Inferential updating in narrative texts: an ERP study

Pérez, A., Cain, K., Castellanos, M.C., & Bajo, T
University of Granada (Spain)

Acknowledgements

This study has been supported by the Spanish Ministry of Education and Science (FPU grant to A.P.), by grants EDU2008–01111 and CSD2008–00048 from the Spanish Ministry of Education; and grants PSI2012–33625 and P08–HUM–03600 from the Andalusian Government to TB and by the department of Psychology of Lancaster University (UK). In addition, we would like to thank Thomas Lorsbach for providing his materials, Kate Nation and Holly Joseph to contribute in the construction of some of the additional materials, and Sasha Smith for help with piloting.

Abstract

We evaluated the process of inferential updating during text comprehension in adults. The introduction supported two plausible concepts (e.g., '*guitar/violin*'), although one was more probable ('*guitar*'). There were three possible continuations: a neutral sentence, which did not refer back to either concept; a no update sentence, referred to a general property consistent with either concept (e.g., '…*beautiful curved body*'); and an update condition, referred to a property that was consistent with only the less likely concept (e.g., '…*matching bow*'). Both working memory groups took longer to read the sentence in the update condition. In a final sentence, a target noun referred to the alternative concept supported in the update condition ('*violin*'). Only high working memory readers updated their initial incorrect interpretation (P3b) and integrated this new inference (N400) reading the update sentence. Low working memory readers had problems accurately representing semantically related inferential concepts, failing to update their situation model.

Keywords: updating information; inference making; working memory; P3b; N400.

3.3.1. Introduction

Successful text comprehenders construct an integrated, coherent and accurate mental representation of the state of affairs described by the text. The construction of this situation model requires the reader to go beyond a representation of the surface characteristics of the text, and to incorporate world knowledge from long–term memory (Kintsch & van Dijk, 1978). This is a dynamic process; the situation presented in a text is described incrementally and can, therefore, change as the text unfolds. As each new piece of information is processed, it must be integrated with the mental representation constructed thus far, which involves revising and updating the current situation model (O'Brien, Rizzella, Albrecht, & Halleran, 1998; Zwaan & Radvansky, 1998).

Situation model updating can involve the addition of new concepts or information, a change in the encoded relations between events, and also the elimination of information from the representation if it is no longer relevant (Radvansky & Copeland, 2001). Skilled comprehenders evaluate new information against the current representation. When consistent with the current situation model, readers readily map the new information onto the current structure, drawing inferences as necessary (Morrow, Bower, & Greenspan, 1989). When new information is inconsistent with the current model, comprehension difficulties may occur (O'Brien et al., 1998). In such circumstances, a successful comprehender may revise the current situation model, which may involve inferential processing or shifting to build a new substructure if there is a significant change in topic (Gernsbacher, 1990).

Despite widespread agreement that comprehension is a dynamic process that involves inferential processing and updating, readers may not always successfully perform these processes when new information contradicts the current model. In a classic example, O'Brien et al. (1998) found that participants took longer to read a sentence regarding a person's behavior when the behaviour contradicted earlier information, e.g., reading '*Mary ordered a cheeseburger and fries*' after '*Mary, a health nut, had been a strict vegetarian for ten years*'. This finding indicates that participants experienced difficulty integrating the new information into their mental representation, because the new information was inconsistent with the earlier inference that '*Mary did not eat meat*'. This comprehension difficulty was reduced, but still evident, in a qualified condition that provided an additional explanation for the character's behavior encouraging a revision of the previously made inference ('*Nevertheless, Mary never stuck to her diet when she dined out with her friends*'). If participants had successfully updated their mental representation to incorporate this qualification, there would have been no comprehension difficulty. Thus, when new information is inconsistent with prior parts of the text, successful understanding requires the revision of the situation model.

The combination of both inference making and updating information is what we have called inferential updating. That is, text comprehension sometimes involves updating the mental representation created from information explicitly stated in the text or from an inference that is supported by the text and incorporated in the situation model. Evidence to date from a number of different paradigms has shown that readers do not always successfully revise and update their mental representation (e.g., Rapp & Kendeou, 2009). In this paper, we aim to understand better the dynamics of the inferential updating process and to explore one reader characteristic that might explain why some readers have difficulties with this process: working memory.

The construction of the situation model draws on working memory resources, particularly those related to the central executive (Baddeley & Hitch, 1974) and executive functions (Lehto, 1996; Morris & Jones, 1990). Working memory has been found to influence both inference making (e.g., Morrow et al., 1989) and updating (e.g., Carretti, Cornoldi, De Beni, & Romanò, 2005; Dutke & von Hecker, 2011). Consequently, there are two (not mutually exclusive) reasons why working memory may constrain an individual's ability to update inferences that they have encoded in their mental representation, when new information prompts a different or more specific interpretation. First, the reader has to activate and maintain inferred information generated from previous parts of the text to evaluate incoming information, a typical process during narrative comprehension. Thus, a person with low working memory might not be able to detect inconsistencies in the text because he/she has not accurately activated and/or maintained previous non–explicit information (MacDonald, Just, & Carpenter, 1992; Whitney, Ritchie, & Clark, 1991). Second, if new information disconfirms a previous interpretation, the revision process involves not only the activation and detection of the new information that prompts the update, but also the inhibition of the previous incorrect interpretation, a process that readers with low memory may find hard (Carretti et al., 2005; Dutke & von Hecker, 2011).

In general, measures of working memory that tap the central executive component are related to reading comprehension (Daneman & Merikle, 1996). Research has demonstrated that there is a specific link between working memory and the essential processes of reading comprehension such as maintenance of the current ideas of the text, activation of new information, and inhibition of outdated information. Readers with high working memory capacity are more likely than those with weaker memory skills to make elaborative inferences when these are strongly supported by discourse context (St George, Mannes, & Hoffman, 1997). In addition, when presented with ambiguous sentences that allow more than one interpretation, readers with high memory capacity are better able than readers with low memory capacity to maintain competing representations until the ambiguity can be resolved (MacDonald et al., 1992), or they are less likely to commit to a specific interpretation than are low memory readers (Whitney et al., 1991). Finally, readers with high working memory capacity are better able to inhibit information that is no longer relevant than are those with weaker memory skills (Carretti et al., 2005; Dutke & von Hecker, 2011). This latter finding is congruent with studies showing that both adults and children with poor reading comprehension also have difficulties inhibiting no longer relevant information from memory (e.g., Cain, 2006; De Beni, Palladino, Pazzaglia, & Cornoldi, 1998).

From this evidence, it appears that it will be easier for readers with good working memory capacity to revise or update their interpretation of a text when they encounter new information because it will be consistent with one of the currently active interpretations of events, and/or the prior interpretation will be quickly discarded when the new information is activated. For readers with weaker working memory, a comprehension difficulty will occur if the new information disconfirms their (single) interpretation and/or the initial incorrect interpretation is maintained for too long, because of slow inhibitory processes.

Surprisingly, there are very few studies that have examined the process of inferential updating. One relevant study conducted by Dutke and von Hecker (2011) provides evidence that an individual's memory capacity is associated with updating the situation model. In their paradigm, a narrative provided information about social relations between protagonists. Participants with high memory scores were better able than a group with lower memory scores to maintain and recall the situation model and to discard an earlier representation that was incompatible with new information. Thus, in situations where updating is required – when the relations between concepts and events become better specified and less ambiguous as the text unfolds, readers with low working memory have difficulties updating at the situation model level.

Dutke and von Hecker's (2011) materials concerned the relational structure of text and, therefore, do not speak directly to the role of inferential updating during the construction of the situation model. One of the few studies exploring some aspects of what we have called inferential updating is the developmental study of Lorsbach, Katz, and Cupak (1998). In their experiment, children and adults read garden path passages, in which the introductory sentences supported two different interpretations of an object, one more likely than the other. All participants were equally likely to set up the expected inference, but the children were less able to inhibit this when subsequent text favoured the alternative concept and, therefore, required an update to the situation model. Although the processing of these texts clearly required listeners to update their situation model, the authors discussed their results in terms of competing inferences and did not explain their results in relation to the updating process.

Inferences and the updating of the contents of the situation model are critical to the construction of a coherent situation model (Graesser, Singer, & Trabasso, 1994). Therefore, an understanding of the accuracy and time course of updating an inference is fundamental for comprehensive theories of text processing. The study reported here investigated the relation between inferential updating and working memory in two important ways. First, we manipulated whether or not the updating of inferences was required at the situation model level. Second, we examined if participants with poor working memory skills were also poor at updating their situation model during reading. The purpose of our study was to explore the interplay between these processes by recording electrical brain activity during reading.

Event–related brain potentials (ERPs) are a robust means to study the time course of inference making and whether or not an update has been made. In relation to our goal, an interesting theoretical framework is the context–updating theory (Polich, 2003; 2007). This framework distinguishes two subcomponents of the P300: a central–frontal positivity or "P3a" (e.g., Debener, Makeig, Delorme, & Engel, 2005), which appears when an incoming information is evaluated as new or different with respect the current representation demanding attentional control; and a posterior positivity or "P3b" (e.g., Hartikainen & Knight, 2003; Kok, 2001), which is found when the context of the incoming stimulus involves updating by memory processes. Polich (2003) suggests a brain circuit between a top–down process in the frontal lobe driven by focal attention (P3a) and a bottom–up process in the temporo–parietal areas guided by memory updating operations (P3b). More important, some studies have demonstrated a selective relationship between the reduction of the P3b amplitude and a poor execution in several capacities as working memory (e.g., Evans, Sellinger, & Pollak, 2011); comprehension monitoring (e.g., Getzmann & Falkestein, 2011[1]); or conflicting response–performance (e.g., Trewartha, Spilka, Penhune, & Phillips, 2013). This evidence is crucial for the present experiment. Another component of interest in our study is the N400, which is an index of the ease with which the meaning of a word can be integrated into the current situation model (see Kutas & Federmeier, 2009; 2011, for a review). The amplitude of the N400 is attenuated when there is a good fit between the word being processed and the context, in comparison to a poorly fitting word. This attenuation of the N400 component has been demonstrated for words related to a causal inference supported by the previous text, both during text reading (Kuperberg, Paczynski, & Ditman, 2011; St George et al.,

[1] Getzmann and Falkestein (2011) also observed a right lateralized activation of the P3a in high–performing older listeners compared to low–performing older and high/low–performing younger listeners, which they interpreted as a compensatory mechanism of age to improve comprehension monitoring.

1997) and subsequent lexical decision (Steele, Bernat, van den Broek, Collins, Patrick, & Marsolek, 2012).

In our study, we used ERPs to gain a more detailed view of when readers update their situation model using the following paradigm. First, readers were presented with a three–sentence introductory text that could support two different concepts, e.g., '*guitar'* and '*violin*'. The introduction was written such that neither concept was mentioned explicitly, both were plausible, but one was considered more likely by independent judges. Readers were then presented with one of three possible conditions: a neutral sentence, which did not refer directly or indirectly to either concept (e.g., '*The concert takes place at…*'); a no update sentence, which mentioned a property consistent with either concept (e.g., '*…made of maple wood, with a beautiful curved body*'); and an update condition, in which the sentence referred to a property that was consistent with only the less likely concept (e.g., '*…made of maple wood, with a matching bow*'). The less likely concept was the one least supported by the introduction such that readers in this condition were required to update their situation model to ensure good comprehension. In the concluding sentence of the text, the final word was consistent with the concept supported in this latter update condition (e.g., '*violin*'). A preliminary proposal was that readers who successfully update their situation model should elicit an increase of the N400 in the no update compared to the update condition, since the semantic representation of the disambiguating word has not been incorporated to the situation model yet. This is because the N400 is an indicator of integration difficulty: the update condition should enable readers to integrate easily the target word into their situation model, whereas the no update condition should lead to difficulty because the target word is semantically inconsistent. In addition, considering that there is a strong relation between working memory and critical reading comprehension processes, such as updating and inference making, we also hypothesized working memory differences in the N400. Specifically, we predicted that those with high working memory scores would be more likely to update their situation model when reading the update sentence than those with low working memory scores, resulting in larger N400 (in the no update condition) for the high working memory group.

In relation to these hypotheses, a preliminary study of inferential updating process in adults (Pérez, Cain, Castellanos, & Bajo, 2012) was carried out. In this study, we compared the performance of adults with high and low working memory on two of the conditions described above: texts that prompted an update in the fourth sentence following the introduction and those that did not. We did not include the neutral condition described above in this pilot. We found that memory did not influence reading times for the update sentence, suggesting that readers with high and low working memory were equally likely to engage in additional processing, most likely detecting an inconsistency between the update information and their memory representation and/or revising their situation model in line with that information. The measurement of the N400 event–related potential revealed that both memory groups showed evidence of detecting a mismatch between the disambiguating word and the situation model. However, the electrophysiological pattern that emerged for each group was very distinct: the high memory group manifested a large N400 in the no update compared to the update condition in posterior regions; while the low memory group presented a relative difference between conditions generalized to all regions (main effect of condition). Therefore, we decided to carry out another study to clarify these results.

Table 4. Example of text used in the situation model updating task.

Dan was a gypsy who had played flamenco since childhood. Now he is a popular musician who plays all over the world. Today, he is giving a recital of his favourite works.	Introduction (bias *guitar*)
- The concert takes place at the prestigious national concert hall.	Neutral
- His instrument is made of maple wood, with a *beautiful curved body*.	No update
- His instrument is made of maple wood, with a *matching bow*.	*Update*
The public was delighted to hear Dan playing the **violin**.	ERP
In the recital, Dan played his favourite works.	Comprehension sentence

Note. Participants saw the text in the neutral, no update or update condition. The word in bold in the ERP sentence was the disambiguating word.

The current study

Our pilot confirmed the situation model updating task as a promising framework for the study of inferential updating. However, in order to clarify the cognitive processes underlying the reading time data, as well as the relative difference found in the ERP amplitude, it is necessary to include a critical third condition – the neutral condition described above, in which the fourth sentence does not directly or indirectly refer to either of the two critical concepts. We included this third condition in the study we report here. We measured the reading time for sentence 4, which could either require no update to the situation model (as in the no update and neutral conditions) or an update (as described above and shown in Table 4; see Appendix F for full materials, p. 205). We also measured ERPs on the final word of sentence 5, in which the concept matched the inference prompted in the update condition but did not match the concept most strongly supported by the introduction for the other two conditions. According to the context–updating theory, our general proposal is that a frontal activation (P3a) will be found for the disambiguating word when readers have encountered the no update and neutral conditions compared to when they have read the update condition. This result would indicate the detection of information that is new or different with respect to the representation active in working memory. Furthermore, if readers successfully updated their situation model in line with the information provided by the update condition, they will demonstrate a reduced parietal activation (P3b) on reading this final disambiguating word compared to when they read the no update and neutral conditions. This is because readers would not draw on additional memory processes to revise their situation model, since the meaning of the disambiguating word is already encoded. Similarly, but as a consequence of a lack of updating, the same disambiguating word should elicit a N400 only in those conditions that did not help to update the context (no update and neutral) compared to the update condition, because the semantic representation of the less likely concept is not yet in their situation model.

Our predictions were as follows. First, in relation to the reading time for the sentence 4, we predicted longer reading times in the update condition relative to both other conditions. This prediction holds for both high and low memory readers, in line with our pilot. Longer reading times could reflect several (not mutually exclusive) processes: identification that the information is inconsistent with the current situation model, inhibition of the no–longer relevant inference that was originally encoded in the model, and the additional processing involved in generating an inference to update the situation model. Second, in relation to the ERP measures, we predicted no working memory group differences in the P3a subcomponent, since our preliminary study showed (behavioural results) that all participants were able to detect an inconsistency between the new concept and the general context of the introduction. Nevertheless, we predicted working memory differences in both the P3b and N400: high working memory readers will be able to accurately update their situation model when reading the update sentence compared to low working memory readers. If this is true, a reduction of the P3b in the update condition (e.g., reading '*violin*' after '*matching bow*') and an increase of the N400 in the no update and neutral conditions should occur for the high but not for the low working memory group. To our knowledge, this is the first electrophysiological study investigating working memory differences of the inferential updating process in reading comprehension.

3.3.2. Method

Participants

Seventy–seven people living in the city of Granada (Spain) with a mean age of 22.5 years old (range: 18–37 years) were recruited by an internet advertisement to participate for money. All were native English speakers and gave their consent to participate in the experiment. After they performed the two memory tasks (memory updating and reading span tasks), only participants with extreme working memory scores (see below) were selected to participate in the situation model updating task.

Materials

Memory updating task. We developed an English version of Carretti, Belacchi, and Cornoldi's (2010) memory updating task. Participants read lists of words, one word at a time, which increased in length as the trials progressed. The words were concrete nouns referring to objects of different sizes (large or small, e.g., '*ship*' or '*pea*'). The task was to recall the smallest object/s in the list. The number to be recalled was stated before each list and increased from 1 (Level 1) to 5 (Level 5), with a fixed presentation order. The recall set had to be updated as new words were presented. Participants were required to maintain the items in working memory to compare their sizes, to maintain activation of the smallest items in the specified set size, and to inhibit any previously activated words that no longer meet the criteria (that is to inhibit a large–size item when they heard the name of a smaller item). All participants completed all trials.

Reading span task. We used Daneman and Carpenter's (1980) reading span task. Participants read sets of sentences presented one by one and were required to recall the last word of each sentence, at the end of each set of sentences. The order of recall was not important but participants could not start with the last word of the last sentence. There were five levels increasing in difficulty from 2 (Level 1) to 6 (Level 5) sentences. A level was considered correct if participants recalled correctly each last word of at least three out of five (maximum) sets of sentences.

The scores of both memory tasks were the total number of words correctly recalled minus the total number of words incorrectly recalled (intrusions) in each memory task. These scores were used to divide participants into the high and low working memory groups.

Situation model updating task. We constructed 93 (3 practice, 90 experimental) five–sentence narrative texts, some modified from texts used by Lorsbach et al. (1998). The first three sentences (introduction) biased an inference generated at the situation model level (e.g., '*guitar*'). There were three versions of the fourth sentence: 1) in the neutral condition the sentence did not refer back to the inferable concept. Therefore, this sentence was neither consistent nor inconsistent with the introduction (e.g., '*The concert takes place at the prestigious national concert hall*'); 2) in the no update condition the information was consistent with the concept primed in the introduction (e.g., '*His instrument is made of maple wood, with a beautiful curved body*'); 3) in the update condition the information presented in sentence 4 prompted readers to revise their situation model so that only the alternative concept was encoded, rather than the concept supported by the introduction (e.g., '*His instrument is made of maple wood, with a matching bow*'). Reading times were the dependent variable for this sentence.

A pilot study with a two–alternative forced choice task confirmed that the two critical concepts were both supported by the fourth sentence (e.g., '*guitar*' for the no update, and '*violin*' for the update). Participants read the 3–sentence introduction, followed by one of the two versions of the sentence 4. They were instructed to mark the concept that the text was about. Seven participants completed each version of each text. In the final study, we included only texts for which the appropriate word was selected in both versions by a minimum of five participants. The word frequency for each of the two critical concepts was examined using the Word Frequency Guide database (WFG; *Zeno*, Ivens, Millard, & Duvvuri, 1995) and did not differ (*M* no update = 56.58; *M* update = 47.18, $t(89) = 0.27$, $p = .79$). The word length of the fourth sentence did not differ between conditions (*M*s = 11.70, 11.46, and 11.81, for the neutral, no update and update conditions respectively: $F(2, 178) = 1.74$, $p = .18$). The fifth and final sentence concluded with a disambiguation word (e.g., '*violin*'), which was always incongruent with the concept supported by the introduction and congruent with the concept of the update condition in the fourth sentence. Consequently, the final disambiguation word was unexpected in the no update and neutral conditions and expected in the update condition. At the end of the text, a comprehension sentence was shown which required a true or false judgment, (e.g., '*In the recital, Dan played his favourite works*'). This sentence was included to encourage participants to read for meaning.

To provide empirical confirmation of concept preferences in our situation model updating task, we conducted an additional pilot study with twenty–two participants. Participants read the introduction of each text and were then presented with a single word. Their task was to decide (yes/no) if it fitted with the sense of the story. The word was either the concept in the no update condition, which was most strongly supported by the first three sentences (e.g., '*guitar*'), the concept in the update condition (e.g., '*violin*'), or another noun that did not fit the context (e.g., '*poker*'). Accuracy and response times were analysed separately. Results of the one–way ANOVA performed on the accuracy data showed a main effect of word type: $F(2,42) = 92.92$, $p < .001$, $p\eta^2 = .82$, because the participants were more likely to correctly accept the word in the no update condition ('*guitar*'; $M = 17.63$, $SD = 1.62$) and to correctly reject the word that did not fit the story ('*poker*'; $M = 18.68$, $SD = 1.52$) than to accept the update word ('*violin*'; $M = 10.59$, $SD = 2.61$). Further, when participants did accept the word in the update condition, they took longer to do so ($M = 2079.43$ ms) compared with response times to the no update word ($M = 1612.33$ ms), $t(21) = 3.72$, $p < .001$)[2]. This difference suggests that the concept in the no update condition was significantly more likely to be activated than the update concept, after reading the introduction, as intended.

Procedure

Materials were administered in two sessions. The first session took approximately 30 minutes and included the two memory tasks. The memory updating task was administered first. Before each word list, participants were informed of the number of words in the list and how many objects to recall. Each word was presented for 2 seconds. A question mark prompted recall and the participant said their response out loud. A practice trial preceded the experimental trials. The reading span task was completed next. Participants were instructed to recall the last word of each sentence and, before each block, they were informed of the number of sentences (and words to recall) in the trial. Participants read each sentence at their own pace. At the end of the trial, a white screen appeared and participants said aloud the words that they could remember. A practice trial preceded the experimental trials.

Before the second session, the scores of both working memory tasks were used to divide participants into two groups. The mean number of words recalled for the low working memory group was 21.11 ($SD = 2.74$; range = 16–24) in the memory updating task and 29.50 ($SD = 8.03$; range = 16–44) in the reading span task; the mean number of

[2] Because the rejection search process for words that did not fit the meaning is different from the confirmatory search process for words that fit, we did not include the correct rejection of the nonstory condition in the response time analysis.

words recalled for the high group was 26.39 (*SD* = 1.50; range = 24–29) in the memory updating task and 68.39 (*SD* = 12.10; range = 47–86) in the reading span task. A t–test between groups confirmed our group selection procedure, where significant group differences were apparent in both tasks: memory updating, $t(34) = 7.17$, $p < .001$; and reading span, $t(34) = 11.36$, $p < .001$.

In the second session, participants completed the situation model updating task. This session took approximately 90 minutes and only participants in the low and high working memory groups took part. For this task, we placed the electrode cap onto the participant's head to record the EEG. Each trial started with a fixation point ('+') that remained on the screen until the participant pressed the 'B' key on the keyboard to present the first sentence. Participants read the first four sentences at their own pace, pressing the space bar to display the next sentence. The reading time of the fourth sentence (neutral vs. no update vs. update) was measured. The fifth sentence was presented word by word with a fixed duration of 300 ms per word. In addition, there was a delay of 700 ms after the disambiguation word to ensure that the electrophysiological activities of the event–related potential were registered. Finally, participants were presented with the true/false comprehension sentence. This always referred to information in the introductory sentences (equally distributed across the three sentences) and was included to encourage participants to read for meaning. Participants pressed the designated true or false key to respond. Each of 90 experimental texts was presented to each participant only once in one of the three conditions (neutral, no update or update) counterbalanced across participants. The task was administered in three blocks, keeping the same proportion (10 texts) in each condition per block. The same number of participants completed each condition, and the presentation of texts was randomized within block. A practice of 3 trials ensured that instructions were understood.

Apparatus

All tasks were presented by the E–prime software (*Schneider*, et al., 2002), administered on a 19" CRT video monitor (refresh rate = 75 Hz). For the situation model updating task, scalp voltages were recorded from a SynAmps2 64 channels Quik–Cap, plugged in to a Neuroscan SynAmps RT amplifier. The electrical signal was amplified with a 1–30 Hz band–pass filter and a continuous sample rate of 250 Hz. Ocular movements and blinks were also collected by two pairs of channels: a) the vertical electrooculogram situated in the left eye of the participant, with one electrode supra and another infraorbitally to measure blink artifact; b) the horizontal electrooculogram placed in the external canthi, with one electrode on the left and another on the right side to register eye movements. Impedances were kept below 5kΩ. Both blinks and ocular movements were corrected. In addition, trials with artifacts were rejected (3.12%), and in those cases where electrodes had a high level of artifacts (>1%), these were substituted by the average value of the group of nearest electrodes. Epochs with an interval between –200 and 800 ms with respect to the presentation of the target word (disambiguating word) were averaged and analysed. Baseline correction was applied using the average EEG activity in the 200 ms preceding the onset of the target as a reference signal value. Separate ERPs averages were developed for each condition for each participant. Individual averages were re–referenced off–line to the average of left and right mastoid. Six regions of interest (ROI) were extracted from the 64 channels (see Figure 6), keeping the criteria of 1) symmetry between hemispheres and 2) same number of electrodes (five sites)[3]: left frontal or LF (F1, F3, F5, FC3 and FC5); right frontal or RF (F2, F4, F6, FC4 and FC6); central or C (C1, C2, CZ, FCZ and CPZ); left parietal or LP (P1, P3, P5, CP3 and CP5); right parietal or RP (P2, P4, P6, CP4 and CP6); and occipital or O (O1, O2, POZ, PO3 and PO4).

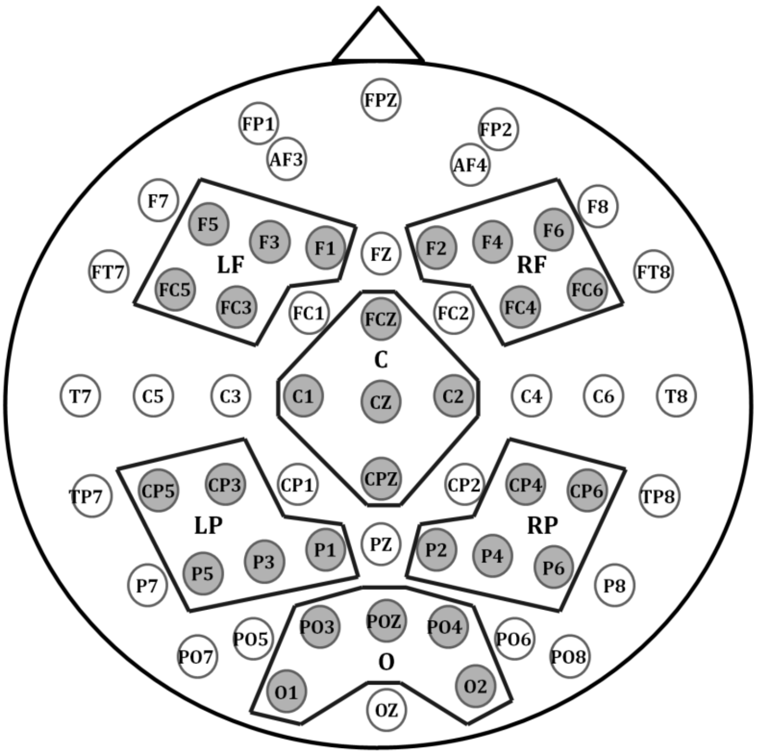

Figure 6. The six regions of interest (ROI): left frontal (LF); right frontal (RF); central (C); left parietal (LP); right parietal (RP); and occipital (O).

[3] Because the magnitude of the components involved in the inferential updating process is still unknown, our regions of interest were selected taking into account a good representation of the different parts of the scalp.

Statistical analyses

We report statistical analyses of thirty–six participants for all trials[4]. Working memory group was a between–subjects factor in all analyses. The behavioural analysis of the situation model updating task was conducted on RT (milliseconds) per sentence. In the ERP analyses, the mean amplitude was calculated in the window of 220–300 ms (P3a and P3b) and the window of 300–550 ms (N400) after the disambiguating word onset (see Figure 7). Outlier amplitude data per continuation, group and ROI was detected by the Box–Whisker plot, and replaced by the mean for both the P300 (3.70%) and the N400 (2.47%).

[4] Responses to the comprehension question were recorded only to ensure that participants were attending. The comprehension question always referred to the introduction (first three sentences), so did not affect either the behavioural (fourth sentence) or the electrophysiological (fifth sentence) data. T–test comparison on RT confirmed no differences between the sample without incorrect responses and the whole sample, $t(35) = 0.07$, $p = .95$.

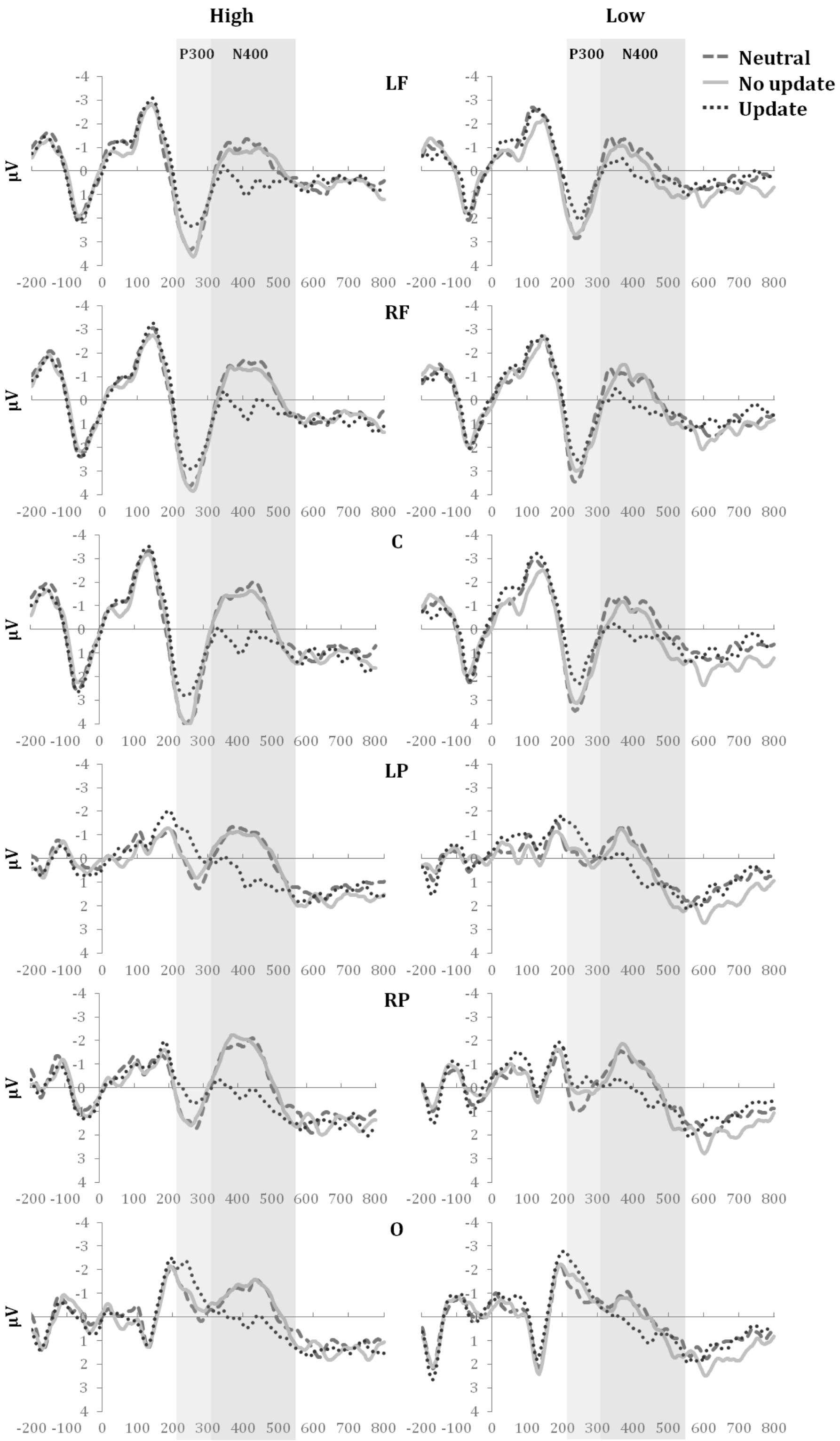

Figure 7. Graphical representation of the electrophysiological activity (in microvolts) found in both working memory groups (High and Low) in each region of interest (LF, RF, C, LP, RP and O). The columns indicate the temporal window of the

components: the pale grey column shows the P300, and the dark grey column the N400. Conditions are also represented by different colors (see legend).

3.3.3. Results

Behavioural analysis: reading times in sentence 4

A mixed model ANOVA with working memory group (high vs. low) and condition (neutral vs. no update vs. update) was performed on RT per sentence. There was only a main effect of condition, $F(2, 68) = 11.27$, $p < .001$, $p\eta^2 = .25$, where the update required longer reading times (3075.75 ms) than the other two conditions: neutral (2800.56 ms), and no update (2714.43 ms). The memory group effect, $F(1, 34) = 1.82$, $p = .19$, and the group x condition interaction, $F(2, 68) = 1.82$, $p = .17$, were not significant (see Table 5 for means). T–test comparisons to confirm the locus of differences in the main effect of condition revealed that the update condition significantly differed from the neutral, $t(35) = 3.01$, $p = .005$, and the no update, $t(35) = 4.21$, $p < .001$, conditions. The comparison between the neutral and the no update condition was not significant, $t(35) = 1.43$, $p = .17$.

Table 5. Interaction between working memory group (High vs Low) and condition (Neutral vs No update vs Update).

		M	***SD***	**Range**
High	Neutral	2675.46	890.95	1534.57–4954.17
	No update	2630.92	763.23	1588.33–4148.77
	Update	2845.11	533.26	2002.83–3836.77
Low	Neutral	2925.66	651.37	1750.97–4406.50
	No update	2797.95	574.52	1684.80–3963.13

Update	3306.39	761.91	1740.60–4647.20

Event–related potential analysis: amplitude in sentence 5

First, in order to see if both subcomponents of the P300 (P3a and P3b) could be distinguished in our data, we carried out a one–way ANOVA on the mean amplitude data in the time window of 220–300 ms, dividing ROI in central–frontal (C, LF and RF) and posterior (LP, RP and O) regions. The analysis showed a significant main effect of location, $F(1, 35) = 173.56$, $p < .001$, $p\eta^2 = .83$, because the central–frontal regions were significantly more positive than the posterior regions. Therefore, we conducted separate analyses for the P3a (C, LF and RF regions) and the P3b (LP, RP and O regions).

P3a analysis. A mixed model ANOVA with working memory group (high vs. low), condition (neutral, update, no update), and the three ROI[5] associated with the P3a (LF, RF, and C) was performed on the mean amplitude data in the time window of 220–300 ms. There was a tendency towards a larger positivity in the high memory group compared to the low memory group, $F(1, 34) = 3.36$, $p = .08$, $p\eta^2 = .09$. The main effect of condition was significant, $F(2, 68) = 3.87$, $p = .03$, $p\eta^2 = .10$, where, as predicted, the amplitude for the disambiguation word following the neutral and no update versions of sentence 4 resulted in larger positivity than that found in the update condition. There was also a main effect of ROI, $F(1.83, 62.28) = 5.83$, $p = .005$, $p\eta^2 = .15$, with larger positivity in the C and RF regions than in the LF region. No interactions were significant (all $p > .35$; see Figure 8a).

P3b analysis. A second mixed model ANOVA with working memory group, condition, and the three ROI related to the P3b (RP, LP, and O) was performed on the mean amplitude of the same temporal window. The main effect of group did not reach significance, $F(1, 34) = 1.00$, $p = .33$. There was a significant effect of condition, $F(2, 68) = 7.42$, $p = .002$, $p\eta^2 = .18$, because as predicted the amplitude in the neutral and no update conditions was more positive than in the update condition. There was also a main effect of ROI, $F(1.88, 64.02) = 72.11$, $p < .001$, $p\eta^2 = .68$, because the two parietal regions (LP and RP) were significantly more positive than the O region. In addition, there was a significant two–way interaction between the working memory group and condition, $F(2, 68) = 3.79$, $p = .03$, $p\eta^2 = .10$, where, as expected, only the high memory group presented

[5] ROI values are based in a Greenhouse–Geisser correction.

a more positive amplitude in the neutral and no update conditions compared to the update condition (see Figure 8b). No other interactions reached significance (all $p > .10$).

To identify the locus of the interaction between working memory group and condition, planned comparisons between conditions were carried out for each memory group separately (with a Bonferroni correction setting the alpha at .008). For the high memory group, significant differences between the update and both the neutral and the no update conditions were apparent, $t(17) = 4.02$, $p < .001$; and $t(17) = 3.13$, $p = .007$, respectively; whereas the neutral and no update conditions did not differ $t(17) = 1.26$, $p = .22$. A different pattern was apparent for the low memory group: none of the contrasts reached significance (all *p*s > .44).

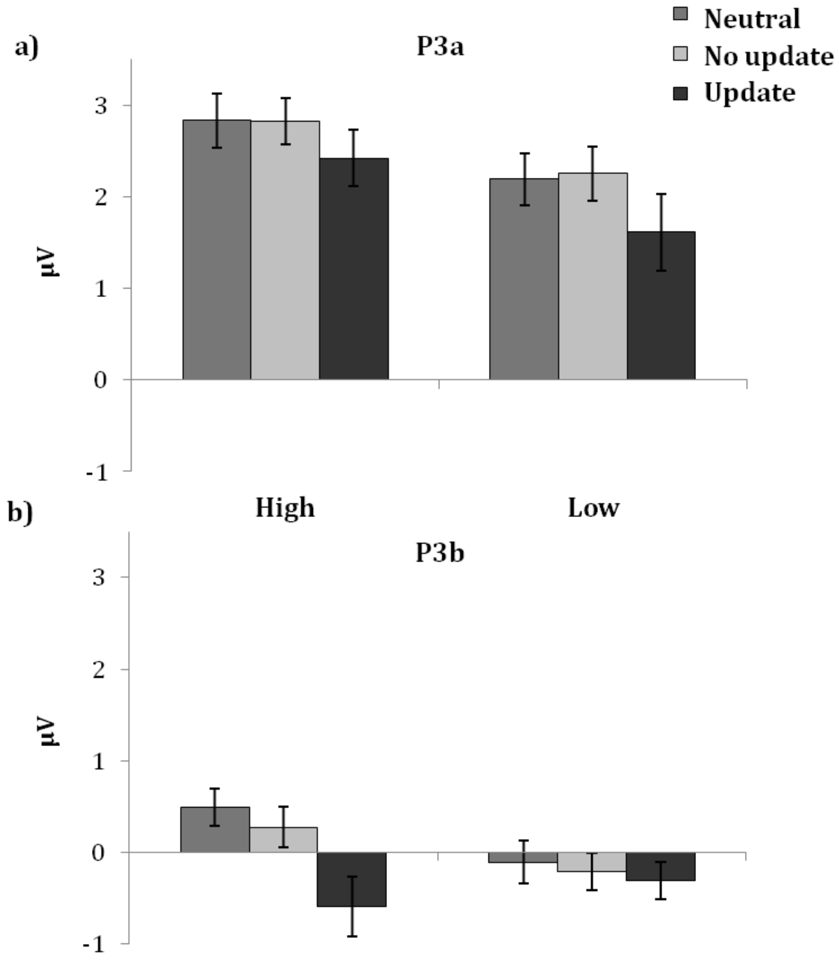

Figure 8. ERP results of the interaction between working memory group (High vs Low) and condition (Neutral, No update and Update) in the P300 component. The panel ‘a’ shows the activation of the P3a subcomponent (central–frontal regions), while the panel ‘b’ shows the brain activity of the P3b subcomponent (posterior regions).

N400 analysis. A third mixed model ANOVA with working memory group, condition and the six ROI was performed on the mean amplitude data in the time window of 300–550 ms. The main effect of group did not reach significance, $F(1, 34) = 0.91$, $p = .35$. There was a main effect of condition, $F(2, 68) = 21.84$, $p < .001$, $p\eta^2 = .39$, because as predicted the amplitude in the neutral and no update conditions was more negative than in the update condition. There was also a tendency toward a main effect of ROI, $F(2.36, 80.10) = 2.65$, $p = .07$, $p\eta^2 = .07$, with less negativity in the LP region. In addition, there were two significant interactions: one between condition and ROI, $F(6.02, 204.77) = 3.94$, $p = .001$, $p\eta^2 = .10$, with the neutral and the no update conditions always being more negative than the update condition, particularly in the RP region, $t(35) = 7.23$, $p < .001$, and $t(35) = 6.60$, $p < .001$, respectively; and the other between group and condition, $F(2, 68) = 3.85$, $p = .03$, $p\eta^2 = .10$, where only the high memory showed more negative amplitude in the no update condition compared to the update condition. (see Figure 9). The two–way interaction between group and ROI and the three–way interaction were not significant, $F(5, 170) = 2.08$, $p = .07$, $p\eta^2 = .06$; and $F(10, 340) = 1.29$, $p = .23$, respectively.

Once more, to identify the locus of the interaction between working memory group and condition, we conducted planned comparisons between conditions for each memory group separately. As before and as predicted, the high memory group showed larger negativity (N400) in the neutral condition, $t(17) = 6.80$, $p < .001$, and the no update condition, $t(17) = 6.02$, $p < .001$, compared to the update condition. Furthermore, the neutral and the no update conditions did not differ, $t(17) = 0.19$, $p = .85$. In contrast, the low memory group showed larger negativity (N400) in the neutral compared with the update condition, $t(17) = 3.44$, $p = .003$, but there was no difference between the no update and the update condition, $t(17) = 1.46$, $p = .16$; nor between the neutral and the no update condition, $t(17) = 1.49$, $p = .15$.

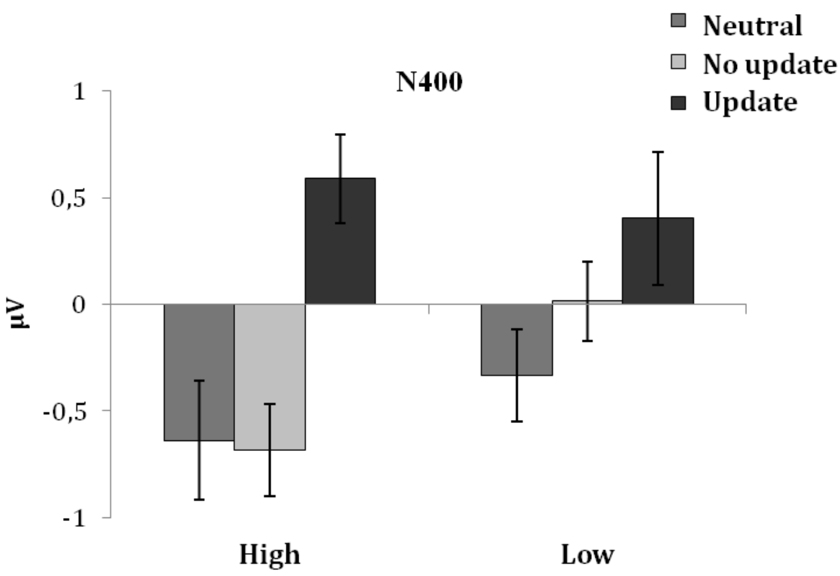

Figure 9. ERP results of the interaction between working memory group (High and Low) and condition (Neutral, No update and Update) in the N400 component.

3.3.4. Discussion

Our goal was to investigate the dynamics of inferential updating in relation to adult readers' working memory capacity, using both behavioural and electrophysiological measures. The behavioural results reflected longer reading times in both working memory groups when they read a sentence that prompted updating in comparison to a neutral and a no update sentence. On the other hand, the electrophysiological results indicated differences between the working memory groups for specific ERP components. The division by subcomponents in the P300 demonstrated that memory groups did not differ in the amplitude of the P3a: both memory groups presented larger positivity in the neutral and no update conditions compared to the update condition. The pattern of findings for the P3b differed by working memory group. The high memory group showed significantly larger positivity in the neutral and no update conditions compared to the update condition. In contrast, the low memory group did not differ between conditions. Similarly, for the N400 component, the high memory group demonstrated larger negativity in the neutral and no update conditions than in the update condition. In contrast, the low memory group did not show a difference between the neutral and the no update condition, although a difference was apparent between the neutral and the update condition.

In our situation model updating task, the first three sentences (introduction) provided a general context that facilitated at least two plausible inferences, one of which was more likely than the other. The fourth sentence was either neutral, in that it did not refer to the critical concept in the story; did not prompt an update, because it was consistent with both concepts supported by the introduction; or it prompted an update to the situation model because the description was consistent with only the less likely concept. In line with our pilot study, we found a large cost effect when participants read sentence 4 in the update condition, compared to the neutral and no update conditions. This effect demonstrated that participants detected an inconsistency between the introduction and the concept supported in the update condition. This cost effect was apparent only in the update condition. The reading times for the no update condition did not differ from the neutral condition, demonstrating that the context of the introduction primed only the most plausible concept. If both concepts were active in memory, participants should have shown no differences in RTs between the update and no update conditions. Further, there were no differences in RTs between the working memory groups: all participants took longer to read the sentence in the update condition than the neutral and the no update conditions. This was congruent with our preliminary study.

The reading time data alone do not identify if both groups engaged in processing additional to the detection of a mismatch, such as successfully updating their situation model and inhibiting the earlier interpretation of the concept. The ERP data recorded for the subsequent disambiguating word speak to that. As we have already pointed out, several components were of interest here. The analysis of the P300 window demonstrated a clear distinction between the P3a (central–frontal activation) and the P3b (posterior activation) subcomponents. Critically, there were no differences between the working memory groups for the P3a, while these differences were apparent for the P3b. According to the context–updating theory (Polich, 2003; 2007) the P3a takes place when an incoming information demands attentional control because is evaluated as ‘new’ or ‘different’ with respect the current memory representation (top–down process); in contrast, the P3b appears when that incoming information forces subsequent attentional resources to favour context updating by memory operations (bottom–up process). Although this theoretical framework has been developed using a traditional attentional task (oddball paradigm), the results found with our situation model updating task fit perfectly within this framework. First, the common pattern of the larger P3a in the no update and neutral conditions for both memory groups indicates that all readers required greater attentional control when the earlier text had not prompted an update. This means that both high and low working memory readers were able to evaluate the disambiguating word as new or different with respect to their memory representation (detection of inconsistency). From our point of view, this is convergent with the reading time data: both memory groups were aware that the concept encoded in their situation model was not appropriate when they read sentence 4 in the update condition. Additionally, according to the context–updating theory, the tendency towards larger positivity in the high memory group suggests that this group required fewer attentional resources than the low group to evaluate the disambiguating word as 'new'. Second, the group differences found for the P3b component indicate that the two groups engaged in different processing when they read sentence 4 in the update condition. The smaller P3b found for the high working memory group in the update condition relative to the other two conditions suggests that this group not only detected that the information in sentence 4 was not consistent with the concept activated in their situation model (as indicated the longer RTs and the P3a), but had successfully updated their situation model when reading this information. As a result, the disambiguating word did not require additional memory processes after the update condition because that information was already incorporated into the mental representation. In contrast, although the low working memory group also showed the detection of inconsistent information with their memory representation (longer RTs and P3a), they appeared not to have updated their situation model when reading sentence 4, because there were no differences in the amplitudes of the P3b for any conditions. The distinction found between the P3a and P3b clearly demonstrates a dissociation between the process to detect information that does not match with the current memory representation (top–down process), and the process to update inferential information by integrating this into the mental representation of the text (bottom–up process). We further explain these results in terms of the generation of the situation model. Since low working memory readers can detect the presentation of new information, we believe that they built a coherent situation model from the introduction of the text. However, because they seem to have problems incorporating that information into their representation, we also believe that they have difficulty constructing a 'precise' and 'accurate' situation model. The N400 component sheds light on this issue.

The analysis of the N400 also reflected a different pattern depending on the working memory group. As we predicted, the high memory group showed larger negativity in the neutral and no update conditions compared to the update. This finding signals their ability to detect a semantic inconsistency between the disambiguating word and their current representation, when an earlier update was not prompted. In addition, it also demonstrates that high working memory readers were able to integrate the prior updating sentence into their situation model of the story. In contrast, the low working memory group manifested a different electrophysiological pattern. As expected, these readers showed no significant difference between the no update and the update condition, indicating that they had not successfully revised their situation model when prompted to do so. However, in contrast to our predictions, they presented a larger negative amplitude in the neutral condition compared to the update condition. From our point of view, different explanations underlie these results.

On the one hand, the lack of difference between the no update and the update condition strongly suggests that the low memory group had difficulty in accurately representing the specific concept in their situation model. Interestingly, both critical concepts (e.g., '*guitar*/*violin*') shared similar semantic properties, which could potentially interfere and make it difficult to integrate the specific concept into an accurate situation model. In accordance with this, some studies have shown the pervasive effect of semantic interference to updating the contents of working memory (e.g., Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). On the other hand, the difference found between the neutral and the update condition suggests that low memory readers were able to detect a semantic inconsistency between the context of the introduction (e.g., idea of '*guitar*') and the final word (e.g., '*violin*').

We have already discarded the possibility that both concepts were activated in memory, since the behavioural results demonstrated a clear cost effect for only the update condition. The neutral condition differed from the other two conditions because it presented information that did not refer to the specific concept. If this difference resulted in a change of story focus in the neutral condition, the inconsistency between the concept supported by the introduction and the disambiguating word would be less apparent because that aspect of the story would not be foregrounded and therefore less activated in memory (Glenberg, Meyer, & Lindem, 1987; Radvansky & Copeland, 2001). This reduced focus or activation may be more apparent for readers with low working memory capacity who have been found to experience difficulties detecting change when across several sentences (Daneman & Carpenter, 1980; Oakhill, Hartt, & Samols, 2005).

Finally, it is important to remember that text comprehension is not an all–or–nothing process. In this sense, the updating of inferential information can vary according to the degree of accuracy with which a situation model is generated. This supports the distinction found here between the P3a and P3b, and accounts for the issue that amplitude differences found between working memory groups are due to differences in magnitude rather than qualitative differences. This is consistent with other studies of adult language processing, where differences between high and low working memory groups are typically quantitative rather than qualitative (e.g., Daneman & Carpenter, 1983).

Conclusions

In sum, this is the first study to report ERP data for inferential updating of readers' situation models. We have identified where in the reading process poor working memory readers have difficulties with situation model updating and also demonstrated distinct electrophysiological patterns that can be used to study this phenomenon further. Our results strongly suggest that readers with low working memory are able to detect coherence breaks when processing text, but that they have difficulties excluding a prior wrong interpretation because they fail to update and integrate new information to ensure an accurate situation model.