible heuristics is large, two specific strategies have been described in more detail. The first “NVN strategy” involves treating the first noun, the verb, and the second noun as referring to agent, action, and patient roles, respectively. The second “plausibility heuristic” entails that readers combine the lexical items of a sentence in the most plausible way. Here sentences are treated as unordered lists of words. A string like *cat–milk–drink* can only have one plausible meaning and, to derive this meaning, a syntactic parse is not necessary. Ferreira (2003) has provided evidence for the use of both strategies in normal speakers. Furthermore, the use of heuristics plays an important role in the explanation of agrammatic comprehension (e.g., Kolk & Weijts, 1996; Caramazza & Zurif, 1976).

If we assume that normal sentence processing entails both the use of parsing algorithms and the use of heuristics, the question arises how the two are related. A first possibility is a cascade-like model in which heuristics constrain the initial search space of the subsequent more time-consuming algorithmic analysis, so that “semantics proposes and syntax disposes” (Townsend & Bever, 2001). A second possibility is that heuristic and algorithmic processing take place in parallel. Thus, in analogy with the well-known dual-route model of reading aloud (e.g., Coltheart, Curtis, Atkins, & Haller, 1993), sentence processing may proceed through two routes, which together determine the final interpretation of the sentence (Kolk, Chwilla, van Herten, & Oor, 2003).

The hypothesis that sentence processing proceeds through two routes implies the possibility of a conflict. What would happen if the heuristic and the algorithmic routes lead to different interpretations? This is particularly clear in the case of the plausibility heuristic. If the plausibility heuristic produces the most plausible interpretation of the set of content words that occur in the sentence (e.g., the words *deer–hunter–chase* lead to the interpretation that the hunter is chasing the deer), then highly implausible but grammatical and unambiguous sentences (e.g., “The deer was chasing the hunter”) will produce a conflict. It is this type of sentences that has been subject to a number of recent ERP studies. Despite differences in sentence materials and language (English and Dutch), the general finding was that implausible sentences relative to their plausible counterparts (e.g., “The hunter was chasing the deer”) did *not* elicit an N400 effect—as would have been expected, given that semantic anomalies typically elicit an N400 effect—but instead a P600 effect (Kuperberg, Caplan,

Sitnikova, Eddy, & Holcomb, in press; Kim & Osterhout, 2005; van Herten, Kolk, & Chwilla 2005; Hoeks, Stowe, & Doedens, 2004; Kolk et al., 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). This result was highly unexpected as P600 effects have been shown to consistently occur to syntactic anomalies, but not to semantic anomalies. How can one account for this paradoxical finding and how does it relate to the presence of a conflict between algorithmic and heuristic processing routes?

The accounts that have been proposed for these phenomena have two important assumptions in common. The first is that individual word meanings “cue,” “suggest,” or “prime” a plausible role assignment for both plausible and implausible sentences, even in syntactically unambiguous sentences. For instance, Kim and Osterhout (2005) speak of “semantic attraction” and suggest that this reflects “the activation of highly stable representations in world knowledge” (p. 216). As a result, no N400 effect is obtained. It seems then that all accounts—implicitly or explicitly—adhere to the notion of a plausibility heuristic. The second assumption is that the P600 effect reflects an immediate consequence of the situation that the parse and the individual word meanings suggest different interpretations: an implausible one in the first and a plausible one based on world knowledge in the second case. The accounts diverge, however, in their description of this immediate consequence.

A first possibility has been suggested by Kuperberg, Sitnikova, et al. (2003). Because semantic relationships between the individual words suggest one set of role assignments and the regular parse suggests another, a “mismatch” occurs. In response to this mismatch, the processing system is said to “repair the anomaly by reassigning thematic roles” (p. 128). This repair process is of a syntactic nature: It involves a process of restructuring, similar to what happens in garden path situations, which also elicit P600 effects (e.g., Osterhout & Holcomb, 1992). This interpretation has the advantage that it connects to the dominant view that the P600 has a syntactic function (e.g., Hagoort, Brown, & Groothusen, 1993). However, the sentences we are dealing with (see references above) are not ambiguous, like garden path sentences. In the context of ambiguous sentences, the notion of repair makes sense as it refers to the replacement of one sentence parse by another. In syntactically unambiguous sentences, there is nothing to replace because the syntactic structure allows only one role assignment or interpretation. Becoming involved in such restructuring would lead the system away from a veridical sentence interpretation and this would bring participants to erroneously evaluate the sentences as plausible. There is, however, no evidence that participants actually do this because, as the authors themselves admit, they almost always classify the sentences correctly in a judgment task.

A second possible consequence of the mismatch between lexical and syntactic analysis was investigated by van Herten et al. (2005) (see Kim & Osterhout, 2005, for a related idea). They proposed that P600 effects to semantically anomalous sentences could arise if the interpretation provided by the lexical analysis leads the participant to expect a particular grammatical morphology. The discrepancy between the expected and the observed morphology would then be responsible for the P600 effect. This hypothesis was tested by manipulating grammatical number. In a sentence such as *De kat die voor de muizen vluchtte...* (literal translation: “The cat that fled for the mice fled...”; paraphrase: “The cat that fled for the mice...”), the plausible interpretation is that the mice are fleeing, and this would lead one to expect a plural inflection of the verb. Because the Dutch verb *vluchtte* carries the singular inflection, the syntactic prediction is violated. However, in sentences in which subject and object noun phrase (NP) carry the same number, such violations should not be noticeable. Therefore, if an unexpected grammatical morphology gave rise to the P600 effect, then a P600 effect should be present in the conditions in which subject and object carry a different number, but not in the conditions in which they carry the same number. However, a P600 effect was present not only in the different number sentences but also in the same number sentences. This showed that the P600 effect to reversal anomalies was not due to a syntactic mismatch but was a response to the semantic anomaly (the meaning of the unexpected verb) as such.

It appears that neither of the two accounts described above can fully explain the P600 effects to semantic anomalies. There is, therefore, reason to consider a third approach, already alluded to above, that is, that we are dealing with conflict monitoring. Faced with a conflict between the outcomes of the heuristic and the parser, the language system attempts to resolve this conflict simply by reprocessing the sentence. That is, the conflict triggers a process that checks upon the veridicality or truthfulness of the reader’s analysis. After all, an inconsistency can have two sources. It can be real, in the sense that an unexpected event did indeed occur. On the other hand, it can also stem from a processing error. To be sure that no erroneous information is integrated into the current discourse, the reader will generally check upon the correctness of his or her analysis in case of a conflict.

Assuming that sentence processing depends upon the joint action of algorithm and heuristic routes, three possible situations exist. We hereby assume, in line with Townsend and Bever (2001), that the algorithmic parser always comes up with the right answer (presuming that it is given a sufficient amount of time). Additionally, because the heuristic is a plausibility heuristic, it will always come up with a semantically plausible reading. A first situation is that the algorithm and the heuristic

both deliver a semantically plausible sentence interpretation. In this case, neither an N400 effect nor a P600 effect is expected. Second, the algorithm and the heuristic both deliver a semantically implausible sentence interpretation. Now, no P600 effect should occur as there is no conflict. However, an N400 effect is predicted because semantic integration is hampered. Third, the algorithm delivers a semantically implausible sentence interpretation, whereas the heuristic delivers a semantically plausible sentence interpretation. In such a situation, a P600 effect should appear because the dissimilar outcome of the algorithm and the heuristic yields a conflict. Furthermore, no N400 effect should be present because the heuristic routine delivers a plausible interpretation not only for the plausible but also for the implausible sentences. As the literature described above shows, the third situation reliably elicits a P600 effect, in most cases, in the absence of an N400 effect.

The purpose of the current study is to further investigate the second situation in which the algorithm and the heuristic both deliver a semantically implausible sentence interpretation. These sentences are expected to elicit an N400 effect without a P600 effect, and this is what has been found in most of the studies. Kim and Osterhout (2005), for example, presented participants with sentences such as “The sealed envelope was devouring. . .” Compared to appropriate controls, these sentences elicited no P600 effect, but an N400 effect. However, in the study of van Herten et al. (2005), an N400 effect was found which was followed by a P600 effect. van Herten and colleagues employed sentences such as *De boom die in het park speelde. . .* (paraphrase translation: “The tree that played in the park. . .”). The words “played,” “tree,” and “park” cannot be integrated to form one semantically plausible unit. We accordingly predicted an N400 effect and no P600 effect. In contrast to this prediction, a P600 effect occurred in addition to an N400 effect. This seems difficult to explain by a hypothesis that couples the P600 with the existence of a conflict between an algorithmic and a heuristic route.

Experiment 1 was set up to investigate whether the second situation (algorithm and heuristic both deliver a semantically implausible sentence interpretation) reliably elicits a P600 effect. Experiment 1 thus replicates the van Herten et al. (2005) study with two major modifications. First, the semantic violations were created not by changing the verb, as was the case in the previous study, but by changing the subject NP (for example sentences, see Table 1). This results in the critical verb being the same for acceptable and unacceptable sentences. Second, because we wanted to compare two kinds of semantic anomaly, it was important to employ the same kind of violation for the two sentence types. So far we had been using implausibilities in our critical sentences (e.g., “The fox that *bunted* the poacher”), we now shifted to selectional restriction

**Table 1.** Examples of the Sentence Material from Experiment 1

<i>Reversal Condition</i>	
Acceptable	De schilder die op de ladder <i>klom</i> viel plotseling.
	The painter that on the ladder <i>climbed</i> fell suddenly. <sup>a</sup>
	The painter who <i>climbed</i> the ladder suddenly fell. <sup>b</sup>
Unacceptable	De ladder die op de schilder <i>klom</i> viel plotseling.
	The ladder that on the painter <i>climbed</i> fell suddenly. <sup>a</sup>
	The ladder that <i>climbed</i> the painter suddenly fell. <sup>b</sup>
<i>Nonreversal Condition</i>	
Acceptable	De eekhoorn die in de boom <i>klom</i> zag er schattig uit.
	The squirrel that in the tree <i>climbed</i> looked cute. <sup>a</sup>
	The squirrel that <i>climbed</i> the tree looked cute. <sup>b</sup>
Unacceptable	De appel die in de boom <i>klom</i> zag er sappig uit.
	The apple that in the tree <i>climbed</i> looked juicy. <sup>a</sup>
	The apple that <i>climbed</i> the tree looked juicy. <sup>b</sup>

The critical words are italicized.

<sup>a</sup>Word-by-word translation.

<sup>b</sup>Paraphrase.

violations (e.g., “The fox that *shot* the poacher”). So the comparison will be between sentences of the latter type, which will be labeled *reversal anomalies*, and sentences such as “The tree that shot the poacher,” which will be labeled *nonreversal anomalies*. We expect reversal anomalies to create a conflict between heuristic and algorithmic routines, and therefore, to elicit a P600 effect without an N400 effect. For nonreversal anomalies, we predict that they will not induce a conflict and will therefore give rise to an N400 effect without a P600 effect.

## EXPERIMENT 1

### Methods

#### *Participants*

There were 26 participants (mean age = 22 years; 20 women). All were native speakers of Dutch, had no

reading disabilities, were right-handed, and had normal or corrected-to-normal vision.

### Materials

All sentences consisted of center-embedded subject-relative sentences. Sentence acceptability was experimentally manipulated: A semantically acceptable variant and a semantically unacceptable variant were created for each sentence. Semantically unacceptable sentences always contained a selectional restriction violation. For the reversal condition, the selectional restriction violations resulted from reversing the subject and the object NP of semantically acceptable sentences that express a plausible and familiar event. The example sentence in Table 1, for example, depicts the likely concept of a painter climbing a ladder. The unacceptable reversed sentence, on the other hand, depicts a very unlikely and even impossible event, that is, a ladder climbing a painter. For the nonreversal condition, the selectional restriction violations resulted from changing the first NP of a semantically correct sentence into an NP that violated the selectional restrictions of the verb. For example, the NP “squirrel” was changed into “apple” in the example sentence in Table 1 to create the semantically unacceptable sentence “The apple that climbed the tree. . . .” Note that in the current condition, reversing subject NP and object NP does not lead to a correct sentence as was the case in the reversal condition. In the reversal and nonreversal conditions, the subject and object NP always had the same number; this was singular in about half (reversal: 52%, nonreversal: 57%) of the sentences and plural in the other half. Furthermore, about equal numbers of animate and inanimate nouns were employed in the reversal and nonreversal conditions. Finally, in all sentences, the violation was not evident before the relative clause’s verb.

The reversal and nonreversal conditions were presented in separate blocks. The number of trials was the same in both blocks. For each block, the experimental sentences were divided equally into two lists. None of the items was repeated, so each participant only saw one variant of a sentence. Each list contained 60 experimental sentences, of which 30 were acceptable and 30 were semantically unacceptable sentences. An equal number of filler sentences was added to each list: 15 acceptable right-branching sentences, 15 semantically unacceptable right-branching sentences, 15 acceptable conjunctions, and 15 semantically unacceptable conjunctions. Semantically unacceptable filler sentences contained selectional restriction violations, which were of the reversal or the nonreversal type dependent on the block in which they were presented (for examples, see Table 2). Experimental reversal sentences were accompanied by filler reversal sentences, whereas experimental nonreversal sentences were accompanied by filler nonreversal sentences.

**Table 2.** Examples of the Filler Sentences

<i>Experiment 1</i>	
Reversal Condition	
Right-branching	# De rechter luisterde naar de beklagde die opkwam voor zijn advocaat. # The judge listened to the defendant who stood up for his lawyer. <sup>a</sup>
Conjunctions	# De zeehonden doken in het water en vingden de ijsbeer. # The seals plunged into the water and caught the ice bear. <sup>a</sup>
Nonreversal Condition	
Right-branching	# De tuinmannen baalden van de struiken die de tuinen verhoorden. # The gardeners were fed up with the shrubs that interrogated the gardens. <sup>a</sup>
Conjunctions	# De reizigers overnachtten in het hotel en verdampten de volgende ochtend. # The travelers stayed the night at the hotel and evaporated the next morning. <sup>a</sup>
<i>Experiment 2</i>	
Acceptable	Jan zag dat de schildpadden wandelden over het zand dat heet was van de zon. John saw that the turtles walked on the sand that was hot from the sun. <sup>a</sup>
Unacceptable	
Violation mid-sentence	# Jan zag dat de dief kwispelde naar de bewakers die kwamen aanrennen. # John saw that the thief wagged to the guards that came running. <sup>a</sup>
Violation end-sentence	# Jan zag dat de poes sloop naar de merels die zachtjes neuriëden. # John saw that the cat stalked the blackbirds that hummed softly. <sup>a</sup>

The symbol # is used to indicate semantically implausible sentences.

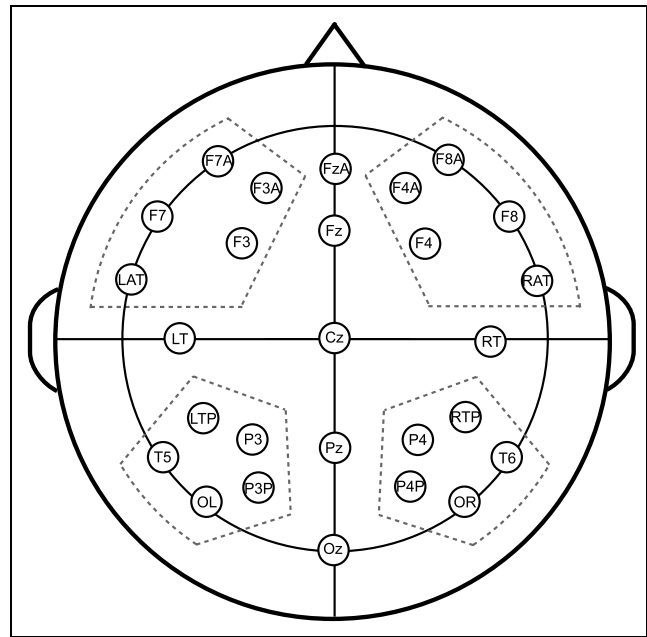
<sup>a</sup>Paraphrase.

## Procedure

Sentences were presented word-by-word in serial visual presentation mode at the center of a Macintosh monitor. Word duration was 345 msec and stimulus-onset asynchrony was 645 msec. Sentence final words were followed by a full stop. The intertrial interval was 2 sec. Words were presented in black capitals on a white background in a 9-by-2-cm window at a viewing distance of approximately 1 m. Each sentence was preceded by a fixation cross (duration 500 msec), followed by a 500-msec blank screen. A set of practice trials preceded the experimental trials. The two conditions were presented in separate blocks. As ERPs have been shown to be sensitive to list composition (e.g., Chwilla, Kolk, & Mulder, 2000), the more salient semantic violations that formed the nonreversal condition always followed the less salient reversal violations. There was a short pause between blocks. Participants were instructed to read the sentences and were told that they should do this attentively and thoroughly to be able to answer the content questions that followed the experiment. Thirty-six content questions were added at the end of the experiment which were responded to with “yes” or “no” by pressing a button on a button box. Because eye movements distort the electroencephalogram (EEG) recording, participants were trained to make eye movements, blinks in particular, only in the period between the end of the last sentence and the beginning of the next sentence.

## EEG Recording and Data Analysis

EEG was recorded with 27 tin electrodes mounted in an elastic electrode cap (Electrocap International). For electrode positions, see Figure 1. The left mastoid served as reference. An electrode was also placed on the right mastoid. Electrode impedance was less than 3 k $\Omega$ . The electrooculogram (EOG) was recorded bipolarly; vertical EOG was recorded by placing an electrode above and below the right eye and the horizontal EOG was recorded via a right to left canthal montage. The signals were amplified (time constant = 8 sec, bandpass = 0.02–30 Hz), and digitized on-line at 200 Hz. EEG and EOG records were examined for artifacts and for excessive EOG amplitude (>100  $\mu$ V). Averages were aligned to a 150-msec baseline preceding the critical verb. Before the analyses, the signals were referenced to the average of the right and left mastoid. ERPs were analyzed by calculating mean amplitudes in the 400–500 msec window and the 600–800 msec window, capturing N400 and P600 effects, respectively. The mean amplitudes were entered into a repeated-measures multivariate analysis of variance (MANOVA). The multivariate approach was used to avoid problems concerning sphericity. Midline and lateral sites were analyzed in separate MANOVAs so that laterality effects could be examined.



**Figure 1.** Electrode positions.

The midline analysis included five levels of site, whereas the lateral analysis included five levels of site, two levels of hemisphere (left, right), and two levels of region of interest (ROI; anterior, posterior; see Figure 1). If the analyses yielded interactions with the factor site, paired *t* tests were performed at the single-site level. Additionally, the analyses included two levels of acceptability (acceptable, unacceptable).

## Validation Study of the Materials

An RT study was conducted to check for the presence of unsuitable items and to test whether participants were successful in detecting the semantically unacceptable sentences. A separate group of 20 participants was tested that fulfilled the same criteria as those participating in the ERP experiment. The procedure was identical to the ERP experiment, except for one point. That is, the task for the participants consisted of a speeded acceptability judgment task. Participants were instructed to attentively read each sentence and indicate as fast as possible during reading of the sentence whether the sentence “had an odd meaning” or not by pressing one of two pushbuttons.

First, items that were miscategorized by at least half of the participants (unsuitable items) were omitted (three items for the reversal condition, three items for the nonreversal condition) before analyzing the RT and error data presented below. For the EEG experiment, these items were replaced by new items. MANOVAs with repeated measures on condition (reversal, nonreversal) and acceptability (acceptable, unacceptable) were performed for RT and error data. As Table 3 shows, unac-



**Table 3.** Response Times and Error Percentages (in brackets) from the RT Validation Study

Reversal Condition		Nonreversal Condition	
Acceptable	Unacceptable	Acceptable	Unacceptable
732	938	767	916
(6,67)	(3,67)	(8,17)	(3,67)

ceptable sentences were responded to slower [ $F(1,19) = 8.24, p < .05$ ] but elicited less errors [ $F(1,19) = 9.73, p < .01$ ] than acceptable sentences, indicating a speed-accuracy tradeoff. No differences between conditions were found (all  $F_s < 1$ ), and no Condition  $\times$  Acceptability interactions were present (all  $F_s < 1.5$ ). Most important for our present purpose is that the participants were successful in detecting the semantic violations in both the reversal and the nonreversal conditions as the error percentages were far above chance level (reversal: 5.17%, nonreversal: 5.92%).

## Results

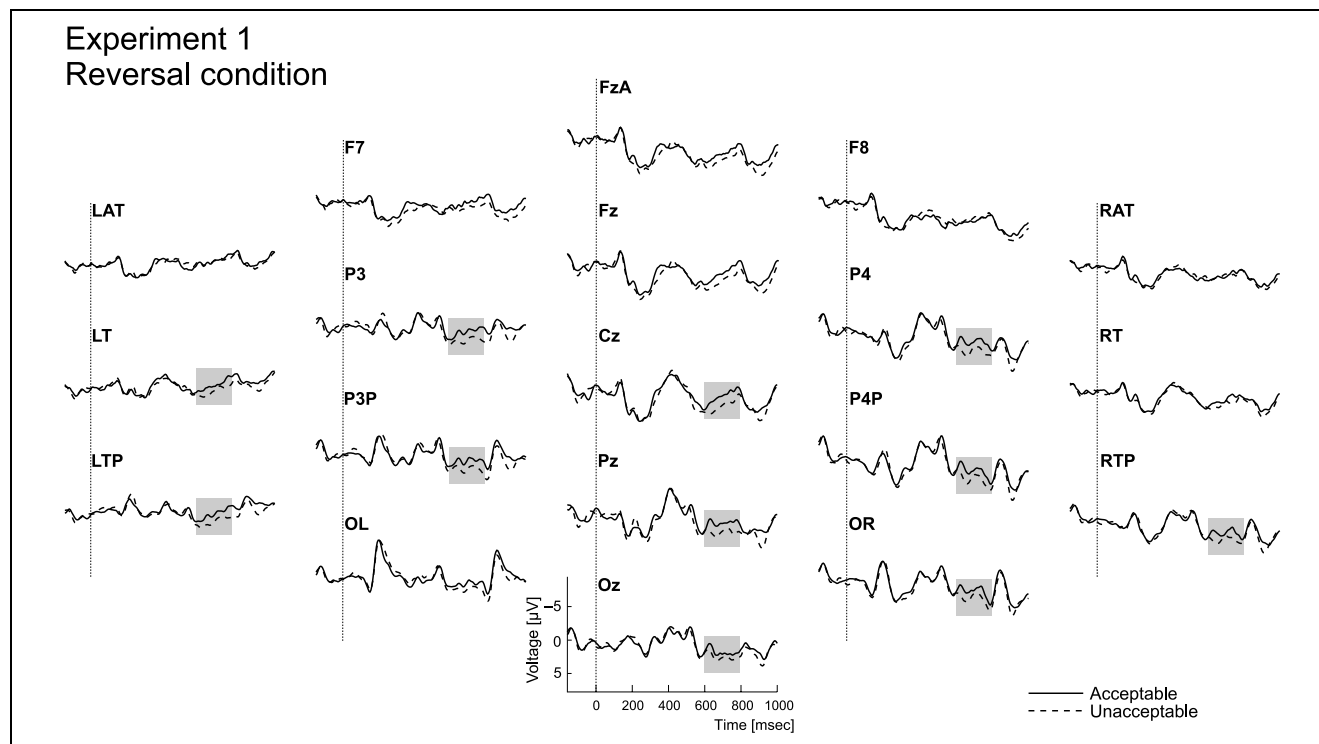
### Performance on Content Questions

Mean error rate to the content questions was 27.19% (reversal condition: 31.13%, nonreversal condition: 23.13%). The reason why participants made so many errors is likely due to the fact that the questions were asked at the end

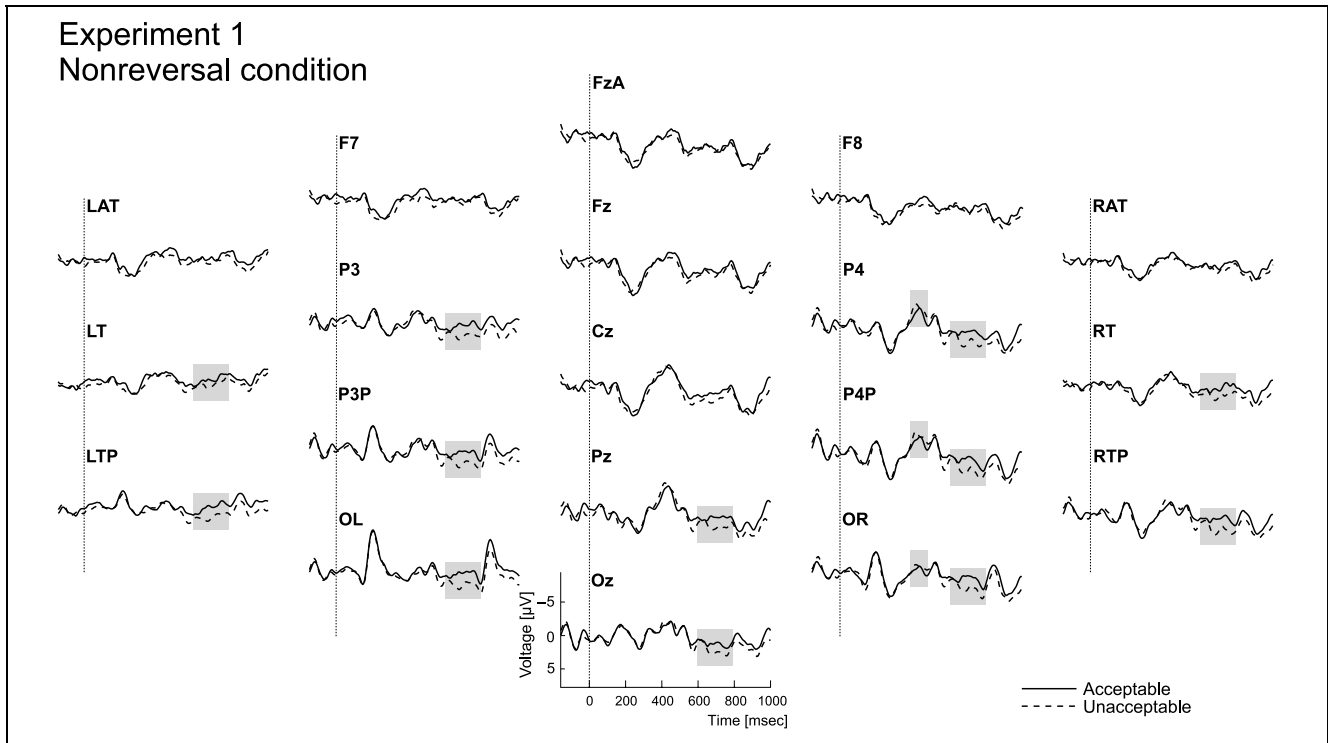
of the experiment and not because the sentences as such are hard to understand. After all, error percentages are much lower in the RT validation study. Nevertheless, error percentages are below chance level, and this indicates that participants attentively read the sentences during the EEG experiment. Response time and error analyses were not performed because the number of trials per condition (four to five) was very small.

### Event-related Potentials

The grand-average waveforms were time-locked to the critical verb for all midline sites, and a representative subset of lateral sites for the reversal condition and the nonreversal condition are displayed in Figures 2 and 3, respectively. All conditions elicited for visual stimuli a characteristic early ERP response—that is, an N1 followed by a P2, which at occipital sites was preceded by a P1 component. These early components were followed by a negative-going wave that peaked at about 425 msec (N400 component) largest at central and posterior sites, which was followed by a slow positive shift starting at about 600 msec and extending up to 1000 msec (P600 component). Inspection of the waveforms for the “reversal condition” suggests that no N400 effect (more negative-going amplitudes for unacceptable than acceptable verbs) was present but that a P600 effect (more positive-going amplitudes for unacceptable than acceptable verbs) was present. The P600 effect in terms of its



**Figure 2.** Grand ERP averages for all midline and a subset of lateral sites for the reversal condition of Experiment 1. Averages are time-locked to the onset of the critical verb and superimposed for the two levels of acceptability.



**Figure 3.** Grand ERP averages for all midline and a subset of lateral sites for the nonreversal condition of Experiment 1. Averages are time-locked to the onset of the critical verb and superimposed for the two levels of acceptability.

timing (maximal differences between 600 and 800 msec) and scalp distribution (the effect was largest over central and posterior sites) resembled the P600 effect observed in a variety of syntactic violations (e.g., Hagoort et al., 1993; Osterhout & Holcomb, 1992). The statistical analyses for the N400 window (400–500 msec) for the midline and the lateral sites revealed no effects of acceptability (midline and lateral:  $F_s < 1$ ) or Acceptability  $\times$  Site interactions (midline:  $F < 1$ ; lateral:  $F < 1.5$ ). No other interactions were observed (all  $F_s < 2$ ). The statistical analyses for the P600 window (600–800 msec) yielded a main effect of acceptability for the midline sites [ $F(1,25) = 5.13, p < .05$ ], indicating that a P600 effect was obtained. For the lateral sites, no acceptability effect ( $F < 2.9$ ), but an Acceptability  $\times$  Site interaction, was found [ $F(4,100) = 3.39, p < .05$ ]. Single-site analyses revealed P600 effects for the following lateral sites: LTP, RTP, P3, P4, T5, T6, P3P, P4P, OR. No other interactions were present (all  $F_s < 2.2$ ).

Visual inspection of the waveforms for the “nonreversal condition” suggests that no N400 effect was present at the midline and left posterior sites, whereas a small N400 effect appeared to be present for a subset of right posterior sites. The most distinguishing feature in the waveforms, however, seemed to be a P600 effect which followed the N400. Statistical analyses for the N400 window did not yield main effects of acceptability, neither for the midline nor for the lateral sites ( $F_s < 1$ ). For the lateral sites, an Acceptability  $\times$  Site interaction [ $F(4,100) = 3.49, p < .05$ ; midline:  $F < 1.2$ ] and

a trend toward an Acceptability  $\times$  Hemisphere interaction [ $F(1,25) = 3.86, p < .07$ ] were found. Follow-up single-site analyses showed that an N400 effect was present at three sites of the right hemisphere: P4, P4P, and OR (all  $p$  values  $< .05$ ). No other interactions were present (all  $F_s < 2.6$ ). Statistical analyses for the P600 window disclosed a main effect of acceptability for the lateral sites [ $F(1,25) = 10.15, p < .005$ ] and a trend for the midline sites [ $F(1,25) = 3.56, p < .08$ ]. The analyses thus confirmed that a P600 effect was present. No Acceptability  $\times$  Site interactions were present (midline and lateral sites:  $F_s < 1$ ). In addition, for the lateral sites, an Acceptability  $\times$  ROI interaction was found [ $F(1,25) = 6.00, p < .05$ ]. Separate analyses for the two levels of ROI (anterior, posterior) revealed an acceptability effect for the posterior sites [ $F(1,25) = 17.58, p < .001$ ], but not for the anterior sites ( $F < 1$ ).

## Discussion

As predicted, and in line with the recent ERP literature reviewed above, semantically implausible reversal sentences elicited a P600 effect instead of an N400 effect. Thus, the basic observation that gave rise to the different hypotheses described in the Introduction has been replicated. Semantically implausible nonreversal sentences, on the other hand, elicited an N400 effect. As in the van Herten et al. (2005) study, however, this N400 effect was again followed by a P600 effect. The appearance of a P600 effect does not seem to fit with the idea—as

was reasoned above—that there should be no conflict between sentence interpretations in nonreversal sentences, unless, of course, our nonreversal anomalies would still elicit some kind of conflict.

To investigate the latter possibility, we carefully checked our materials. It appeared to us that a large part of our sentences (about half) are at least partially plausible. Let us look, for example, at the semantically implausible sentence from Table 1, “The apple that climbed the tree. . . .” This implausible sentence contains a very plausible sentence part, which forms a meaningful and familiar unit, namely, climbing a tree. We would like to propose that the data pattern observed points to an intermediate possibility between Situations 2 and 3 described in the Introduction. That is, that a conflict can also arise between a *partially* plausible sentence interpretation (caused by the presence of a highly plausible unit) and the outcome of the parsing process—which indicates that the sentence is implausible—and that it is this conflict that gives rise to the P600 effect for the nonreversal anomalies in Experiment 1 and in the previous study of van Herten et al. (2005).

From a monitoring point of view this hypothesis makes sense, as in sentences with a plausible verb phrase, it could be that *the subject NP has been misread*. The biphasic N400/P600 pattern in the nonreversal sentences in Experiment 1 could subsequently be explained in the following way. The sentences that included a plausible unit elicited a P600 effect, whereas the sentences that did not include a plausible unit elicited the predicted N400 effect. Averaging both sentence types would then superimpose the P600 effect on the N400 effect, mimicking a biphasic pattern. To summarize, we hypothesized that the unexpected P600 effect could be the result of plausible sentence parts that were present in a large number of our semantically implausible nonreversal sentences.

## EXPERIMENT 2

In Experiment 2, it was investigated whether plausible sentence parts can indeed create a conflict, which in turn yields a P600 effect. We tested this by changing the verb in semantically plausible sentences in such a way that the combination of this verb and the object NP either formed a plausible or an implausible unit. This is not a trivial manipulation. Sentences employed in a typical study on semantic anomalies that demonstrated N400 effects may also contain such highly plausible units (e.g., “He drank his coffee with cream and dog”), but to our knowledge, the effect of the presence of plausible sentence parts in sentences, which as a whole violate the selectional restriction of the main verb, has not been systematically investigated yet. If the unexpected P600 effect is caused by plausible sentence parts, a P600 effect should correspondingly occur in sentences in which these plausible sentence parts are present, as opposed

to sentences that do not contain such a plausible sentence part. Plausibility of a sentence part was assessed by computing semantic relatedness (semantic relatedness value [SRV]) between object noun and verb by using latent semantic analysis (LSA; Landauer & Dumais, 1997). LSA is a technique that measures co-occurrence relationships between pairs of words. Although one could argue that LSA captures word associations rather than plausibility per se, Chwilla and Kolk (2002) showed that LSA is a sensitive measure for detecting subtle differences in (semantic) relatedness between words that were not associatively related. In addition, in a recent study (Chwilla & Kolk, in press) examining the access of world knowledge, word triplets were used that described a conceptual script (e.g., DIRECTOR–BRIBE–DISMISSAL). Such conceptual scripts are comparable to our plausible units as a script describes a typical—thus familiar and plausible—life event. A free association task in which multiple associates to the three words comprising the script-related triplets were required assured that the word triplets were *not* associatively or semantically related. Nevertheless, the LSA values were higher to those triplets that formed a script compared to control items. This finding bolsters the claim that LSA is sensitive to more abstract kinds of knowledge, such as script information.

As reasoned above, we hypothesized that only sentences that contain a plausible unit may create a conflict, and thus, elicit a P600 effect. Correspondingly, we predict that a P600 effect is only present in sentences in which an object noun and a verb are highly related (high SRV sentences), whereas it should be greatly reduced or absent in sentences in which object NP and verb are not highly related (low SRV sentences). In contrast, an N400 effect should be absent or reduced in the high SRV, but present in the low SRV sentences. This is because the unit of meaning selected in the high SRV sentences, which supposedly gives rise to the conflict, is a plausible one (e.g., playing in the park) and therefore does not cause integration difficulty.

## Methods

### *Participants*

Thirty-six participants (mean age = 23 years; 29 women) were tested. They fulfilled the same criteria as those in Experiment 1.

### *Sentence Material*

The sentence material consisted of Dutch sentences with the sentence structure *Jan zag dat NP<sub>subject</sub> NP<sub>object</sub> V en. . .* (translation: *John saw that NP<sub>subject</sub> NP<sub>object</sub> V and. . .*) (for an example, see Table 4). A different sentence structure than in Experiment 1 was used because for the LSA manipulations that we planned, we

**Table 4.** Examples of the Sentence Material from Experiment 2

<i>Condition</i>	
Acceptable	Jan zag dat de olifanten de bomen <i>omduwden</i> en hun mars door het oerwoud vervolgden.
	John saw that the elephants the trees <i>pushed-over</i> and their march through the jungle continued. <sup>a</sup>
	John saw that the elephants <i>pushed-over</i> the trees and continued their march through the jungle. <sup>b</sup>
Unacceptable, high SRV	Jan zag dat de olifanten de bomen <i>snoeiden</i> en hun mars door het oerwoud vervolgden.
	John saw that the elephants the trees <i>pruned</i> and their march through the jungle continued. <sup>a</sup>
	John saw that the elephants <i>pruned</i> the trees and continued their march through the jungle. <sup>b</sup>
Unacceptable, low SRV	Jan zag dat de olifanten de bomen <i>verwenden</i> en hun mars door het oerwoud vervolgden.
	John saw that the elephants the trees <i>caressed</i> and their march through the jungle continued. <sup>a</sup>
	John saw that the elephants <i>caressed</i> the trees and continued their march through the jungle. <sup>b</sup>

The critical words are italicized.

<sup>a</sup>Word-by-word translation.

<sup>b</sup>Paraphrase.

preferred using verbs without prepositions. In Dutch, however, verbs without prepositions implemented in subject-relative sentences form syntactically ambiguous sentences. As we wanted our sentences to be syntactically unambiguous, we had to change the sentence structure. Sentence acceptability was manipulated. That is, for each sentence, three variants were made; a semantically acceptable variant and two semantically unacceptable variants. The semantically unacceptable sentence variants both contained a selectional restriction violation (see Table 4; elephants cannot prune trees and neither can they caress them). The two variants differed in whether they included a plausible unit or not. The combination of the object NP and the verb either formed a plausible unit (e.g., pruning the trees, mowing the lawn, baking a cake) or not. Plausibility was assessed by computing semantic relatedness between object NP and verb by using the LSA method. LSA is a mathematical technique that generates a high-dimensional semantic space from the analysis of a large corpus of written texts. The meaning of a word is defined as a vector in this semantic space. Semantic relatedness of two words can be determined by calculating the cosine between their two vectors. The higher the cosine, the more semantically related words are. In the current study, semantically acceptable sentences were

transformed into semantically unacceptable sentences by changing the verb. The semantic relatedness was calculated between the object noun and a set of new transitive verbs. From this set, two new verbs were chosen: one verb whose semantic relatedness with the object NP was high (LSA value  $> +0.25$ ) and one verb whose semantic relatedness with the object NP was low (LSA value  $< +0.20$ ). The former sentences were termed “high SRV sentences,” and the latter were termed “low SRV sentences.” As topic space, “General Reading up to First Year of College” was used. The verbs of the three sentence variants were matched for length and (lemma) frequency (see Table 5). The subject and object NPs always carried the same number, in 56% both NPs were singular and in 44% they were plural. The experimental sentences were divided equally into three lists, each list contained only one variant of a sentence. Each list contained 90 experimental sentences: 30 acceptable sentences, 30 unacceptable high SRV sentences, and 30 unacceptable low SRV sentences. Ninety filler sentences were added, of which 60 were acceptable sentences and 30 were semantically unacceptable sentences. All filler sentences had a structure that was different from the experimental sentences in that the verb preceded the second noun. The unacceptable sentences included

**Table 5.** Description of the Sentence Material

<i>Condition</i>	<i>LSA Value</i>	<i>Word Length</i>	<i>Log Frequency</i>
Acceptable	0.181 (0.119)	7.433 (1.594)	1.358 (0.736)
Unacceptable high SRV	0.464 (0.163)	7.156 (1.498)	1.225 (0.673)
Unacceptable low SRV	0.077 (0.005)	7.422 (1.683)	1.213 (0.662)

Standard deviations are in brackets. All LSA values differ significantly ( $p < .001$ ).

Word length and log frequency do not differ.

a selectional restriction violation, which was either at the mid-sentence position or at the end of the sentence. This was done to encourage participants to pay attention to the entire sentence (for examples of filler sentences, see Table 2).

### Procedure

The procedure for Experiment 2 was the same as for Experiment 1, except for one change. For reasons outlined above, instead of a fixation cross, each sentence was preceded by a three-word-long carrier phrase *Jan zag dat* (“John saw that”), which had a duration of 765 msec.

### EEG Recording and Data Analysis

The EEG recording and the data analysis for Experiment 2 was the same as for Experiment 1.

### Validation Study of the Materials

A separate group of 26 participants was tested that fulfilled the same criteria as those in Experiment 1. The procedure was identical to the procedure for the validation study for Experiment 1.

Thirteen items that were miscategorized by at least half of the participants were omitted, and for the EEG experiment, were replaced with new items. MANOVA, including the factor sentence type, with levels “acceptable,” “unacceptable high SRV,” and “unacceptable low SRV,” revealed effects of sentence type [RT:  $F(2,50) = 14.15, p < .001$ ; Error percentages:  $F(2,50) = 10.79, p < .001$ ]. Follow-up paired  $t$  tests indicated that participants responded slower and made more errors to acceptable sentences than to unacceptable high SRV sentences (RT:  $t = 5.35, df = 25, p < .001$ ; Error percentages:  $t = 4.61, df = 25, p < .001$ ) and to unacceptable low SRV sentences (RT:  $t = 5.11, df = 25, p < .001$ ; Error percentages:  $t = 4.47, df = 25, p < .001$ ). The unacceptable high SRV and unacceptable low SRV sentences were responded to equally fast and accurate (RT:  $t = -1.15, df = 25, > .2$ ; Error percentages:  $t = -0.737, df = 25, p > .1$ ; see also Table 6). In short, the validation study shows that participants were successful in detecting the semantic violations, as the error percentage for the unacceptable sentences was, on average, only 5.44%. Unacceptable high SRV and unacceptable low SRV sentences did not differ in response time or number of errors, indicating that the difficulty level of these sentences was matched.

## Results

### Event-related Potentials

The grand-average waveforms are displayed in Figure 4 (acceptable sentences vs. unacceptable high SRV sen-

**Table 6.** Response Times and Error Percentages (in brackets) from the RT Validation Study

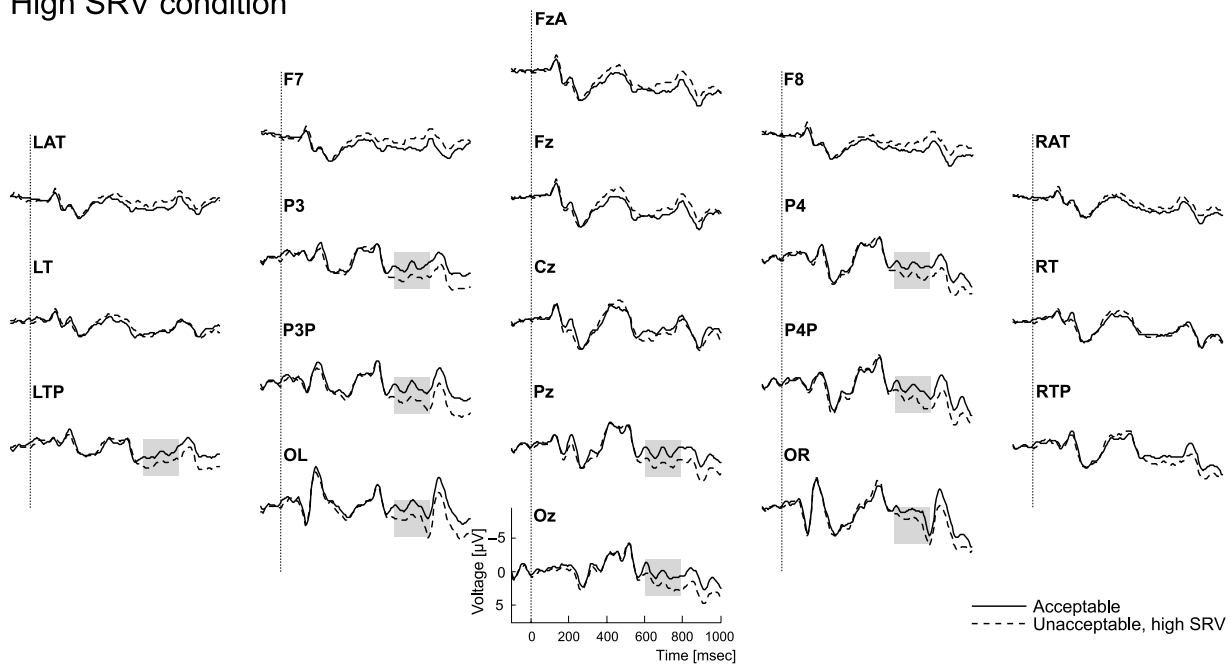
Acceptable	Unacceptable	
	High SRV	Low SRV
768	642	655
(15,98)	(5,94)	(4,94)

tences) and Figure 5 (acceptable sentences vs. unacceptable low SRV sentences). The overall form of the ERPs was similar to that in Experiment 1. Visual inspection of Figures 4 and 5 suggests that a standard N400 effect with maximal effects at central/posterior midline and bilateral posterior sites was present for the low SRV verbs, but not for the high SRV verbs. The P600 seems to be affected both by acceptability and SRV: A P600 effect seemed to be elicited by the high SRV verbs but not by the low SRV verbs. This P600 effect looked similar to the standard P600 effect elicited by syntactic anomalies in terms of its timing and scalp topography. To statistically confirm these apparent different ERP signatures, separate analyses were conducted in which the two levels of unacceptable SRV sentences (low SRV vs. high SRV) were compared with the acceptable versions of the sentences.

The analyses for the unacceptable *high* SRV sentences for the N400 window (400–500 msec) revealed a trend toward a main effect of acceptability for the lateral sites [ $F(1,35) = 3.75, p < .07$ ; midline:  $F < 2.3$ ], indicating that mean amplitudes tended to be more negative-going for unacceptable verbs than for acceptable verbs. To explore whether a significant difference between conditions was present at any of the sites, single-site analyses were performed. These additional analyses revealed a significant difference at two frontal sites (F3 and F4,  $p < .05$ ). The topography of this negative effect does not match that of the standard N400 effect, but may reflect a left anterior negativity effect, which typically shows an anterior distribution. No other interactions were present ( $F_s < 1.5$ ). To further determine that no N400 effect was present in Experiment 2, supplementary analyses were conducted, using a broader latency window of 300–500 msec. These analyses confirmed that no N400 effect was present for the high SRV sentences (Acceptability: midline:  $F < 1.5$ , lateral:  $F < 2.6$ ; In addition, no significant interactions of acceptability with site, hemisphere, and/or ROI were present, all  $F_s < 2.6$ ).

The statistical analyses for the P600 window (600–800 msec) yielded no main effects of acceptability (midline and lateral:  $F_s < 1$ ) but Acceptability  $\times$  Site interactions both for midline [ $F(4,140) = 5.36, p < .005$ ] and lateral sites [ $F(4,140) = 3.31, p < .05$ ]. In addition, an Acceptability  $\times$  ROI interaction was found for the lateral sites, indicating that a P600 effect was present at posterior sites [ $F(1,36) = 14.52, p < .005$ ], whereas a reversed

## Experiment 2 High SRV condition

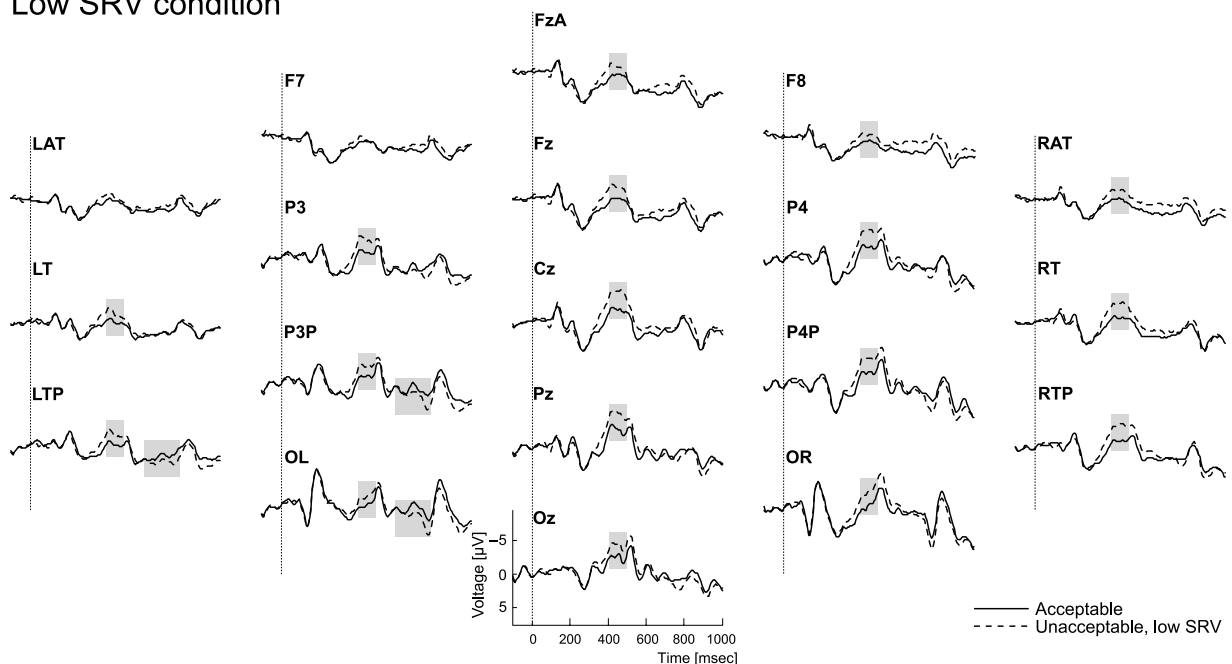


**Figure 4.** Grand-average waveforms for all midline sites and a subset of lateral sites. Averages are time-locked to the onset of the critical verb and superimposed for the acceptable and the unacceptable high SRV verbs.

effect was present at anterior sites [i.e., more positive-going mean amplitudes for acceptable than for unacceptable verbs;  $F(1,36) = 6.23, p < .05$ ]. This reversed

effect seemed to reflect a prolongation of the early negative anterior distributed effect observed in the N400 window. Follow-up single-site analyses for the midline

## Experiment 2 Low SRV condition



**Figure 5.** Grand-average waveforms for all midline sites and a subset of lateral sites. Averages are time-locked to the onset of the critical verb and superimposed for the acceptable and the unacceptable low SRV verbs.

sites revealed P600 effects at Pz and Oz ( $p < .005$ ) and for all lateral sites ( $p < .05$ , except for RTP:  $p < .06$ ), whereas reversed effects were present for a number of anterior sites.

The analyses for the unacceptable *low* SRV sentences for the N400 window (400–500 msec) revealed clear N400 effects as reflected by main effects of acceptability for the midline [ $F(1,35) = 17.61, p < .001$ ] and for the lateral sites [ $F(1,35) = 16.44, p < .001$ ]. In addition, two-way interactions of Acceptability  $\times$  Hemisphere [ $F(1,35) = 5.84, p < .05$ ], Acceptability  $\times$  Site [midline:  $F(4,140) = 4.70, p < .005$ ; lateral:  $F(4,140) = 3.19, p < .05$ ], as well as a three-way interaction of the latter factors with ROI [ $F(4,140) = 3.95, p < .05$ ], were present. Single-site analyses revealed N400 effects at all midline sites ( $p < .05$ ). The interactions for the lateral sites reflected that the N400 effect was broadly distributed across the scalp with largest effects at bilateral posterior sites. Only for the right hemisphere were reliable N400 effects also obtained at anterior sites (F8, F4A, F4, RAT;  $p < .05$ ). The statistical analyses for the P600 window (600–800 msec) showed that unacceptable *low* SRV sentences did not elicit main effects of acceptability (midline and lateral:  $F_s < 1$ ) or Acceptability  $\times$  Site interactions (midline and lateral:  $F_s < 1$ ). For the lateral sites, interactions between acceptability and hemisphere [ $F(1,35) = 8.06, p < .01$ ] and between acceptability and ROI [ $F(1,35) = 20.91, p < .001$ ] were found. Follow-up analyses indicated that no acceptability effects were present neither for the left nor for the right hemisphere (both  $F_s < 1$ ). Acceptability  $\times$  ROI interactions [left hemisphere:  $F(1,35) = 17.04, p < .001$ ; right hemisphere:  $F(1,35) = 17.09, p < .001$ ] were, however, disclosed. Supplementary analyses showed that a P600 acceptability effect was present at posterior sites for the left hemisphere [ $F(1,35) = 7.23, p < .05$ ]. To test whether the P600 effects found for high SRV verbs were significantly larger than for low SRV verbs,  $t$  tests were performed on the difference scores (unacceptable mean amplitude minus acceptable mean amplitude). Follow-up tests showed that the P600 effect was larger for the high SRV verbs than for the low SRV verbs at two of four sites (P3P and OL;  $p < .05$ ).

## Discussion

As predicted, sentences in which object and verb formed a plausible unit elicited a P600 effect in the absence of an N400 effect. Conversely, in sentences in which object and verb did not form a plausible unit, the P600 effect was greatly reduced. It disappeared completely at the midline and was present only at four left posterior sites.<sup>1</sup> Moreover, at two of these four sites, the effect was significantly smaller than the P600 effect in sentences that included a familiar unit. In addition, a standard N400 effect was elicited. We conclude that our hypothesis was confirmed. It seems that there is an intermedi-

ate possibility between Situations 2 and 3 indeed. In particular, a conflict between a heuristic and an algorithmic route can also be created if the algorithmic route produces an implausible interpretation and the heuristic route a partially plausible interpretation.

Before turning to the General Discussion, however, an alternative hypothesis about the absence of the N400 effect in the high SRV sentences should be considered. One could argue that the absence of the N400 effect may be the result of word association priming. More precisely, the N400 effect to an implausible word completing a sentence is reduced if this implausible completion is associatively (Kutas & Hillyard, 1983) or categorically (Federmeier & Kutas, 1999) related to the expected completion. In a similar way, in the high SRV sentences, the object noun “trees” may have primed the associatively related verb “pruned,” which would then give rise to a reduction of the N400 effect. This could have eliminated any effect of the anomaly on the N400, which would have led us to the erroneous conclusion that readers (temporarily) missed the anomaly. The associative priming hypothesis can explain the absence of the N400 effect equally well. However, we favor the monitoring hypothesis for the following reason. Although the associative priming view can account for the absence of an N400 effect in Experiment 2, it cannot account for the absence of an N400 effect in the reversal sentences of Experiment 1 (as correct and incorrect sentences included the same set of critical words) and it cannot also account for the presence of a P600 effect instead of an N400 effect in Experiments 1 and 2.

## GENERAL DISCUSSION

Much recent work, in particular in the action domain, has demonstrated that the brain is involved in error monitoring. So far, however, this evidence relates to errors in production only. The current study indicates that the brain also monitors for errors of perception. If an interpretation based upon the ensemble of individual word meanings clashes with one based upon a full syntactic analysis of the sentence, a P600 effect occurs while the N400 effect of semantic anomaly is absent. According to the monitoring framework presented here, the conflict between the two interpretations gave rise to a monitoring response in which the previous input string is reprocessed to check for a possibility of a processing error. Our findings in Experiment 2 extend the number of situations under which semantically anomalous sentences can show this pattern. Even implausible sentences that include plausible sentence parts can elicit P600 instead of N400 effects. Apparently, the presence of a highly familiar and in that sense meaningful concept, such as “playing in the park” or “milking the cows,” in syntactically unambiguous sentences is

sufficient to lead to a conflict. As the present data show, the influence or semantic attractiveness of such meaningful concepts can be so strong that syntactic processing is overruled. As Kim and Osterhout (2005) forcefully argue, these findings have important implications for our thinking about sentence processing. They appear incompatible not only with syntax-first, but also with constraint-based satisfaction models. Further implications of our data concern (a) the nature of heuristic strategies, (b) the notion of conflict monitoring in language perception, and (c) the functional significance of the P600.

### **Heuristics and Shallow Processing in Sentence Understanding**

In our opinion, the results have important consequences for the hypothesis that sentence processing includes the use of heuristics. First of all, the results provide compelling evidence for the existence of a plausibility strategy: How else can one explain that the language system—at least temporarily—misses the anomaly of “The fox that shot the poacher” or even of “John saw that the elephants pruned the trees”? Second, it appears that the plausibility heuristic can overrule the word-order heuristic, as the preferred interpretation of “the fox that shot the poacher” corresponded to the noncanonical word order. Third, there are also consequences for what a plausibility strategy can consist of. As originally formulated (Bever, 1970), it takes the lexical items of the sentence and combines them in the most plausible way. The results of our second experiment suggest that the language system not only attempts to combine *all* items into a single representation, but also looks if *subsets* of these items can be meaningfully related.

If language users would employ a plausibility strategy, they would ignore—at least temporarily—syntactic structure. Is it reasonable to assume that such a rich source of information is completely bypassed? In a recent review paper on shallow processing in language understanding, Sanford and Sturt (2002) argue that fully specified representations are often not possible or not useful to construct. They are not possible, for example, whenever insufficient information is provided to resolve an ambiguity, which happens regularly in natural language settings. They are not useful when full specification requires a large amount of working memory capacity, as will readily happen, for instance, with the use of multiple quantifiers. In that case, it is more profitable to allow more than one interpretation. That language users actually employ such underspecified representations has been demonstrated in a number of different ways. A well-known example comes from a study by Barton and Sanford (1993) who asked participants to respond to questions such as “After the air crash, where were the survivors buried?” and observed that half of

the participants did not notice the anomaly. A recent study by Nieuwland and Van Berkum (2005) demonstrated the existence of semantic illusions such as that above in an ERP paradigm. In this study, they employed sentences such as “She told the suitcase. . . .” Presented in isolation, such sentences elicited an N400 effect. However, embedded in the context of a conversation at—in the case of the example—a check-in desk at an airport, there was a striking absence of a standard N400 effect. Instead of this, a late positivity emerged around 500–600 msec, with a peak latency within the 900–1100 msec window. The absence of an N400 effect was taken as evidence for a “semantic illusion.” The absence of an N400 effect in our present nonreversal semantic anomalies can similarly be taken as evidence for the existence of such semantic illusions. However, our results, as well as those obtained by Nieuwland and Van Berkum (2005), also suggest that this kind of shallow processing goes on simultaneously with syntactic parsing. It is the conflict between the two kinds of processing that, we propose, underlies the late positivity. It is this ERP effect to which we will turn now.

### **Conflict Monitoring in Sentence Perception**

Monitoring refers to a process of cognitive control aimed at output optimization. In the above, we have argued for the existence of a monitoring process in language perception. This monitoring would be triggered by a conflict, as has been proposed for the action domain. At the sentence level, this conflict can probably be described best as one between different response tendencies: the tendency to see the sentence as plausible on the one hand, and the tendency to interpret the sentence as implausible on the other hand. In the reversal sentences of Experiment 1 (e.g., “The fox that shot the poacher”), the tendency to see the sentence as plausible is strengthened by the fact that the interpretation provided by the lexical strategy (that of poachers shooting foxes) refers to a very likely event in the real world. The tendency to see the sentence as implausible is also strengthened, this time by the fact that the interpretation provided by the parser (that of foxes shooting poachers) refers to a very unlikely real world event, as it constitutes a selectional restriction violation of the main verb. In the high SRV sentences of Experiment 2 (e.g., “. . . that the elephants pruned the trees”), the tendency to see the sentence as plausible is strengthened by the fact that the partial interpretation provided by the lexical strategy (that of “pruning the trees”) refers to a highly likely event. In contrast, the tendency to see these sentences as implausible is strengthened by the fact that the interpretation of the whole sentence provided by the parser (that of elephants pruning trees) refers to a very unlikely real-world event, as it again constitutes a selectional restriction violation of the main verb. In the low SRV sentences



of Experiment 2, on the other hand, the contrast between the two response tendencies seems much smaller, resulting in a reduction of the P600 effect. In these sentences (e.g., "...that the elephants caressed the trees"), the tendency to see the sentence as plausible is not strengthened by the fact that a partial interpretation provided by the lexical strategy (that of "caressing the trees") refers to a likely real-world event—although the event is possible. The tendency to see these sentences as implausible, on the other hand, is strengthened by the fact that the interpretation of the whole sentence provided by the parser (that of elephants caressing trees) refers to a very unlikely real-world event. Thus, whereas in the reversal and the high SRV sentences, the critical contrast would be one between highly plausible and highly implausible, the contrast in the low SRV sentences would be one between quite implausible and very implausible.

We have proposed that the conflict between two response tendencies triggers a monitoring reaction. This process would involve reprocessing the memory trace of the linguistic string that led to the conflict. The notion that readers actually go back to what was processed a moment ago is supported by results of eye movement studies. Not only do eyes make saccades and fixate stimuli for a certain amount of time, they also occasionally return to a position already fixated before. This phenomenon is referred to as "regression." Regressions occur among others when the stimulus display is changed during a saccade (Binder, Pollatsek, & Rayner, 1999). It seems clear that in such circumstances, readers refixate to check for the possibility of a processing error. It is as if the reader asks herself: Can I believe my eyes? If the proposal that a conflict in language perception leads to reprocessing of the memory trace of the input string is right, one would expect that if the input string would still be visible, the eyes would refixate this string. This is indeed what seems to happen. Garden path sentences (e.g., "The woman persuaded to answer the door...") not only elicit P600 effects, as we saw above, but also an increase in the number of regressions (e.g., Frazier & Rayner, 1982). The same holds for morpho-syntactic anomalies which not only elicit P600 effects but also co-occur with an increase in the number of regressions (Ni, Fodor, Crain, & Shankweiler, 1998). Below, we will argue that these syntactic violations also embody a conflict, a conflict which could be responsible both for the P600 effect and for the regression increase.

As described above, monitoring in the action domain manifests itself in an ERN. It is worth noticing that, although most EEG studies on error monitoring have concentrated on the ERN, errors in choice RT tasks are generally followed by a late positivity as well (for a review, see Falkenstein, Hoormann, Christ, & Hohnbein, 2000). This late positivity occurs between 200 and 500 msec after an incorrect trial, has a parietal distribution, and peaks at 300 msec, with a maximum at the vertex. Falkenstein

et al. discuss several possible functions of this "error positivity" and conclude that it probably reflects some kind of posterior processing, the nature of which is still unclear. We would like to propose here that this positivity and the P600 effect we observed in our studies carry similar functions. One possibility—but one that needs further study—is that the error positivity reflects reprocessing of the stimulus—or of its memory trace—to assess whether the erroneous response was due to faulty stimulus processing. One might criticize this hypothesis because of the large differences in the timing of these positive waves (from about 300 up to 1100 poststimulus). However, just as the latency of the P300 has been shown to be a function of the stimulus evaluation time (e.g., Donchin, 1979), which is unrelated to response selection processes (e.g., McCarthy & Donchin, 1981) and independent of behavioral response time (e.g., Verleger, 1997), the latency of these positivities may vary with the difficulty of checking the perceptual input or its memory trace for possible processing errors.

### **The Functional Significance of the P600**

One current view on the functional significance of the P600 is that it reflects syntactic (re)processing of some kind (Kuperberg et al., 2003; Hagoort et al., 1993). Accordingly, attempts have been made to account for the occurrence of P600 effects of semantic anomaly by assuming particular syntactic processes to be triggered by these anomalies, but as was discussed above, these proposals are not without problems. Furthermore, the occurrence of a late positivity after sentences such as "She told the suitcase..." provided such sentences were embedded in a biasing context (Nieuwland & Van Berkum, 2005), also seems hard to reconcile with an exclusively syntactic function of the P600. Our view on the P600 (van Herten et al., 2005; Kolk et al., 2003) is that it does reflect reprocessing indeed, but that its function is a general one and refers to checking the possibility of processing errors. This monitoring process is triggered by a conflict between two or more incompatible representations. These conflicting representations may stem from the use of two different processing routes, as we argued was the case with the present semantic reversal and nonreversal anomalies. Alternatively, the conflict may be between a lexical representation strongly predicted by the discourse context and the lexical representation activated by the actual word (Nieuwland & Van Berkum, 2005).

What about other situations known to elicit P600 effects, can they also be characterized as representing some kind of conflict? It seems that this is the case indeed. In the case of grammaticality violations, the conflict may be between the strongly predicted grammatical morpheme and the actually observed morpheme. However, in a list of sentences with ungrammatical morphemes, P600 effects are elicited by grammatical morphemes (Coulson, King,

& Kutas, 1998b), due to a conflict between a strongly expected ungrammatical and an observed grammatical morpheme. Garden path sentences (e.g., “The woman persuaded to answer the door. . .”) generally elicit P600 effects (e.g., Osterhout & Holcomb, 1992) and this would be due to a conflict between the preferred parse and the less preferred one. Finally, P600 effects are seen in some grammatical but complex sentences, as compared to less complex controls (Kaan, Harris, Gibson, & Holcomb, 2000). The research on action monitoring that was discussed above has suggested that monitoring—as indicated by ACC activity—occurs in difficult tasks even in trials in which no error was made. It makes sense that one monitors for errors when the task is difficult as the chance of making an error is large. Such monitoring may underlie the P600 effect in highly complex sentences.

In our view, the function of the P600 is more general than just a syntactic one (see for a similar proposal Coulson, King, & Kutas, 1998a, 1998b).

### Acknowledgments

This research was supported by Grant 425-202-02 from the Netherlands Organization for Scientific Research (NWO). Portions of this research were presented at the 11th Annual Meeting of the Cognitive Neuroscience Society in San Francisco, USA, 2004 and at the 18th Annual CUNY Conference on Human Sentence Processing in Tucson, USA, 2005. We thank Fernanda Ferreira and Lee Osterhout for their helpful comments on earlier versions of this article. We are grateful to the ERG group of the NICI for technical assistance.

Reprint requests should be sent to Marieke van Herten, IWTS, Max-Planck-Institut für Psycholinguistik, P.O. Box 310, 6500 AH Nijmegen, The Netherlands, or via e-mail: marieke.vanherten@mpi.nl.

### Note

1. A biphasic N400/P600 pattern is also reported in metaphor comprehension (Coulson & Van Petten, 2002), as in sentences like “My lawyer is a shark.” The authors argue that the enlarged N400 and late positivity seen to these sentences represent the more effortful processing that is required for metaphorical comprehension. An alternative viewpoint thus could be that the N400 effect in the low SRV sentences, instead of representing problematic semantic integration due to the inclusion of a semantic implausibility, could also represent—together with the late positivity—effortful semantic processing due to participants trying to make sense out of our sentences by checking whether they can be understood metaphorically (elephants gently touching trees?). We cannot exclude this possibility but, on the other hand, it is not necessarily incompatible with our view. On the contrary, the monitoring theory can account for the ERP pattern in metaphor comprehension in the following way. Initially, the literal meaning is considered and the sentence is interpreted as incorrect (as sharks are fish, not people), which would elicit an N400 effect. The quick discovery that the sentences can be understood metaphorically would then trigger a reprocessing/monitoring process to check upon this possibility.

### REFERENCES

- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477–487.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.
- Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1162–1172.
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*, 572–582.
- Chwilla, D. J., & Kolk, H. H. J. (in press). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*.
- Chwilla, D. J., & Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Memory & Cognition*, *30*, 217–225.
- Chwilla, D. J., Kolk, H. H. J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*, 314–341.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed processing approaches. *Psychological Review*, *100*, 589–608.
- Coulson, S., King, J., & Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, *13*, 653–672.
- Coulson, S., King, J., & Kutas, M. (1998b). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–38.
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*, 958–968.
- Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begeleiter (Ed.), *Evoked potentials and behavior* (pp. 13–75). New York: Plenum.
- Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*, 87–107.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.
- 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.
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113–157.

- Hoeks, 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.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.
- Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36.
- Kolk, H. H. J., & Weijts, M. (1996). Judgements of semantic anomaly in agrammatic patients: Argument movement, syntactic complexity and the use of heuristics. *Brain and Language*, *54*, 86–135.
- Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (in press). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.
- Kutas, M., & Hillyard, S. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*, 539–550.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*, 77–80.
- Ni, J., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515–539.
- Nieuwland, M. S., & Van Berkum, J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, *24*, 691–701.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 282–386.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge: MIT Press.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.
- Verleger, R. (1997). On the utility of P300 latency as an index of mental chronometry. *Psychophysiology*, *33*, 131–156.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.

## AUTHOR QUERIES

### **AUTHOR PLEASE ANSWER QUERIES**

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. The sections must be in the order Introduction, Results, Discussion, Methods. Please check if the section headings can be rearranged as per journal specifications for this article.
2. Please provide location of manufacturer (city and state) Electrocap International.
3. Experiment 1, Event-related Potentials section: The sentence “All conditions elicited for visual stimuli characteristic early ERP response—that is, an N1 followed by a P2, which at occipital sites was preceded by a P1 component.” was changed to “All conditions elicited for visual stimuli a characteristic early ERP response—that is, an N1 followed by a P2, which at occipital sites was preceded by a P1 component.” Please check if the original message was retained.
4. Experiment 1, Discussion: VP was changed to verb phrase.
5. Please update the following references Chwilla & Kolk, in press; Kuperberg et al., in press if it has already been published.

**END OF ALL QUERIES**