

Reasoning with Exceptions: An Event-related Brain Potentials Study

Judith Pijnacker¹, Bart Geurts¹, Michiel van Lambalgen²,
Jan Buitelaar^{1,3,4}, and Peter Hagoort^{1,5}

Abstract

■ Defeasible inferences are inferences that can be revised in the light of new information. Although defeasible inferences are pervasive in everyday communication, little is known about how and when they are processed by the brain. This study examined the electrophysiological signature of defeasible reasoning using a modified version of the suppression task. Participants were presented with conditional inferences (of the type “if p, then q; p, therefore q”) that were preceded by a congruent or a disabling

context. The disabling context contained a possible exception or precondition that prevented people from drawing the conclusion. Acceptability of the conclusion was indeed lower in the disabling condition compared to the congruent condition. Further, we found a large sustained negativity at the conclusion of the disabling condition relative to the congruent condition, which started around 250 msec and was persistent throughout the entire epoch. Possible accounts for the observed effect are discussed. ■

INTRODUCTION

In everyday communication, the meaning of an utterance usually goes beyond its explicit linguistic meaning. For example, take the following situation. I have invited my friend for dinner at 7 o'clock. At 6:45 my telephone rings. It is my friend and he says “I have a flat tire.” To understand his utterance, there is a lot more required than just combining word meanings and syntactic structure. Background knowledge and inferences are involved too. As I know that my friend does not have a car, I will infer that it is my friend's bike that has a flat tire. Moreover, I might infer that if he has a flat tire, he will arrive later. Further, I reason that if he is late, it will be better to turn off the oven, otherwise the meal might burn.

Although such inferences are pervasive in everyday communication, little is known about how and when they are processed by the brain. This study examines the electrophysiological signature of everyday inferences. The focus is on one particular form of inference, namely, conditional inferences, which always have “If p then q” as one of their premises. One classical inference associated with conditionals is modus ponens, in which the

second premise (1b) confirms the antecedent of the conditional premise (1a):

- (1a) If Mary has an exam, she will study in the library.
- (1b) Mary has an exam.
- (1c) Mary will study in the library.

Most people accept this inference (for a review, see Evans, Newstead, & Byrne, 1993). According to classical logic, modus ponens is considered valid, which means that the conclusion necessarily follows from the premises. This makes classical logic monotonic: Additional information cannot render an inference invalid if it was valid previously. However, many studies (e.g., Pijnacker et al., 2009; Bonnefon & Hilton, 2002; Politzer & Bourmaud, 2002; Dieussaert, Schaeken, Schroyens, & d'Ydewalle, 2000; Byrne, Espino, & Santamaria, 1999; Chan & Chua, 1994; Cummins, Lubart, Alksnis, & Rist, 1991; Byrne, 1989) have found that modus ponens inferences can be suppressed in the light of extra information:

- (2a) If Mary has an exam, she will study in the library.
- (2b) If the library is open, Mary will study in the library.
- (2c) Mary has an exam.
- (2d) Mary will study in the library.

Byrne et al. (1999) presented this reasoning problem with and without premise (2b). As soon as the extra premise (2b) came in, the number of people concluding that Mary will study in the library dropped to about 50%, whereas without the extra premise most people accepted the conclusion. Thus, the addition of an extra premise such as (2b) leads to a significant decrease of the rate at which a modus ponens inference is accepted. This has

¹Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands, ²Institute for Logic, Language and Computation, Amsterdam, the Netherlands, ³Radboud University Nijmegen Medical Centre, the Netherlands, ⁴Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, the Netherlands, ⁵Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

been called the suppression effect (Byrne et al., 1999; Byrne, 1989, 1991). Example (2) clearly illustrates that conditional reasoning is not monotonic, but defeasible: New information can cause people to retract an inference. Everyday inferences are usually defeasible, too. In the above-mentioned example, I reasoned that if my friend has a flat tire, he will arrive later, but I will withdraw this inference if he tells me that he took a bus.

Although there is an extensive literature on the suppression effect, it is still unclear how defeasible inferences are processed in the brain. ERPs have been proven to be a useful tool for investigating the time course of information processing because they have a good temporal resolution and can provide on-line information about cognitive processes. As far as we know, there is just one study that examined conditional reasoning using ERPs: Qiu et al. (2007) found that conditional inferences elicited a larger negativity than a baseline task between 500 and 700 msec, and between 1700 and 2000 msec after onset of the second premise. However, it is not clear how much significance we should attach to these findings, as the premises were presented in their entirety, and therefore, this study does not give us a precise idea of the time course of conditional reasoning.

Perhaps more informative are the studies that have used ERPs for investigating the processing of linguistic information, as integration of linguistic information is an essential part of reasoning. One ERP component that is related to linguistic processing is the N400 effect, which is a negative shift that has a peak around 400 msec after the critical word, and has a centro-parietal scalp distribution. The N400 effect was first found for sentences ending with a semantically inappropriate word, like “He spread his warm bread with *socks*”¹ in contrast to “He spread his warm bread with *butter*” (Kutas & Hillyard, 1980). Later, it was found that the N400 effect also occurs in sentences that are semantically appropriate but where words conflict with (i) expectancy, such as *pocket* in “Jenny put the sweet in her *mouth/pocket* after the lesson” (Hagoort & Brown, 1994); (ii) world knowledge, such as *white* in “Dutch trains are *yellow/white* and very crowded”² (Hagoort, Hald, Bastiaansen, & Petersson, 2004); and (iii) discourse context, such as *slow* in “As agreed upon, Jane was to wake her sister and her brother at five o’clock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally *quick/slow*” (Van Berkum, Zwitserlood, Hagoort, & Brown, 2003; Van Berkum, Hagoort, & Brown, 1999). Furthermore, St. George, Mannes, and Hoffman (1997) showed that implicit information, such as bridging inferences, affects the N400 amplitude as well. In general, the N400 effect is seen as an index of processes involved in the integration of the meaning of a word into a representation of its preceding context. As integration of a word into the context becomes harder because it does not satisfy semantic expectations, the amplitude of the N400 increases (Van Berkum, Hagoort, & Brown, 1999; Brown & Hagoort, 1993).

Another negative ERP component that has been found in linguistic tasks is the sustained negativity. This negative shift occurs in a similar latency window as the N400 effect, but is more sustained, does not have a clear peak, and usually has a more anterior topography. Sustained negativities have been found in a wide variety of linguistic phenomena, although it is less clear how this ERP component should be interpreted. For example, Van Berkum, Brown, et al. (1999, 2003) found a sustained anterior negativity when a noun’s referent is temporarily ambiguous, such as *girl* in “David had told the two girls to clean up their room before lunchtime. But one of girls had stayed in bed all morning, and the other had been on the phone all the time. David told *the girl* that ...,” in contrast to a referentially unambiguous noun. Van Berkum, Brown, et al. (1999, 2003) suggested that the negative shift is due to an extra load on working memory, because two possible candidates for a referent must be maintained. Also, object-relative sentences such as “The reporter who the senator attacked admitted the error” elicited a sustained negativity compared to subject-relative sentences such as “The reporter who attacked the senator admitted the error” (Müller, King, & Kutas, 1997; King & Kutas, 1995), which may be due to greater working memory load, or perhaps to additional processing, because in object-relative sentences, the head noun phrase (“the reporter”) has two different grammatical roles. Münte, Schiltz, and Kutas (1998) found a sustained anterior negativity for sentences that present events out of chronological order, such as “Before the psychologist submitted the article, the journal changed its policy” relative to “After the psychologist submitted the article, the journal changed its policy.” They attributed this effect to additional discourse-level computation. Furthermore, Ye and Zhou (2008) demonstrated that participants with high cognitive control showed a sustained anterior negativity for implausible active sentences such as “The thief kept the policeman in the police station” versus its plausible counterpart “The policeman kept the thief in the police station.” They suggested that the sustained negativity may reflect inhibitory processes to suppress a representation that is in conflict with world knowledge. Finally, Baggio, van Lambalgen, and Hagoort (2008) found a sustained anterior negativity in defeasible inferences. Participants read sentences such as “The girl was writing a letter when her friend spilled coffee on the *paper/tablecloth*.” In the case of spilling on the tablecloth, it can be inferred that the girl wrote a letter, whereas in the case of spilling on the paper, the inference that the girl wrote a letter is suppressed, indicating that the inference is defeasible.

Taken together, although these linguistic phenomena are quite heterogeneous, what they appear to have in common is that there are two possible representations competing with each other. There are two possible referents for the ambiguous noun in Van Berkum, Brown, et al. (1999, 2003); in object-relative sentences, the head noun phrase has both a subject and object role; in the study of Münte et al. (1998), there may have been a default representation

of events in chronological order as well as the actual, non-chronological order of events; and finally, in Baggio et al. (2008), there is a default representation that “writing a letter” has a completed letter as goal state, whereas the “spilling coffee event” requires one to revise this implication. In sum, the sustained negativity may reflect recomputation or additional processing in order to come up with a meaningful representation, or extra working memory load due to multiple representations.

The Present Study

This study was designed to investigate the electrophysiological signature of defeasible reasoning by using a modified version of the suppression task (Byrne, 1989). Participants read modus ponens inferences preceded either by a congruent context or a disabling context (see Table 1 in Materials: ERP Experiment section). The disabling context contained a possible exception with regard to the conditional, and was introduced to elicit suppression of modus ponens.

We argue that there is a default representation that entails the modus ponens inference. However, this representation becomes problematic when the inference is preceded by a disabling context, which causes people to consider revising the inference. This is in line with the framework by Stenning and van Lambalgen (2005, 2008), who argued that conditionals may contain a marker for exceptions [for instance, (2a) could be interpreted as *If Mary has an exam and nothing abnormal is the case, then she will study in the library*]. In the congruent context, people can apply so-called closed-world reasoning to exceptions, that is, exceptions are considered to be not the case, as long as there is no evidence for any exceptions (i.e., the default representation). However, in the disabling context, the original closed-world assumption cannot be maintained anymore because a possible exception has now become salient, namely, that the library may be closed. This prevents people from drawing the conclusion that Mary will study in the library (2d). Hence, modus ponens is suppressed (for a detailed de-

scription of this framework, see Stenning and van Lambalgen, 2005, 2008).

For the present experiment, we hypothesized that there are two possibilities. First, the disabling condition could elicit a discourse-induced semantic N400 effect at the conclusion relative to the congruent condition, due to a difficulty with integrating the conclusion into the preceding discourse as expectations are not fulfilled. Second, the disabling condition could elicit a sustained anterior negativity at the conclusion associated with two competing representations: the default representation and the revised one incorporating the possible exception.

The aim of this study is to investigate how and when the brain integrates information about exceptions that are relevant to arrive at a conclusion. Although we think it is most likely that the ERP effect will occur at the final word of the conclusion, because it is only at this position that it becomes clear that the conclusion clashes with the preceding information, we will also take the final word of the first and second premise into consideration. To check whether the ERP effect has the signature of a standard N400 effect, we added two control conditions that consisted of sentences that ended with a word that was either semantically congruent or incongruent. Moreover, a Reading Span Test was included as an index of verbal working memory performance to investigate its role in suppression.

METHODS

Participants

Twenty right-handed participants took part in the study. All participants were native speakers of Dutch, had no language disorders, had no known neurological history, and had normal or corrected-to-normal vision. Participants were recruited from the Donders Institute subject pool. Two participants were excluded from analysis due to an excessive number of artifacts in the EEG signal. The remaining 18 participants (9 men) were aged between 19 and 32 years, and their mean age was 23 years. All participants signed informed consent and received reimbursement or

Table 1. Experimental Conditions

<i>Condition</i>	<i>Sentence Type</i>	<i>Example</i>
MP-disabling	Disabling context	Lisa probably lost a contact lens.
	Premise 1	If Lisa is going to play hockey, then she will wear contact lenses.
	Premise 2	Lisa is going to play hockey.
	Conclusion	Lisa will wear contact lenses.
MP-congruent	Congruent context	Lisa has recently bought contact lenses.
	Premise 1	If Lisa is going to play hockey, then she will wear contact lenses.
	Premise 2	Lisa is going to play hockey.
	Conclusion	Lisa will wear contact lenses.

MP = modus ponens.

course credits for participation. The study was approved by the local ethics committee.

Reading Span Test

We used a computerized Dutch version of the Reading Span Test to measure verbal working memory (for a detailed description, see Van den Noort, Bosch, Haverkort, & Hugdahl, 2008; Van den Noort, Bosch, & Hugdahl, 2006). Participants had to read aloud 100 sentences, which appeared on a computer screen. Sentences were presented in different set sizes of 2, 3, 4, 5, or 6 sentences in random order. When participants had finished a sentence, they pressed the space bar for the next sentence to appear. If participants could not finish the sentence within 6.5 sec, then the next sentence showed up automatically. After completion of a set of sentences, the word “recall” occurred on the screen. At that point, participants had to recall the final word of each sentence in the set (in free order). Participants were instructed to read the sentences aloud at normal speed, and to remember the final word of each sentence. Reading span was determined as the total number of correctly recalled words.

Materials: ERP Experiment

We created 80 reasoning problems in Dutch. All reasoning problems had the inference form of modus ponens (if p then q; p, therefore q), and were preceded by a congruent context or a disabling context (Table 1).

The disabling context contained a possible exception or precondition with regard to the conditional. Congruent contexts and disabling contexts were kept as similar as possible with regard to syntactic structure and sentence

length. There were no significant differences in sentence length between the congruent and disabling contexts ($p > .1$). Final words of the sentences were never longer than 12 letters to avoid eye movements and average final word length was 6.7 letters.

In addition to the 80 experimental reasoning problems, 80 filler reasoning problems were used, which included 40 modus ponens inferences with an incongruent conclusion, 20 affirmation of the consequent inferences with a congruent conclusion, and 20 with an incongruent conclusion (see Table 2 for examples). All fillers were preceded by a congruent context. Fillers were included to reduce the predictability of the materials and to balance for response types (i.e., to evoke “maybe” and “no” responses). In total, each participant read 160 reasoning problems: 40 reasoning problems in a disabling context, 40 reasoning problems in a congruent context, and 80 fillers.

The two versions of the reasoning problems were counter-balanced across two lists. Thus, no participant saw the same reasoning problem more than once. Each list was presented to an equal number of participants. Moreover, four versions were created in which the order of the items was reversed according to a Latin square design. Thus, no reasoning problem always occurred at the same position in the experiment. Finally, reasoning problems were presented in pseudo-random order with the constraint that the same condition never occurred more than twice in a row.

In addition, we included two control conditions in which semantic congruency was manipulated in order to elicit a standard N400 effect. These additional two conditions contained 80 Dutch sentences that ended with a word that was either semantically congruent (40 items) or incongruent (40 items), such as “Finally the climbers reached the top of the *mountain/tulip*.” Materials were taken from a

Table 2. Fillers

<i>Fillers</i>	<i>Sentence Type</i>	<i>Example</i>
MP-incongruent	Congruent context	Mark lives on a farm far away from the town.
	Premise 1	If Mark is going to the town, then he will go by scooter.
	Premise 2	Mark is going to the town.
	Incongruent conclusion	Mark will go by bike.
AC-congruent	Congruent context	Golf is becoming a popular sport
	Premise 1	If Luc is going to play golf, then he will wear a hat.
	Premise 2	Luc will wear a hat.
	Congruent conclusion	Luc is going to play golf.
AC-incongruent	Congruent context	Miriam likes water sports.
	Premise 1	If Miriam is going to the lake, then she will go rowing.
	Premise 2	Miriam will go rowing.
	Incongruent conclusion	Miriam is going to the forest.

MP = modus ponens; AC = affirmation of the consequent.

previous study by Van den Brink, Brown, and Hagoort (2001). Final words were matched for number of letters and frequency. Mean cloze probability³ of the sentences was 94% (range 80–100%). As in the reasoning conditions, two different stimulus lists were created to counterbalance congruency so that no participant saw the same item more than once. Each stimulus list was presented to an equal number of participants. Moreover, four versions were created in which the order of the items was reversed according to a Latin square design. Thus, no sentence always occurred at the same position in the experiment. Finally, congruent and incongruent sentences were presented in pseudorandom order with the constraint that the same condition never occurred more than three times in a row.

Procedure

After participants had completed the Reading Span Test, the electrodes were placed on the scalp. Participants received written instructions in which they were informed that they had to decide whether a conclusion followed from short stories. They were instructed to read all sentences carefully and to respond by pressing one of the buttons “yes,” “no,” or “maybe” on a button box. Participants were instructed to sit quietly in a comfortable position and not to blink during the word-by-word presentation of the sentences. Stimuli were presented in white font against a black background using Presentation 10.2 software.

The materials were partly presented in whole sentences and partly word-by-word when good time-locking was critical. The trial sequence was as follows (Figure 1). Each trial started with a 3000-msec fixation cross (+) on the screen. Then the context sentence was presented for a duration of 2000 msec plus an additional 250 msec times the number of words. After the context sentence, the first part of the conditional (“If... then”) appeared for 2000 msec plus an additional 250 msec times the number of words. Subsequent sentences were presented word-by-word. Each word was displayed for 300 msec, followed by a blank screen for another 300 msec, after which the next word appeared. The conclusion was preceded by three hash marks (###) to indicate that the conclusion was following. After the final word of the conclusion, there was a 1000-msec blank screen before the response options MAYBE–YES–NO appeared on the screen for 4000 msec. A blank screen ap-

peared between sentences. Reasoning problems were presented in blocks of 10 trials. After each block there was an optional break. The session started with a practice block of 10 reasoning problems to familiarize the participant with the procedure.

After the reasoning problems, the control conditions were presented in serial visual presentation (300 msec + 300 msec interstimulus interval, and a 3000-msec fixation cross between sentences). Participants were instructed to read for comprehension only, and to minimize eye blinks during the word-by-word presentation. No additional task demands were imposed. There were five blocks of sentences with optional breaks in between. The whole EEG session lasted approximately 75 min without breaks.

EEG Recording

The EEG was recorded from 29 electrode sites across the scalp using an EasyCap with Ag/AgCl electrodes. Recordings were referenced to the left mastoid. Three additional electrodes were placed to monitor eye movements. Vertical EOG was recorded by placing an electrode below the right eye and Fp1 was used for above the eye. Horizontal EOG was recorded via a right-to-left canthal montage. All EEG and EOG channels were amplified with BrainAmp DC amplifiers, using a band-pass filter from 10 sec to 125 Hz. The EEG and EOG signals were recorded and digitized using Brain Vision Recorder software with a sampling frequency of 500 Hz. Impedances were kept below 10 kΩ for EOG and below 5 kΩ for all other channels.

Data Analysis

Both behavioral responses and ERPs were analyzed. Because each condition consisted of 40 items, percentages of accepted items (“yes” responses) per condition were calculated. A nonparametric Mann–Whitney test (exact, two-tailed) was used to examine whether responses were different across contexts.

Prior to analyzing, EEG data were preprocessed using Brain Vision Analyzer software. EEG data were re-referenced to the mean of the two mastoids, and corrected for eye movement artifacts using an algorithm described by Gratton, Coles, and Donchin (1983). Data were filtered off-line with a 30-Hz low-pass filter. Data were segmented from 150 msec

+	Context		Premise 1a		Premise 1b		Premise 2		###	Conclusion		Response
3000	2000 + 250w	600	1000 + 250w	300	word-by-word	1000	word-by-word	1000	300	word-by-word	1000	4000
	Lisa probably lost a contact lens.		If Lisa is going to play hockey, then		she will wear contact lenses.		Lisa is going to play hockey.			Lisa will wear contact lenses.		

Figure 1. Setup of how stimuli were presented. Times are in milliseconds (msec), *w* stands for the number of words per sentence, white boxes represent blank screens. Premise 1b, Premise 2, and the conclusion were presented word-by-word for 300 msec + 300 msec interstimulus interval (ISI) per word.

before to 1000 msec after the onset of the critical words (final words of Premise 1, Premise 2, and conclusion). Baseline correction used the 150-msec interval preceding the onset of the critical word. Trials containing artifacts were rejected (11%).

For each participant, average waveforms were computed across all remaining trials per condition. The average waveforms were analyzed over the 1000-msec latency window using a cluster-based random permutation procedure, implemented in the Fieldtrip toolbox (Maris & Oostenveld, 2007), which has the advantage that it controls for Type I error rates involving multiple comparisons. In this procedure, clusters are identified that differ significantly between conditions in the temporal and/or spatial domain. Specifically, *t* statistics were computed for each data point, and a clustering algorithm formed clusters of data points based on significant *t* tests between conditions in a contrast. For each cluster, a cluster-level statistic was calculated by taking the sum of all the individual *t* statistics within that cluster. The Type I error rate was controlled by evaluating the cluster-level statistics under a randomization null distribution of the maximum cluster-level statistic. This randomization null distribution was obtained by randomizing the data between the two conditions across participants in 1000 randomizations. For each of these randomizations, cluster-level statistics were computed and the largest cluster-level statistic was entered into the null distribution. Finally, the actually observed cluster-level statistic was compared against the randomization null distribution. Clusters that had a *p* value below .05 were considered significant. The procedure is more fully described in Maris and Oostenveld (2007), as well as in the documentation available at www.ru.nl/fcdonders/fieldtrip.

RESULTS

Behavioral Results

Reading span scores ranged from 44 to 86 ($M = 68$, $SD = 10.2$). On the reasoning problems, participants accepted significantly fewer inferences in the disabling context than in the congruent context ($U = 28.5$, $p < .001$; Figure 2). A Pearson's correlation revealed no significant correlation between reading span and percentage of accepted inferences ($p > .1$).

Although we used a delayed response paradigm to prevent artifacts related to responses in the critical ERP latency window, reaction times in the disabling condition were still significantly longer in the disabling condition ($M = 1325$ msec, $SD = 296$ msec) than in the congruent condition [$M = 1207$ msec, $SD = 351$ msec; $t(17) = 3.47$, $p = .003$].

ERP Results

Figure 3A–C displays the grand-average ERPs of the disabling and congruent condition time-locked to the onset of the last word of the first premise, the second premise, and the conclusion. Visual inspection of the waveforms shows a clear N1

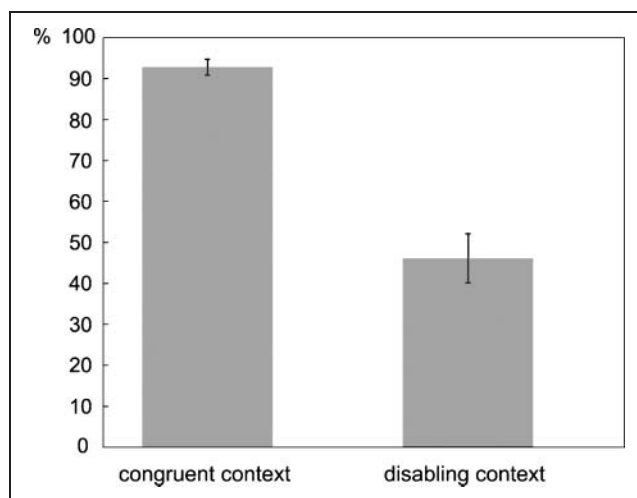


Figure 2. Percentage of accepted inferences (“yes” responses) for the congruent context and the disabling context. Error bars represent 1 SE of the means.

followed by P2, which are characteristic for visual stimuli. Figure 3C reveals a negative shift at the conclusion of the disabling condition relative to the congruent condition.

In the statistical comparisons, this negative shift was expressed by a large, significant negative cluster ($p < .001$). This cluster was present from about 250 to 1000 msec (see Figure 4A), and was most pronounced at the central regions (Figure 5A). There were no other significant clusters found for the contrast. Moreover, statistical analysis of the ERPs time-locked to the final words of the first and second premise did not reveal any significant clusters ($ps > .1$).

We looked for correlations between the negative shift and the percentage of accepted inferences, as well as between the negative shift and reading span, in the latency window from 250 to 1000 msec, using the mean amplitude difference of central electrodes (FCz, FC1, FC2, Cz, CP1, and CP2), based on the topographical distribution of the effect. There was neither a significant correlation between the negative shift and the percentage of accepted inferences nor a correlation between the negative shift and reading span ($ps > .1$).

Figure 3D displays the grand-average ERPs of the control conditions containing semantically congruent and incongruent final words. These waveforms also show an N1–P2 complex, and a clear negative shift for the incongruent sentences relative to congruent sentences with a peak around 400 msec. Cluster-based statistics indeed found a significant, negative cluster ($p = .001$), which was present from about 260 to 470 msec (see Figure 4B), and was maximal over the centro-posterior region (Figure 5B). Furthermore, a late positive cluster was found where the incongruent condition was more positive than the congruent condition, emerging after approximately 500–600 msec ($p = .016$). We will further disregard this late positive cluster as the aim was to elicit a standard N400 effect (for a review on late positive components in N400 paradigms, see Van Petten & Luka, 2006). Based on its latency window and topographical

distribution, it is clear that the negative cluster is an instance of a standard N400 effect.

DISCUSSION

The aim of the present work was to investigate the time course of defeasible reasoning in the brain. For that purpose, ERPs were recorded while participants read modus ponens inferences, which were preceded by a congruent

context or a disabling context that contained a possible exception that could prevent people from drawing the conclusion. We indeed found that people accepted considerably fewer inferences in the disabling condition than in the congruent condition. The ERP findings showed that the disabling condition elicited a widely distributed sustained negativity relative to its congruent counterpart. This negativity started around 250 msec after onset of the final word of the conclusion and was persistent throughout the entire epoch. Participants also read an additional

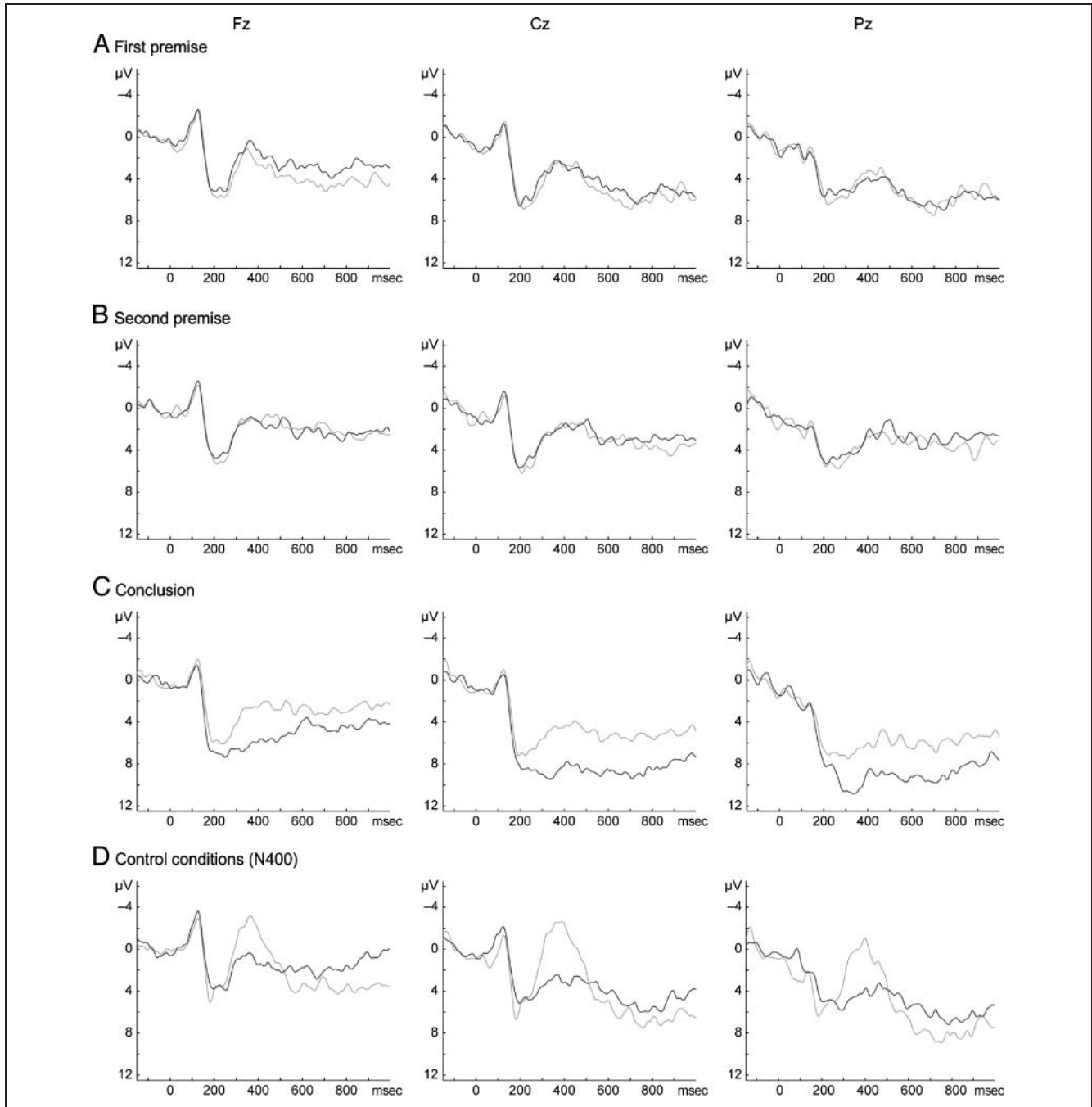


Figure 3. Grand-average ERPs from Fz, Cz, and Pz for the congruent and disabling context time-locked to the final word of Premise 1 (A), Premise 2 (B), and the conclusion (C). Black line = congruent context; gray line = disabling context. (D) The grand-average ERPs for the control conditions time-locked to the sentence-final word. Here, black line = congruent sentence; gray line = incongruent sentence. Negative values are plotted upward.

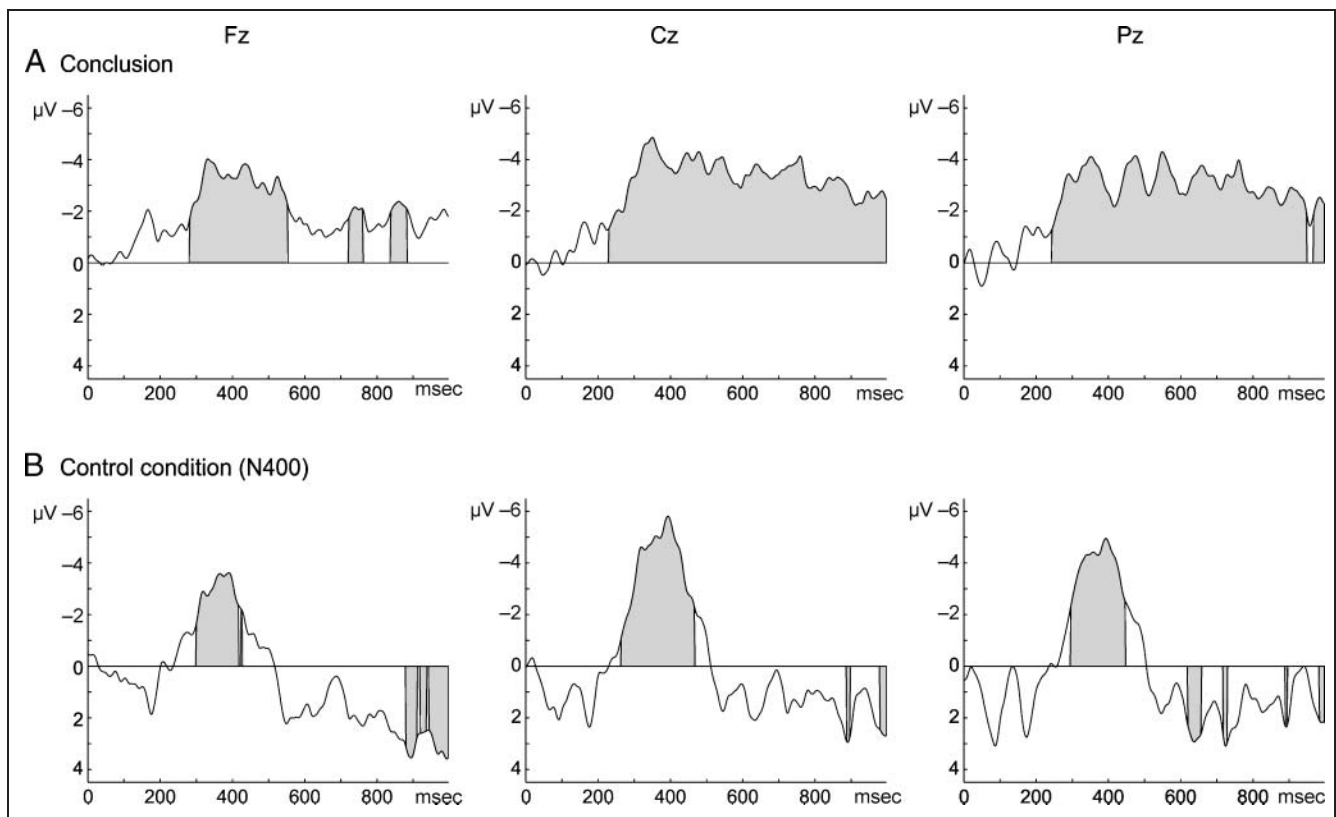


Figure 4. Difference waveform of the reasoning conditions (A) time-locked to the final word of the conclusion, and of the control conditions (B) time-locked to the sentence-final word. Gray blocks represent significant areas. Negative values are plotted upward.

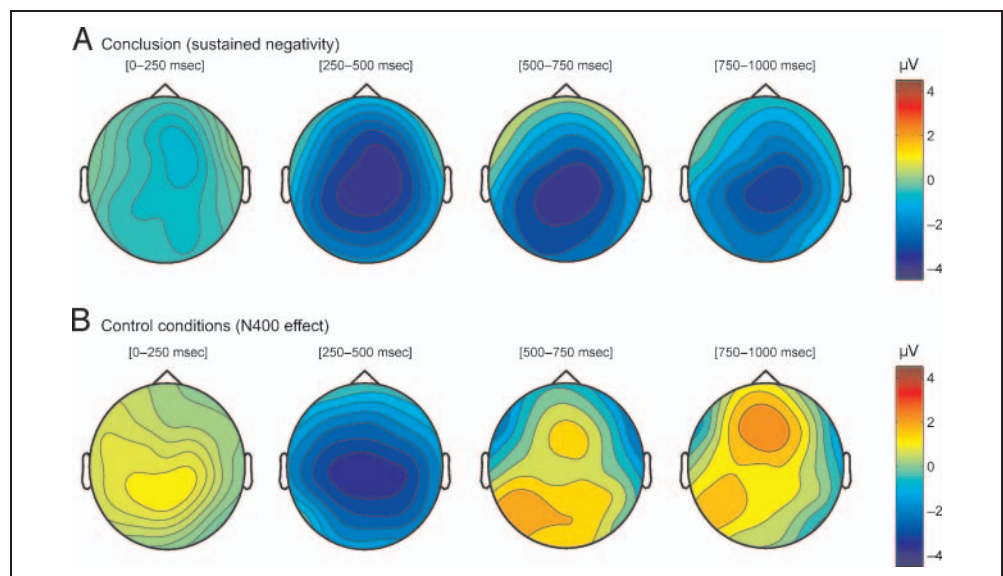
set of control sentences, in which semantic congruency was manipulated in order to elicit a standard N400 effect, which was indeed found, followed by a late positive component for words that were semantically anomalous with the prior sentence relative to semantically congruent words.

It is clear that the observed negativity in the reasoning conditions differs from the standard N400 effect in the control conditions in terms of its morphology and temporal profile: A peak is lacking and the effect is much more sus-

tained than a standard N400 effect. This suggests that the effect evoked by the reasoning conditions is different from that elicited by semantic anomalies. However, the scalp distribution of the effects was very similar. Hence, the sustained negativity may reflect the contribution of the same neural processes as the N400 effect.

Based on other studies that have observed sustained negativities, the observed negativity in the reasoning conditions might reflect additional processing because a

Figure 5. Scalp distribution of the sustained negativity at the conclusion in the reasoning conditions (A), and of the N400 effect in the control conditions (B). Scalp distributions are based on mean amplitude differences in four consecutive time intervals of 250 msec length. Scale values are in microvolts (μV).



default inference must be revised to incorporate an exception. This explanation is supported by Baggio et al. (2008), who also found a sustained negativity when a default inference had to be overridden and revised, albeit that the scalp distribution of the observed effect in our study was more central than in their study. Alternatively, the sustained negativity may reflect an attempt to link the exception with information retrieved from long-term memory, or extra working memory demands to hold information about the exception in mind in order to withdraw the conclusion (Markovits & Potvin, 2001; Vadeboncoeur & Markovits, 1999; Rösler, Heil, & Glowalla, 1993). However, the working memory account seems less likely because we failed to find any relationship between reading span and suppression, and between reading span and the ERP effect. Reading span is usually taken as a measure for verbal working memory capacity (Van den Noort et al., 2008). Although the current evidence does not support any associations between verbal working memory capacity and defeasible reasoning, some caution is required when excluding the working memory account. Because reading span is solely based on the storage and active recall of words, it may not be a good index of the kind of verbal working memory involved in defeasible reasoning (Waters & Caplan, 1996).

As mentioned above, it is not clear whether the observed sustained negativity is different from a standard N400 effect, because it had a scalp distribution that is similar to that of a standard N400 effect. The N400 effect is usually associated with interpretive problems: As the integration of the meaning of a word into a representation of its preceding context becomes harder, the amplitude of the N400 increases. If the sustained negativity is indeed an instance of an N400 effect, then it appears to be associated with interpretive problems. One could argue that the reasoning conditions show a large overlap with “discourse-N400” conditions, in the sense that a number of sentences have to be integrated to arrive at a discourse-level representation. However, several studies have shown that words that conflict with the wider discourse elicit a standard N400 effect instead of a sustained negativity (Nieuwland & Van Berkum, 2006; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003; Salmon & Pratt, 2002; Van Berkum, Hagoort, et al., 1999). In the light of this evidence, it seems less likely that the observed sustained effect is due to difficulties with the integration of the conclusion into the preceding discourse. Moreover, the sustained nature of the negativity also makes it unlikely that it could be attributed only to a conflict between the context and the conclusion while ignoring the premises. In that case, a more discourse-like N400 would have been expected. Perhaps the observed effect is more sustained because the processes that are involved in defeasible reasoning are more demanding than in a “discourse-N400” paradigm. In a recent study, Baggio, Choma, van Lambalgen, and Hagoort (2010) found a similar sustained central negativity for coercion verbs such as *to begin* in “The journalist *began* the article

before his coffee break” compared to “The journalist *wrote* the article before his coffee break.” The first sentence requires the reader to infer what is actually meant by *began*. Thus, coercion verbs involve some semantic enrichment, which seems to require additional processing. Baggio et al. (2010) suggest that the sustained central negativity—or “N400-like shift”—they observed may be associated with more complex, inference-driven integration of information into a semantic representation. In a similar way, the observed sustained negativity in the present study could also reflect more complex, inference-driven interpretive processes, resulting in a sustained N400-like effect.

In conclusion, up to now, little was known about the time course of defeasible reasoning in the brain. Our work demonstrates that just within 250 msec after the onset of the final word of the conclusion, there was an electrophysiological brain response observed when the conclusion does not fit with the context, which was persistent throughout the entire epoch. The observed effect differed from that of semantic anomaly, at least in its morphology and temporal profile. However, we cannot conclude that the effect is qualitatively different from a standard N400 effect, because both effects had a central scalp distribution. Importantly, regardless of the exact nature of the observed sustained negativity, the processing of defeasible inferences seems to be more effortful than default inferences. Because this ERP study on reasoning was done in a largely unexplored field, the exact interpretation of the observed effect remains open. Further research is needed to disentangle processes related to reasoning from linguistic processes.

Acknowledgments

We thank Sander Berends for his assistance during the EEG measurements and Marcel Bastiaansen for his help with Fieldtrip. This research was supported by a grant from the Netherlands Organization for Scientific Research, NWOCOG/04-19.

Reprint requests should be sent to Judith Pijnacker or Peter Hagoort, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands, or via e-mail: j.pijnacker@pwo.ru.nl; p.hagoort@donders.ru.nl.

Notes

1. Critical words are in italics.
2. It is a well-known fact that Dutch trains are yellow, therefore the sentence “Dutch trains are white” is false, although the sentence itself is semantically well-formed.
3. Cloze probability is determined by measuring the probability that a particular word is given on a sentence completion task. The higher the cloze probability, the more a particular word is expected.

REFERENCES

- Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, 22, 2131–2140.

- Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, *59*, 36–53.
- Bonnefon, J., & Hilton, D. J. (2002). The suppression of modus ponens as a case of pragmatic preconditional reasoning. *Thinking and Reasoning*, *8*, 21–40.
- Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34–44.
- Byrne, R. M. J. (1989). Suppression of valid inferences with conditionals. *Cognition*, *31*, 61–83.
- Byrne, R. M. J. (1991). Can valid inferences be suppressed? *Cognition*, *39*, 71–78.
- Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*, 347–373.
- Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, *53*, 217–238.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282.
- Dieussaert, K., Schaeken, W., Schroyens, W., & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking and Reasoning*, *6*, 125–169.
- Evans, S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Psychology Press, Taylor & Francis Group.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.
- Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical-ambiguity resolution and parsing. In L. F. C. Clifton, Jr. & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–80). Hillsdale, NJ: Erlbaum.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 376–395.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, *29*, 736–744.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, *5*, 193–203.
- Münste, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*, 71–73.
- Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098–1111.
- Pijnacker, J., Geurts, B., van Lambalgen, M., Kan, C. C., Buitelaar, J. K., & Hagoort, P. (2009). Defeasible reasoning in high-functioning adults with autism: Evidence for impaired exception-handling. *Neuropsychologia*, *47*, 644–651.
- Politzer, G., & Bourmaud, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology*, *93*, 345–381.
- Qiu, J., Li, H., Huang, X., Zhang, F., Chen, A., Luo, Y., et al. (2007). The neural basis of conditional reasoning: An event-related potential study. *Neuropsychologia*, *45*, 1533–1539.
- Rösler, F., Heil, M., & Glowalla, U. (1993). Monitoring retrieval from long-term memory by slow event-related brain potentials. *Psychophysiology*, *30*, 170–182.
- Salmon, N., & Pratt, H. (2002). A comparison of sentence- and discourse-level semantic processing: An ERP study. *Brain and Language*, *83*, 367–383.
- St. George, M., Mannes, S., & Hoffman, A. J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, *9*, 776–787.
- Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*, 919–960.
- Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. Massachusetts: MIT Press.
- Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking and Reasoning*, *5*, 97–113.
- Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- Van Berkum, J. J., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*, 235–248.
- Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- Van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701–718.
- Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967–985.
- Van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, *24*, 35–42.
- Van den Noort, M., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*, 289–296.
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*, 279–293.
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, *49*, 51–75.
- Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103–115.