Language and Linguistics Compass 3/5 (2009): 1211–1224, 10.1111/j.1749-818x.2009.00163.x

Monitoring in Language Perception

Nan van de Meerendonk,[1]* Herman H.J. Kolk,[1] Dorothee J. Chwilla[1] and Constance Th.W.M. Vissers[2]
[1]*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour and* [2]*UMCN St Radboud*

Abstract

Monitoring is an aspect of executive control that entails the detection of errors and the triggering of corrective actions when there is a mismatch between competing responses or representations. In the language domain, research of monitoring has mainly focused on errors made during language production. However, in language perception, for example while reading or listening, errors occur as well and people are able to detect them. A hypothesis that was developed to account for these errors is the monitoring hypothesis for language perception. According to this account, when a strong expectation conflicts with what is actually observed, a reanalysis is triggered to check the input for processing errors reflected by the P600 component. In contrast to what has been commonly assumed, the P600 is thought to reflect a general reanalysis and not a syntactic reanalysis. In this review, we will describe the different studies that led to this hypothesis and try to extend it beyond the language domain.

Humans adjust themselves to the circumstances they are faced with, and an important information source for these adjustments is errors. For example, an American crossing a street in England, where they drive on the opposite side of the road, may automatically but erroneously look left first. However, with more experience (e.g., an oncoming car honking) he will learn to control this automatic response and implement the correct behavior for the traffic system in England (i.e., look right first). Just as errors are important in daily life, they have an important regulating effect on cognitive processes. The process of executive control through which the quality of our behavior is assured is called 'monitoring'. Monitoring entails the detection of errors and the triggering of corrective actions when there is a mismatch between competing responses or representations.

In this review, the focus will be on monitoring in language perception. We will first describe different models that have been proposed for monitoring as such and relate them to the monitoring hypothesis that we propose for language perception. Then we will go into the studies that led to the proposal of this hypothesis, and describe the experiments that tested the account. Furthermore, we will discuss commonalities and differences regarding the event-related potential (ERP) patterns found in these studies, and end with a possible extension of the hypothesis beyond the language domain.

Models of Monitoring

In different cognitive domains, a monitoring process has been proposed. For example, in the action domain, it has been proposed that ongoing actions and outcomes are monitored and evaluated with respect to internal goals. When an error is made in a choice reaction time task, like the Eriksen flanker task (Eriksen and Eriksen 1974), the error-related negativity (ERN) component is elicited in the electroencephalogram (EEG) around 100 ms after the erroneous response. The ERN is commonly thought to reflect a monitoring process that, when a mismatch is detected between the executed and correct response, signals that an error was made. To date, it is debated what actually triggers this monitoring process and which aspect of action monitoring is reflected by the ERN (for a review, see Yeung et al. 2004). Two views that are important for the present review are the mismatch theory (e.g. Coles et al. 2001) and the conflict monitoring theory (e.g. Botvinick et al. 2001). In the mismatch theory, the monitor is seen as a comparator that compares the executed with the correct response. When there is a mismatch, the comparator sends an error signal to a remedial action system and the ERN is generated. The remedial action system initiates actions to inhibit or correct the error, and makes strategic adjustments to reduce the likelihood for future errors (Coles et al. 2001). Holroyd and Coles (2002) extended the mismatch theory by developing the reinforcement-learning theory. According to this theory, the ERN reflects a reinforcement-learning signal that is carried by the midbrain dopamine system and is transmitted to the anterior cingulate cortex (ACC). The ACC uses the reinforcement-learning signal to recognize the appropriate motor controller and optimize performance accordingly. In contrast to the mismatch theory, in the conflict monitoring theory, monitoring is not seen as a comparison process, but involves the detection of coactivated incompatible responses. After an error has been made, processing continues and the correct response is activated internally. The coactivated responses create a conflict that, when exceeding a certain threshold, is detected by the ACC. The ACC then signals other brain areas to make adjustments in control. The ERN is thought to reflect the response conflict. Therefore, according to this account, the ERN is also a sign for error detection but does not reflect the error-detection process as such (Botvinick et al. 2001).

In the language domain, it is generally accepted that monitoring takes place in production. Monitoring in production becomes apparent when we overtly repair our speech. Take for example the following overt self-repair reported by Levelt (1983):

(1) Go from left again to, uh …, from pink again to blue.

This example shows that we can detect when our utterances deviate from our intentions, and are able to correct them. According to Levelt's (1983, 1989) perceptual loop theory of self-monitoring, speakers continuously compare their intentions with the planned or produced utterance, and when there is a mismatch a repair is made. The monitor in Levelt's model therefore is hypothesized to be a central, conscious process that oversees end-products of speech production (Postma 2000). Another model that deals with speech monitoring is the node structure theory (NST) by MacKay (1987, 1990, 1992). MacKay assumes that errors are detected automatically by the activation patterns in the node system. Take for example the utterance of *srace* instead of *space*. In the node system, there is no mother node with strong connections (committed node) that represents the initial consonant cluster *sr*. Therefore, when daughter nodes *s* and *r* are activated together, an uncommitted mother node is primed that represents initial constant clusters and is weakly connected to the *s* and *r* nodes. Due to these weak connections with the daughter nodes, the uncommitted node cannot be inhibited. Inhibition occurs after a node has been activated, and is needed to prevent the node to become activated again. In the case of inhibition failure, activation is prolonged, inducing awareness leading to error detection. Often, however, a phonological speech error results in a real word and not a non-word (lexical bias effect). An example is the utterance *cool tarts* instead of *tool carts*. According to the NST, these errors are harder to detect, as committed nodes exist for the segments, syllables and words involved. The error could still be detected; since now no committed node exists higher up in the hierarchy, as no noun phrase (NP) node corresponding to *cool tarts* will be available (MacKay 1992). As in the action monitoring literature, these two models of monitoring in language production are in disagreement about the interpretation of the monitoring process. Levelt's perceptual loop theory, as well as the mismatch theory by Coles et al. (2001), assumes a continuous comparator process for monitoring. In contrast, MacKay's NST, as well as the conflict theory by Botvinick et al. (2001), does not assume that such a continuous comparison process is needed to detect errors. Error detection is thought to be automatic when, due to failed inhibition, prolonged activation of a node leads to awareness.

So far, monitoring in the language domain has been exclusively studied in language production. However, in language perception, errors are made as well. For example, during a conversation, we sometimes misunderstand the speaker and experience 'slips of the ear' (Cutler and Butterfield 1992). For example, a mishearing of guests in a restaurant who were listening to a waiter was 'Foot on the table', while he had actually produced 'Food on the table'. Besides these occasional mishearings, we also sometimes make errors while reading a text. For example, Kaufman and Obler (1995) showed that normal adult readers make various 'slips of the eye' (e.g. omissions/insertions of words/letters, parsing errors, homonym errors). These reading errors can occur at various levels, like the syntactic (reading 'posts' as a noun instead of a verb), phonological (reading 'sun' as 'son') and semantic level (reading 'Armenians' instead of 'Americans'). How could these perceptual errors be monitored for? In language perception, it is not possible, as in production, to observe the errors directly, as they do not lead to observable behavior and, furthermore, the intentions of the other person are unknown. Therefore, a comparison between intentions and actual events, as is assumed in the mismatch theory (Coles et al. 2001) and the perceptual loop theory (Levelt 1983, 1989), is not possible. However, a listener can be assumed to have expectations about what the speaker or writer could intend. Therefore, we propose that a strong conflict between expected and perceived representations can trigger a reanalysis of the input to check for processing errors, reflected by the P600 component in the EEG. Therefore, as the conflict monitoring theory proposes for action monitoring (Botvinick et al. 2001), we hypothesize that the conflict between the different representations constitutes the internal signal to detect errors of language perception. The conflict functions as a bottom-up signal, for which no comparison process is needed. Whether conflict detection in the language domain engages the same neural circuitry as proposed for the action domain (e.g. Botvinick et al. 2001) is yet unclear.

P600 Effects to Non-Syntactic Violations

The first proposal of conflict monitoring in language perception arose when different studies unexpectedly reported P600 effects to semantic violations. Until then, it had been generally accepted that semantic violations elicit only an N400 effect. The N400 is a negative-going ERP component that peaks around 400 ms after critical word onset. Its scalp distribution is widespread, but usually more negative over central and parietal electrode sites with a right hemisphere preponderance (Kutas and Van Petten 1994). Kutas and Hillyard (1980c) discovered that the N400 was more negative in response to semantically incongruous words compared with congruous ones. Furthermore, a semantic violation was not necessary; the N400 was also found to be more negative to less-expected plausible sentence endings compared with expected ones. In addition, an unexpected word that was semantically related to the expected word elicited a smaller N400 than unrelated unexpected words (Kutas and Hillyard 1984). This amplitude difference between congruous/expected and incongruous/unexpected words is referred to as the N400 effect. One account of the N400 component is that its amplitude reflects how easily a word can be integrated into the current context (e.g. Holcomb 1993; Chwilla et al. 1998; Van Berkum et al. 1999). An alternative view is the lexical access account, according to which the N400 reflects the ease with which a word can be accessed from long-term memory (e.g. Federmeier and Kutas 1999; Kutas and Federmeier 2000; Lau et al. 2008).

While the N400 is sensitive to semantic violations, the P600 was thought to be an index of syntactic processing. The P600 is a positive-going ERP component that occurs between 500 and 800 ms after critical word onset. It usually has a central–posterior scalp distribution and was discovered by Osterhout and Holcomb (1992). An increase in P600 amplitude has been reported to various syntactic anomalies (e.g. Friederici et al. 1993; Hagoort et al. 1993; Münte et al. 1998), complex sentences (Kaan et al. 2000) and locally ambiguous sentences (Osterhout and Holcomb 1992). Therefore, the P600 has been thought to reflect syntactic reanalysis or repair processes. The amplitude difference between ungrammatical or more complex/ambiguous sentences and grammatical or unambiguous sentences has been referred to as the P600 effect.

However, contrary to what was expected, Kolk et al. (2003) found a P600 effect to semantic anomalies like:

(2) De kat die voor de muizen vluchtte
The cat that from the mice fled$_{[sg]}$ … (literal translation)
The cat that fled$_{[sg]}$ from the mice … (paraphrase)

Likewise, Kuperberg et al. (2003) also found a P600 effect to semantic verb–argument violations (see also Kuperberg et al. 2006, 2007). An example of such a violation is the following:

(3) For breakfast the eggs would only eat ….

Furthermore, Hoeks et al. (2004) and Kim and Osterhout (2005) reported a P600 effect to semantic verb–argument violations like:

(4) De speer heeft de atleten geworpen. (Hoeks et al. 2004)
The javelin has the athletes thrown. (literal translation)
The javelin has thrown the athletes. (paraphrase)

(5) The hearty meal was devouring … (Kim and Osterhout 2005)

Surprisingly, though all studies contained syntactically *unambiguous* sentences with semantic violations, none of them reported an N400 effect, but all reported a P600 effect. To preserve the syntactic interpretation of the P600 effect, in some of these studies it was proposed that the P600 effect reflected a reanalysis triggered by a discrepancy in the probable thematic role of the NP and the thematic role assigned by the verb (e.g. Kuperberg et al. 2003; Hoeks et al. 2004). For instance, in example (3), the inanimate subject NP (*the eggs*) violated the thematic structure of the verb (*eat*), causing an attempt to reassign the thematic role of *the eggs* from agent to theme.

Another account that tried to reconcile the P600 effect to semantic anomalies with the syntactic interpretation of the P600 was the syntactic prediction hypothesis (Kim and Osterhout 2005; Van Herten et al. 2005). According to this hypothesis, it is not the role reassignment as such, but the mismatch between a predicted and observed number inflection on the verb that triggers the P600 effect. The prediction for a certain inflection arises because of a strong semantic relationship between the verb and its preceding argument. Due to this 'semantic attraction', the role assignment signaled by these semantic cues is thought to be pursued [e.g. in example (5), *meal* is assigned a theme role instead of agent] eliciting an expectation for a certain inflection on the verb (*-ed* instead of *-ing*). The sentences that were used in the Kolk et al.'s (2003) study also contained such mismatches. Take example (2), based on general world knowledge, this sentence should be interpreted as if the mice were fleeing from the cat. Therefore, a plural inflection is expected; however, a singular inflection is encountered. To test whether this mismatch elicited the P600 effect, Van Herten et al. (2005) added sentences in which the subject and object had the same number, e.g.:

(6) De kat die voor de muis vluchtte ….
The cat that from the mouse fled$_{[sg]}$ … (literal translation)
The cat that fled$_{[sg]}$ from the mouse … (paraphrase)

If the syntactic prediction hypothesis was correct, only the sentences in which the subject and object had a different number should elicit a P600 effect. However, the results showed that, independent of the presence of a mismatch, a P600 effect was elicited, thereby ruling out the syntactic prediction hypothesis.

Kolk et al. (2003) put forward a different interpretation to explain the P600 effects to semantic violations. They introduced the monitoring hypothesis that was mentioned in the beginning: a strong conflict between representations triggers reanalysis to check for possible processing errors. The competing representations in the Kolk et al.'s (2003) study were elicited on the one hand by a plausibility heuristic, which combines the individual word meanings related to general world knowledge (mice flee from cats). On the other hand, the syntactic parser generates an outcome (cat flees from mice). Therefore, just as the semantic attraction interpretation, the plausibility heuristic is based on the semantic relations between the words, which are combined into a mental model of the conceptual representation. However, according to the monitoring hypothesis, the P600 does not reflect a syntactic reanalysis, but has a more general function.

Vissers et al. (2007) tested whether the P600 effects in the Kolk et al.'s (2003) and Van Herten et al.'s (2005) studies were indeed elicited due to a conflict between the representations resulting from the plausibility heuristic and the parser. To this end, they replicated the Kolk et al.'s (2003) study. However, the participants were now instructed to direct their attention to the syntactic structure of the sentences. This was carried out by informing the participants that semantic reversal anomalies would be present in the experiment, and that they should not be deceived but focus on the sentence structure as such. If the P600 reflects a reanalysis triggered by a conflict, the P600 effect should now be diminished or absent due to a reduced conflict because anomalies were expected. The results of this study indeed showed a significant reduction in the P600 effect.

In addition, Van Herten et al. (2006) showed that, to create a conflict between representations, it is not necessary for the whole sentence to be interpreted as plausible. In the first experiment of this study (Van Herten et al. 2006) and in a previous study (Van Herten et al. 2005), a biphasic N400–P600 pattern was observed to semantically implausible non-reversal sentences to which an N400 effect was expected, like:

(7) De koning die van de baby beviel ....
The king who to the baby gave birth … (literal translation)
The king who gave birth to the baby … (paraphrase)

The materials were investigated and it was found that about half of the sentences contained plausible sentence parts forming meaningful units (e.g. giving birth to a baby). The authors hypothesized that a conflict can also arise between a plausible unit and the implausible sentence interpretation resulting from the parser. This hypothesis was tested by creating sentences that either contained a plausible or implausible unit. For example:

(8) Jan zag dat de boeren de eieren legden/maalden … (plausible/implausible unit)
John saw that the farmers the eggs laid/crushed … (literal translation)
John saw that the farmers laid/crushed the eggs … (paraphrase)

Indeed, the sentences containing a plausible unit elicited a monophasic P600, while the sentences containing an implausible unit elicited an N400 effect and a greatly reduced P600. In general, this study shows that implausibility in itself is not necessary to elicit a P600; a plausible unit in an implausible sentence can also trigger a conflict and elicit a P600.

The studies described so far are all instances in which the conflict is assumed to be at the sentence level. To investigate whether competing representations at other linguistic levels would also trigger a conflict and elicit a P600 effect, two further experiments were conducted. In the first one, Vissers et al. (2006) created high- and low-cloze sentences in which they embedded pseudohomophones (i.e. misspellings of a word but with similar phonology):

(9) De kussens zijn opgevuld met verun … (high-cloze)
The pillows are stuffed with feathurs … (literal translation)

(10) Haar walkman deed het niet meer vanwege de verun … (low-cloze)
Her walkman did not work anymore because of the feathurs … (literal translation)

It was hypothesized that, only for the high-cloze sentences, there would be a strong inclination to accept the pseudohomophone because of its similar phonology and orthography with respect to the expected word (veren–feathers). However, there would also be a strong inclination to reject it because it was misspelled, hereby creating a conflict at the word level and triggering the P600 effect. When comparing these sentences to their correct counterparts, the results indeed revealed a P600 effect only for the high-cloze sentences.[1,2]

In a second experiment, a conflict was created at the conceptual level (Vissers et al. 2008). Participants were shown pictures depicting locative relationships, followed by a sentence that could describe the picture correctly or incorrectly, depending on the preposition that was used. Two types of mismatches were created; intra- and extra-dimensional mismatches. For the intra-dimensional mismatches, the opposite preposition from the same dimension (i.e. horizontal/vertical) compared with the correct preposition was used (e.g. *in front of* instead of *behind*):[3]

(11) □○ De cirkel staat voor het vierkant. (intra-dimensional mismatch)
The circle is in front of the square. (literal translation)

For the extra-dimensional mismatches, a preposition from the other dimension was used (e.g. *above* instead of *behind*):

(12) □○ De cirkel staat boven het vierkant. (extra-dimensional mismatch)
The circle is above the square. (literal translation)

The extra-dimensional mismatches were added as a pure semantic violation to assure that, if a P600 effect would be found, this could not be explained by role reassignment. The intra-dimensional mismatches could in principle be repaired by switching the roles of the NPs. However, for the extra-dimensional mismatches, this would not result in the correct sentence. The authors hypothesized that, in case of a picture–sentence mismatch, a conflict at the conceptual level would be present between the picture representation and the sentence representation. Both the intra- and extra-dimensional mismatches elicited a P600 effect, indicating that besides conflicts at the sentence and word level, conflicts at the conceptual level can trigger reprocessing as well. In addition, the fact that both mismatches elicited a P600 effect ruled out that thematic role reassignment is critical for triggering the P600 effect.

Until now, we have discussed various studies that compared sentences in which conflicts were either present or not. However, in daily conversation and while reading, we often encounter information that is a bit unexpected. This information, however, can be new and important, and should be integrated, otherwise learning is prohibited. On the other hand, we sometimes come across information that is highly unexpected and impossible to integrate in the current context. In this case, to prevent the integration of false information, it would be useful to mistrust what we heard or read and check for possible processing errors. These examples suggest that a gradation in the seriousness of a conflict exists. Therefore, in a recent study (Van de Meerendonk et al. in press), the strength of the conflict was varied by manipulating the plausibility of the sentences (plausible, mildly implausible, deeply implausible). Differences in the degree of the expectancy violations were created, while keeping the expectancy based on the sentence context the same. An example:

(13) Lichaamsdelen zoals een arm, nek en teen … (plausible)
Parts of the body like an arm, neck and toe … (literal translation)

(14) Lichaamsdelen zoals een arm, nek en haar … (mildly implausible)
Parts of the body like an arm, neck and hair … (literal translation)

(15) Lichaamsdelen zoals een arm, nek en telescoop … (deeply implausible)
Parts of the body like an arm, neck and telescope … (literal translation)

It was predicted that only the deeply implausible sentences would create a strong enough conflict to trigger reanalysis and therefore elicit a P600 effect. For the mildly implausible sentences, the conflict would not be strong enough and integration difficulties should be resolved successfully. The results confirmed these predictions. The mildly implausible sentences elicited an N400 effect when compared with the plausible sentences, while the deeply implausible sentences showed a biphasic N400–P600 pattern. This study showed that a strong violation of expectancy is needed to create a conflict that is powerful enough to elicit a reanalysis. The existence of a threshold that has to be passed for monitoring to occur serves a clear function: efficiency. If mild conflicts would also elicit a monitoring response, we would be constantly doubting ourselves ('Did I read/hear that correctly?'), and the integration of new information would be prohibited.

To summarize, a commonality for these different studies is that, in all of them, a certain linguistic event is highly expected but another event is encountered. According to the monitoring hypothesis of language perception, a conflict arises between these representations and a reanalysis is triggered to check the input for processing errors. The P600 amplitude modulation is proposed to reflect this reanalysis. This account could also explain the P600 effects found to syntactic violations and ambiguous sentences, in which a conflict between expectancies arises as well. As to syntactic violations, speech errors are a rare event – less than 1% of our utterances – so the expectation for a sentence to contain a grammatical morpheme is very strong. Furthermore, for more complex sentences, because they are more difficult, the chance of conflicting representations is higher and this may give rise to the P600 effect (Van Herten et al. 2006). It is hypothesized that, in the reanalysis, all aspects of the input are taken into account: the semantic, syntactic, orthographic and phonological aspects. We hypothesize that, as a result of the reanalysis, it becomes clear that no processing error occurred and the perceived error was indeed present.[4] In addition, we assume that depending on the type of error, the reanalysis can be focused on certain aspects of the stimulus (e.g. phonological/orthographic aspects for misspellings (Vissers et al. 2006)). This might explain the fact that in the misspelling study (Vissers et al. 2006) and the picture–sentence mismatch study (Vissers et al. 2008), instead of the generally reported central–posterior scalp distribution, the P600 effect extended to some anterior sites. The type of information that is the focus of the reanalysis could give rise to variations in scalp distribution. However, to be sure, future studies should directly compare the different kinds of violations within the same group of subjects.

Monophasic Vs. Biphasic Patterns

According to the monitoring hypothesis, a strong conflict should elicit a P600 effect. However, some of the studies that were described reported a monophasic P600 effect (e.g. Kolk et al. 2003; Kuperberg et al. 2003, 2007; Hoeks et al. 2004; Kim and Osterhout 2005; Van Herten et al. 2005, 2006; Vissers et al. 2006), while others reported a biphasic N400–P600 pattern (e.g. Van de Meerendonk et al. in press; Vissers et al. 2008). This apparent discrepancy in results could be explained by the fact that in most of the studies that reported a monophasic P600 effect, the individual words could be combined to form a meaningful sentence. However, for the deeply implausible sentences in the Van de Meerendonk et al. (in press) study, this was not possible, causing integration difficulties and eliciting an N400 effect when comparing these sentences to their plausible counterparts. In addition, in the picture–sentence mismatch study (Vissers et al. 2008), an early negativity (between 200 and 400 ms) preceded the P600 effect to the mismatching conditions. It was proposed that this early ERP effect could reflect an early N400 effect elicited by a strong semantic expectancy created by the picture.

Furthermore, contrary to what the monitoring hypothesis proposes, studies including 'strong' semantic violations have all found an N400 effect, but the results are inconsistent regarding the P600 effect. Let us take the following sentence from Kutas and Hillyard (1980c):

(16) He spread the warm bread with socks.

It would seem that, in this sentence, a strong conflict arises between the word 'butter' and 'socks', but no P600 effect was reported. Various studies that contained strong semantic violations, however, did report a positivity following the N400 effect. For example, Ford et al. (1996) and Woodward et al. (1993) presented a subset of the semantically congruous and incongruous sentences that were used in various studies by Kutas and Hillyard (1980a–c). They found that the N400 to incongruous sentence completions was accompanied by a late positivity. Gunter et al. (1992) and Swick et al. (1998) found a positivity following the N400 effect to incongruous sentence endings as well.

Inconsistencies regarding the presence or absence of a concomitant P600 effect to semantic violations have also been shown to critical words in intermediate positions. Kolk et al. (2003) found that in the presence of a plausibility judgment task, a P600 effect followed the N400 effect to strong semantic violations. This effect disappeared when no judgment was required. However, in a later study by Van Herten et al. (2005), the same stimuli were found to elicit a biphasic pattern, while no judgment was asked.

Van Petten and Luka (2006) noted that little research has been carried out to try and determine what factors influence when a monophasic N400 effect or a biphasic pattern occurs. Van de Meerendonk et al. (in press) obtained a biphasic N400–P600 pattern to critical words in intermediate sentence positions. This study showed that plausibility could be one of the factors to discriminate the different patterns. However, the sentence types and/or individual processing strategies, for example, could play a role as well. Another factor that could influence the results is overlap between the N400 and P600 component (see e.g. Schwartz et al. 1996), though this is hard (if not impossible) to prove. It is clear that further research is needed to clarify the commonalities and differences between the monophasic N400, biphasic N400–P600 and monophasic P600 patterns.

Functional Significance of the P600 Component

As discussed in this review, the view that the P600 effect is elicited exclusively by syntactic violations is no longer legitimate. As the presented ERP data show, in addition to syntactic violations, several semantic violations, orthographic violations and picture–sentence mismatches have been found to elicit a P600 effect as well. In addition, earlier studies have reported P600 effects outside the language domain. For example, Patel et al. (1998) found a positivity to musical violations (out-of-key chords), which was similar in scalp distribution, latency and amplitude to the P600 effect elicited by syntactic violations. Furthermore, violations of non-linguistic abstract structures (e.g. ABCBAC and DEFEDF have a different serial structure but they share the same abstract structure 123213;

GHIGHI would violate this structure) have also been found to elicit a positivity similar to the P600 effect (Lelekov-Boissard and Dominey 2002). In addition, Núñez-Peña and Honrubia-Serrano (2004) showed that violations of arithmetic rules (e.g. 4-7-10-13-16-19-**23**) elicit a P600 effect. From these studies, it is concluded by Núñez-Peña and Honrubia-Serrano (2004) that the P600 does not necessarily reflect the violation of a linguistic rule, but that it is a more general index of violations in rule-governed sequences.

How can these various findings be accounted for? One proposal is the account of Kuperberg (2007). She proposes a linguistic explanation of the P600 effect evoked by semantic verb–argument violations. As mentioned in the beginning, Kuperberg et al. (2003) initially proposed that reassignment of thematic roles might underlie the P600 effect. However, when dividing the sentences into whether they would lend themselves to such a reassignment (determined by a plausibility judgment task on passivized versions of the sentences), Kuperberg et al. (2006) found that thematic role reassignment could not be the only trigger for the P600 effect, as the sentences that were not repairable also elicited a P600 effect (see also Kuperberg et al. 2007). In her recent review, Kuperberg (2007) described different factors that can influence the P600 effect to semantic violations. These factors are: the semantic association between verbs and arguments, verb-based restrictions, animacy, implausibility of the final representation, context and the experimental task. She concludes that though all these factors can influence the P600, they are not all necessary at the same time to invoke an amplitude modulation. Kuperberg (2007) proposes that there are at least two neural routes toward language comprehension. The first stream is based primarily on semantic memory, and influences the N400 component. To make predictions about upcoming words, the evolving representation of meaning (the context) is compared with patterns of relationships that are prestored within semantic memory. The second stream is a combinatorial stream that integrates an incoming word with the context, assigning thematic roles based on multiple rule-like constraints (including morphosyntactic and semantic–thematic constraints), to build up a propositional meaning. This updated context is then again compared with information within semantic memory, determining whether it is (im)plausible, and a new cycle begins (Kuperberg 2009). The P600 effect is thought to reflect a continued (re)analysis of the combinatorial stream triggered by a conflict between the outputs of the two processing streams. The outputs conflict when, for example, the semantic memory-based stream comes up with a possible prediction, but the combinatorial stream yields an anomalous or very implausible interpretation.

The focus of Kuperberg's account is on semantic verb–argument violations, but she leaves open the question of how P600 effects to other violations (e.g. orthographic violations) can be accounted for. A possibility could be that, as Núñez-Peña and Honrubia-Serrano (2004) proposed, the P600 reflects a general index of rule violations. Sitnikova et al. (2008) reported a late positivity to unexpected action sequences in videoclips (e.g. someone who tries to shave with a rolling pin). This positivity was proposed to reflect a similar process as the continued analysis of the combinatorial stream in the language domain. The unexpected action sequences violated implicit rules, and to make sense of the event people tried to relate the central action to the objects and people around it.

A second account that could explain the various findings is the monitoring hypothesis discussed in this review. The monitoring account is similar to Kuperberg's account in that both assume that a *conflict* between representations (of which one is based on predictions) triggers a process of continued (re)analysis, reflected by the P600 component. However, a difference between the accounts is that instead of giving a linguistic explanation, in the monitoring account the P600 effects are explained in terms of error monitoring, which is an aspect of executive control. The monitoring hypothesis is therefore able to account for other kinds of violations, besides verb–argument violations, that have been shown to elicit a P600 effect. We think that a strong violation of expectancy and not a rule violation as such is the trigger for the P600 effect. Rules, however, can be imposed on stimuli and create certain expectations. Therefore, when for example grammatical or arithmetic rules are violated, there is also a strong expectancy violation.

Conclusion

Originally the P600 effect was found to be elicited by syntactic violations. However, recently the effect has been found to several other linguistic violations, and in addition a similar positive effect has been reported outside the language domain. Therefore, the empirical generalization seems to be that the P600 effect is not driven by syntactic processes, but due to a strong expectancy violation at different levels of cognitive processing. The monitoring theory of language perception is consistent with this generalization and proposes that strong violations can trigger a conflict, resulting in a reanalysis of the input to check for processing errors, reflected by the P600 component. This indicates that, contrary to what is commonly assumed, language perception is not purely an automatic process, but is in need of executive control. This process of control is not always active but only becomes so when bottom-up information signals the presence of a conflict.

Short Biographies

Nan van de Meerendonk is a PhD student at the Donders Institute for Brain, Cognition and Behaviour at the Radboud University Nijmegen, The Netherlands. She holds a BA and MA in Science from the Radboud University Nijmegen. Her research is concerned with error monitoring in language perception, for which she uses different neuroimaging methods like EEG measurements and fMRI. Currently, a paper of her is in press in the *Journal of Cognitive Neuroscience*.

Herman Kolk is a professor of neuropsychology at the Radboud University in Nijmegen, the Netherlands. He published on language disorders like aphasia, dyslexia and stuttering. The red line in all this research is how persons with a language problem control their language behavior to circumvent the cognitive consequences of their problem. His research with unimpaired language users focuses on monitoring: how persons detect and repair errors, both in production and in perception.

Dorothee Chwilla is a cognitive neuroscientist in the field of language at the Donders Institute for Brain, Cognition and Behaviour. She investigates semantic, syntactic and prosodic processes in language comprehension across contexts (single words, sentences and discourse). A central theme in her work is how and when different kinds of knowledge (associative relations, semantic relations, scripts) are accessed and integrated into context. Recent research areas are the processing of semantics in a second language and the interaction of language and attention.

Constance Vissers holds a PhD in Social Sciences from the Radboud University Nijmegen. Her dissertation deals with Monitoring in Language Perception and includes several EEG experiments. She has authored and co-authored papers in the areas of executive control, language perception and language comprehension; these papers were published in *Biological Psychology*, *Neuropsychologia* and *Journal of Cognitive Neuroscience and Cognitive Brain Research*. She currently works at the University Medical Hospital St Radboud at the Department of Psychiatry as a clinical psychologist and as a post-doc researcher. Her current research deals with the cognitive neuroscience of language processing in psychiatric disorders (depression and autism spectrum disorder) and healthy subjects.

Notes

* Correspondence address: Nan van de Meerendonk, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: n.vandemeerendonk@donders.ru.nl.

[1] The low-cloze pseudohomophones seemed to elicit a reversed P600 pattern. However, closer inspection suggested that this was due to a biphasic N400–P600 pattern in the low-cloze correct word condition (see Vissers et al. 2006).

[2] Gunter et al. (2000) also found that P600 amplitude could be modulated by semantic expectancy. They reported a P600 effect to article–noun gender agreement violations for high-cloze nouns but not for low-cloze nouns.

[3] Before the start of the experiment, subjects were shown three example pictures, and the correct descriptions in terms of 'in front of'–'behind' and 'above'–'below' were explained.

[4] The various studies discussed in this review are related to the detection and recovery from language errors. However, as indicated in the beginning, adjustments are also important to prevent future errors. Experiments that show these adjustments in the language domain have yet to be performed.

Works Cited

Botvinick, Matthew M., Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen. 2001. Conflict monitoring and cognitive control. Psychological Review 108.624–52.

Chwilla, Dorothee J., Peter Hagoort, and Colin M. Brown. 1998. The mechanism underlying backward priming in a lexical decision task: spreading activation versus semantic matching. Quarterly Journal of Experimental Psychology 51A.531–60.

Coles, Michael G. H., Marten K. Scheffers, and Clay B. Holroyd. 2001. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing Biological Psychology 56.173–89.

Cutler, Anne, and Sally Butterfield. 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. Journal of Memory and Language 31.218–36.

Eriksen, Barbara A., and Charles W. Eriksen. 1974. Effects of noise letters upon identification of a target letter in a nonsearch task. Perception & Psychophysics 16.143–9.

Federmeier, Kara, and Marta Kutas. 1999. A rose by any other name: long-term memory structure and sentence processing. Journal of Memory and Language 41.469–95.

Ford, Judith M., Steven H. Woodward, Edith V. Sullivan, Brandy G. Isaacks, Jared R. Tinklenberg, Jerome A. Yesavage, and Walton T. Roth. 1996. N400 evidence of abnormal responses to speech in Alzheimer's disease. Electroencephalography and Clinical Neurophysiology 99.235–46.

Friederici, Angela D., Erdmut Pfeifer, and Anja Hahne. 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cognitive Brain Research 1.183–92.

Gunter, Thomas C., Angela D. Friederici, and Herbert Schriefers. 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. Journal of Cognitive Neuroscience 12.556–68.

——, Janet L. Jackson, and Gijsbertus Mulder. 1992. An electrophysiological study of semantic processing in young and middle-aged academics. Psychophysiology 29.38–54.

Hagoort, Peter, Colin Brown, and Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes 8.439–83.

Hoeks, John C.J., Laurie A. Stowe, and Gina Doedens. 2004. Seeing words in context: the interaction of lexical and sentence level information during reading. Cognitive Brain Research 19.59–73.

Holcomb, Phillip J. 1993. Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. Psychophysiology 30.47–61.

Holroyd, Clay B., and Michael G. H. Coles. 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychological Review 109.679–709.

Kaan, Edith, Anthony Harris, Edward Gibson, and Phillip Holcomb. 2000. The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes 15.159–201.

Kaufman, Renée C., and Loraine K. Obler. 1995. Classification of normal reading error types. Developmental and acquired dyslexia: neuropsychological and neurolinguistic perspectives, ed. by Che Kan Leong and R. Malatesha Joshi, 149–57. Dordrecht: Kluwer Academic Publishers.

Kim, Albert, and Lee Osterhout. 2005. The independence of combinatory semantic processing: evidence from event-related potentials. Journal of Memory and Language 52.205–25.
Kolk, Herman H. J., Dorothee J. Chwilla, Marieke Van Herten, and Patrick J. W. Oor. 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. Brain and Language 85.1–36.
Kuperberg, Gina R. 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Research 1146.23–49.
——. 2009. Understanding events: streams of language processing in the brain. Talk given at the Donders Institute for Brain, Cognition and Behaviour. The Netherlands: Nijmegen.
——, David Caplan, Tatiana Sitnikova, Marianna Eddy, and Phillip J. Holcomb. 2006. Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. Language and Cognitive Processes 21.489–530.
——, Donna A. Kreher, Tatiana Sitnikova, David N. Caplan, and Phillip J. Holcomb. 2007. The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. Brain and Language 10.223–37.
——, Tatiana Sitnikova, David Caplan, and Phillip J. Holcomb. 2003. Electrophysiological distinctions in processing conceptual relationships within simple sentences. Cognitive Brain Research 17.117–29.
Kutas, Marta, and Kara D. Federmeier. 2000. electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Neuroscience 4.463–70.
——, and Steven Hillyard. 1980a. Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychology 11.99–116.
——, and ——. 1980b. Reading between the lines: event-related potentials during natural sentence processing. Brain and Language 11.354–73.
——, and Steven A. Hillyard. 1980c. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207.203–5.
——, and ——. 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307.161–3.
——, and Cyma K. Van Petten. 1994. Psycholinguistics electrified: event-related brain potential investigations. Handbook of psycholinguistics, ed. by Morton Ann Gernsbacher, 83–143. San Diego, CA: Academic Press.
Lau, Ellen F., Colin Phillips, and David Poeppel. 2008. A cortical network for semantics: (de) constructing the N400. Nature Reviews Neuroscience 9.920–33.
Lelekov-Boissard, Taïssia, and Peter F. Dominey. 2002. Human brain potentials reveal similar processing of non-linguistic abstract structure and linguistic syntactic structure. Clinical Neurophysiology 32.72–84.
Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. Cognition 14.41–104.
——. 1989. Speaking: form intention to articulation. Cambridge, MA: The MIT Press.
MacKay, Donald G. 1987. The organization of perception and action: a theory for language and other cognitive skills. New York: Springer.
——. 1990. Perception, action, and awareness: a three-body problem. Relationships between perception and action: current approaches, ed. by Odmar Neumann and Wolfgang Prinz, 269–303. Berlin: Springer-Verlag.
——. 1992. Awareness and error detection: new theories and research paradigms. Consciousness and Cognition 1.199–225.
Münte, Thomas F., Hans-Jochen Heinze, Mike Matzke, Bernardina M. Wieringa, and Sönke Johannes. 1998. Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. Neuropsychology 36.217–26.
Núñez-Peña, M. Isabel, and M. Luisa Honrubia-Serrano. 2004. P600 related to rule violation in an arithmetic task. Cognitive Brain Research 18.130–41.
Osterhout, Lee, and Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language 31.785–806.
Patel, Aniruddh D., Edward Gibson, Jennifer Ratner, Mireille Besson, and Phillip J. Holcomb. 1998. Processing syntactic relations in language and music: an event-related potential study. Journal of Cognitive Neuroscience 10.717–33.
Postma, Albert. 2000. Detection of errors during speech production: a review of speech monitoring models. Cognition 77.97–131.
Schwartz, Tanya J., Marta Kutas, Nelson Butters, Jane S. Paulsen, and David P. Salmon. 1996. Electrophysiological insights into the nature of the semantic deficit in Alzheimer's disease. Neuropsychologia 34.827–41.
Sitnikova, Tatiana, Phillip J. Holcomb, Kristi A. Kiyonaga, and Gina R. Kuperberg. 2008. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. Journal of Cognitive Neuroscience 20.2037–57.
Swick, Diane, Marta Kutas, and Robert T. Knight. 1998. Prefrontal lesions eliminate the LPC but do not affect the N400 during sentence reading. Journal of Cognitive Neuroscience Supplement 5.29.
Van Berkum, Jos J. A., Peter Hagoort, and Colin M. Brown. 1999. Semantic integration in sentences and discourse: evidence from the N400. Journal of Cognitive Neuroscience 11.657–71.
Van de Meerendonk, Nan, Herman H. J. Kolk, Constance Th. W. M. Vissers, and Dorothee J. Chwilla. in press. Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. Journal of Cognitive Neuroscience.

Van Herten, Marieke, Dorothee J. Chwilla, and Herman H. J. Kolk. 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. Journal of Cognitive Neuroscience 18.1181–97.

——, Herman H. J. Kolk, and Dorothee J. Chwilla. 2005. An ERP study of P600 effects elicited by semantic anomalies. Cognitive Brain Research 22.241–55.

Van Petten, Cyma, and Barbara J. Luka. 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language 97.279–93.

Vissers, Constance Th. W. M., Dorothee J. Chwilla, and Herman H. J. Kolk. 2006. Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. Brain Research 1106.150–63.

——, ——, and ——. 2007. The interplay of heuristics and parsing routines in sentence comprehension: evidence from ERPs and reaction times. Biological Psychology 75.8–18.

——, ——, ——, and ——. 2008. Monitoring in language perception: evidence from ERPs in a picture-sentence matching task. Neuropsychologia 46.967–82.

Woodward, Steven H., Judith M. Ford, and Susan C. Hammett. 1993. N4 to spoken sentences in young and older subjects. Electroencephalography and Clinical Neurophysiology 87.306–20.

Yeung, Nick, Matthew M. Botvinick, and Jonathan D. Cohen. 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychological Review 111.931–59.