

Inferential revision in narrative texts: an ERP study

Ana Pérez¹, Kate Cain², María C. Castellanos¹ and Teresa Bajo¹

Lancaster University

University of Granada (Spain)¹

Lancaster University (United Kingdom)²

Word count: 9,618

Corresponding author:

Ana Isabel Pérez Muñoz

Mind, Brain and Behavioral Research Centre (CIMCYC)

Department of Experimental Psychology, University of Granada.

C/ Profesor Clavera s/n, CIMCYC, 18011, Granada, Spain.

E-mail: antiterrenal@gmail.com

Phone: +34660308256

Abstract

We evaluated the process of inferential revision during text comprehension in adults. Participants with high or low working memory read short texts, in which 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 revise sentence, which referred to a general property consistent with either concept (e.g., *'...beautiful curved body'*); and a revise sentence, which referred to a property that was consistent with only the less likely concept (e.g., *'...matching bow'*). Readers took longer to read the sentence in the revise condition, indicating that they were able to evaluate their comprehension and detect a mismatch. In a final sentence, a target noun referred to the alternative concept supported in the revise condition (e.g., *'violin'*). ERPs indicated that both working memory groups were able to evaluate their comprehension of the text (P3a), but only high working memory readers were able to revise their initial incorrect interpretation (P3b) and integrate the new information (N400) when reading the revise sentence. Low working memory readers had difficulties inhibiting the no longer relevant interpretation and thus failed to revise their situation model, and they experienced problems integrating semantically related information into an accurate memory representation.

Keywords: revising information; inference making; working memory; P3a; P3b; N400.

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, by generating inferences and incorporating world knowledge from long-term memory (Kintsch & van Dijk, 1978). The construction of the situation model is a dynamic process (e.g., Rapp & van den Broek, 2005; McNamara & Magliano, 2009; Kintsch, 1998); the text is processed incrementally (e.g., word by word and sentence by sentence) and, therefore, the situation model is constantly being updated as the text unfolds. As each new piece of information is processed, it must be integrated with the mental representation constructed so far. This involves monitoring for comprehension to identify when and where additional processing, such as inference generation, is necessary to ensure coherence (Kintsch, 1998; Perfetti, Stafura, & Adlof, 2013).

Comprehension monitoring is the metacognitive awareness that readers have about what they are reading (Wagoner, 1983). Baker (1985) distinguishes two monitoring phases: evaluation and regulation. Evaluation, more recently defined as validation (Singer, 2013), refers to the process that allows readers to detect an inconsistency or mismatch in the text (e.g., Vauras, Kinnunen, Salonen, & Lehtinen, 2008). Current evidence on this process in adult readers converges on the view that evaluation is a routine, passive and nonstrategic reading activity that depends on both the activation of current information and the integration of that information with previous text information or world knowledge (Kendeou, 2014). On the other hand, regulation is associated with the repair processes that are necessary to incorporate the new information into the current memory representation (e.g., O'Brien, Rizzella, Albrecht, & Halleran, 1998). According to Hacker (1998), comprehenders self-regulate their reading by asking themselves questions (evaluation) and updating their situation model (e.g., revising inconsistent information). Updating is a type of regulation that includes

a broad category of processes and for coherent texts it is achieved with little cognitive cost. When a mismatch between the current situation model and new information is detected in the evaluation phase, the updating process will involve more than simply integrating just-read information into the situation model. In such instances, readers may *revise* the situation model by modifying or replacing information. Revision of the situation model can only be achieved if readers are able to adequately regulate their comprehension. Hereafter, we refer to regulation as a revision process.

The revision process is clearly a specific updating activity that involves the inhibition of an interpretation that was encoded into the situation model in favour of the new information (Rapp & Kendeou, 2007). Interestingly, both the evaluation of mismatches and the revision of no longer relevant information can occur at an inferential processing level. For example, if the text supports the generation of a specific inference (e.g., ‘*A mouse was looking for something to eat while a bigger animal was waiting to hunt it.*’), which supports the inference of ‘*cat*’), only readers who generate that inference will be able to detect a subsequent mismatch (e.g., ‘*The dog jumped out and scared the mouse.*’¹). Readers who detect that mismatch should then revise their situation model by replacing ‘*cat*’ with ‘*dog*’, to ensure that the situation model is an accurate representation of the text. The need to revise may not always be triggered by an explicitly stated concept (such as ‘*dog*’ in the previous example). A continuation that invites an inference (e.g., ‘*The bigger animal barked loudly...*’) would also require readers to revise the earlier inference that the animal was a ‘*cat*’. We have called this process *inferential revision*, and a deeper understanding of how individual differences in working memory affect this process is the main goal of the present study.

A number of different paradigms show that while readers easily incorporate new information into their situation model (e.g., Rapp & Kendeou, 2007; Rapp & Taylor, 2004),

¹ Importantly, the example provided here does not exactly represent a real inconsistency (the concept ‘*dog*’ is plausible in the context of ‘*bigger animal*’), but an alternative concept that differ from the most probable inference (‘*cat*’). This process has been called ‘*inference alteration*’ (e.g., Radvansky & Copeland, 2004).

they do not always successfully revise their mental representation when new information contradicts previously stated information (e.g., Guéraud, Harmon, & Peracchi, 2005; O'Brien, et al., 1998; Rapp & Kendeou, 2007, 2009). In a classic example, O'Brien et al. (1998) found that participants took longer to read a sentence regarding a person's behaviour when that 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 although participants detected the inconsistency between the character's behaviour and earlier information, they experienced difficulty integrating the new information into their mental representation. 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 situation model ('*Nevertheless, Mary never stuck to her diet when she dined out with her friends*'). If participants had successfully revised their situation model 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.

One reason for a failure to revise the situation model is that readers have problems replacing outdated information (Kendeou, Smith, & O'Brien, 2013; O'Brien, Cook, & Guéraud, 2010). The Knowledge Revision Comprehension framework (KReC; Kendeou & O'Brien, 2014) proposes that once information is retrieved from long-term memory, the activation of the new information competes with the no longer relevant information, drawing activation away from the now-outdated information. Importantly, if the activation of the outdated information is not sufficiently reduced by the competition mechanism, this information may interfere with the new information making difficult the revision of the situation model. Empirical support for this comes from studies of children and adults with poor text comprehension. These studies show a relationship between poor comprehension and difficulties with working memory, a key cognitive resource that supports the general

processes involved in the construction of the situation model. Critically, these working memory difficulties are associated with deficient suppression or inhibitory control over the contents of working memory (Cain, 2006; Carretti, Corndoldi, de Beni, & Palladino, 2004; Pimperton & Nation, 2010). For example, Carretti et al. (2004) found that good comprehenders had higher working memory capacity than poor comprehenders and were also better recalling words in a categorization task: they made fewer intrusion errors (words that had been categorised but were not list final words so should not be recalled). These findings suggest that the relationship between poor reading comprehension and poor working memory may be related to difficulties that poor comprehenders have with inhibiting irrelevant information.

These studies linking poor comprehenders' working memory capacity to problems with inhibition fit also well with Gernsbacher's model of text comprehension: the Structure-Building framework (Gernsbacher, 1990, 1997). This framework proposes that readers with low working memory capacity have problems with text comprehension because of difficulties with suppressing irrelevant information. As a result, they generate new substructures rather than integrating new information into the situation model and therefore produce a less coherent situation model than good comprehenders (e.g., Gernsbacher & Faust, 1991; Gernsbacher, Varner, & Faust, 1990). This body of work provides both empirical and theoretical support for the proposal that readers with good working memory may be better able to revise their situation model than readers with poor working memory, because they are more efficient at inhibiting no longer relevant information, an essential process to construct an accurate and coherent situation model.

Surprisingly, there are very few studies investigating how individual differences in working memory are associated with the process of inferential revision. Indirect evidence for such an association comes from a study by Dutke and von Hecker (2011). They investigated how working memory capacity affects the process of revising the situation model when

ambiguous information is read. They presented narrative texts about the social relations between protagonists to adult readers with high or low working memory span. High-span readers were better able than low-span readers to inhibit an earlier representation (e.g., '*Carlo and Salvatore did not like each other.*') that was incompatible with new information (e.g., '*Franco and Salvatore relied on each other.*'). This suggests that an individual's memory capacity is related to their ability to revise a situation model. However, although Dutke and von Hecker's (2011) study focused on the structure of social relations described in the text to investigate inferential revision, they did not use moment-by-moment processing measures to study revision during the construction of the situation model. Moreover, in a separate study, Dutke and colleagues found evidence indicating that some of the situation model revision took place after reading the text (Dutke, Baadte, Hähnel, von Hecker, & Rinck, 2010), which leaves open the question of whether the revision process may occur when reading the text (that is on-line) under certain circumstances.

From our point of view, it is important to understand the time course and accuracy with which readers revise inferential information in their situation model, to elucidate reading comprehension problems at a high-level of processing. Thus, the present study aims to investigate the inferential revision process in two important ways. First, we developed a paradigm that enabled us to dissociate the two key components of comprehension monitoring described earlier: the detection of a mismatch (evaluation process) and the updating of no longer relevant information (revision process) at an inferential processing level. Second, we examined how individual differences in working memory were associated with the process of revising the situation model. To explore these issues we recorded reading times and electrophysiological brain activity during on-line reading.

Behavioural measure

Reading times. Reading times are an established way to assess processing difficulty. Readers typically take longer to read a sentence if they detect a mismatch between the text

and what they have read previously (O'Brien et al., 1998), and they also take longer to read critical sentences when inferential processing is required (McKoon & Ratcliff, 1992). Our texts contained a sentence that either prompted readers to make a revision or not to the current situation model. Reading times for this sentence were compared across these conditions to determine whether participants had successfully evaluated the new information against the existing situation model. The reading times did not enable us to establish whether or not participants had actually revised their situation model and replaced the now-incorrect inference with the new-correct inference. Because event-related potentials (ERPs) are a robust means to study the precise time course of many cognitive processes, we examined distinct ERP components to a subsequent critical word in the text to investigate the revision process.

ERP measures

P3a and P3b subcomponents. A relevant theoretical framework for the purposes of the present experiment 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 is evident when incoming information is evaluated as new or different with respect to the current representation demanding attentional control; and a temporo-parietal positivity or 'P3b' (e.g., Hartikainen & Knight, 2003), which has been found when the context of the incoming stimulus involves updating by memory processes. Generally, the P3a is assumed to reflect mechanisms of attentional orientation driven by a target or novel stimulus (see Friedman, Cycowicz, & Gaeta, 2001 for review), whereas the P3b is related to processing capacity, being affected by the allocation of cognitive resources, the relevance of the stimulus to the task, and the probability of the stimulus (e.g., Kok, 2001). These findings suggest the existence of a brain circuit encompassing a) a top-down stimulus-driven process that takes place in frontal areas (P3a), and b) a bottom-up memory-driven process, which is guided by updating operations and occurs in parietal areas (P3b; see Polich, 2003). Accordingly, we used the P3a as an additional index of the detection

of mismatches (evaluation process), and the P3b as an index of the updating of no longer relevant information (revision process).

Moreover, some studies have demonstrated a selective relationship between the reduction of the P3b amplitude and poor execution in several capacities such as comprehension monitoring (Getzmann & Falkestein, 2011) and working memory capacity (Evans, Sellinger, & Pollak, 2011). In support of this distinction and of relevance to the current study is work by Getzmann and Falkestein (2011). They compared younger (19–25 years) and older (54–64 years) adults' performance on a comprehension monitoring task in which participants had to respond according to the stock price of a specific company, while ignoring other prices and beep sounds. Interestingly, participants did not show significant differences in the behavioural results (accuracy and reading times), but electrophysiological differences emerged. Specifically, the older adults manifested an increased right-frontal P3a (only in high-performing older adults) and a reduced parietal P3b relative to the younger adults. The authors interpreted the P3a result as an age-related compensatory mechanism and the P3b result as an effect of age-related decline in spoken language comprehension. These data signal that the P300 subcomponents may reflect individual differences in language comprehension.

N400 component. Another ERP of interest 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). The amplitude of the N400 is reduced when there is a good fit between the word being processed and the context, in comparison to a poorly fitting word. For example, Kuperberg, Paczynski, and Ditman (2011) demonstrated a N400 for words causally unrelated to an inference supported by the text (e.g., '*Jill's skin always tanned well. She always put on sunscreen. She had sunburn on Monday.*') compared to causally related words (e.g., '*Jill had very fair skin. She forgot to put sunscreen on. She had sunburn on Monday.*'). However, although there have been several electrophysiological studies

demonstrating a relation between working memory and inference making (e.g., St George, Mannes, & Hoffman, 1997) and in evaluation of coherence breaks (e.g., Virtue, Haberman, Clancy, Parrish, & Jung-Beeman, 2006), no study to date has investigated the relationship between working memory and the revision of the situation model using ERPs. Therefore, a second aim of the present study was to explore if these ERP components (P3a, P3b and N400) reflected individual differences in working memory associated with the construction of the situation model.

The current study

To address our aims we developed the ‘situation model revision task’ (see Table 1). Participants read short texts in which sentences 1–3 provided an introduction for which at least two different concepts could be inferred, by means of the generation of knowledge-based elaborative inferences (McKoon & Ratcliff, 1980). Both concepts were plausible, but one was considered more likely (e.g., ‘*guitar*’) by independent judges (see below). There were three versions of the subsequent sentence 4: a neutral condition, which did not refer directly or indirectly to either concept; a no revise condition, which mentioned a property consistent with either concept (e.g., ‘...*beautiful curved body*’); and a revise condition, which referred to a property that was consistent with only the less likely concept (e.g., ‘...*matching bow*’). This latter condition should prompt readers to revise the situation model to ensure good comprehension. Reading times were measured for this sentence. The final word in sentence 5 was always inconsistent with the concept supported in the introduction but consistent with the concept supported in the revise condition (e.g., ‘*violin*’). This word was called the disambiguating word and ERPs were recorded here.

Table 1. Example of text used in the situation model revision 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 <i>guitar</i>)
The concert takes place at the prestigious national concert hall. His instrument is made of maple wood, with a <i>beautiful curved body</i> . His instrument is made of maple wood, with a <i>matching bow</i> .	Neutral No revise Revise
The public was delighted to hear Dan playing the violin .	ERP sentence
In the recital, Dan played his favourite works.	Comprehension sentence

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

Our predictions were as follows. First, in relation to the behavioural data, if readers generate and encode the inference supported by the introduction (*'guitar'*), then they will show longer reading times for sentence 4 in the revise condition (*'matching bow'*) compared to the neutral (*'national concert hall'*) and no revise (*'beautiful curved body'*) conditions. This effect would signal the ability to detect a mismatch (evaluation process) when new information does not match the current situation model (e.g., Bohn-Gettler, Rapp, Van den Broek, Kendeou, & White, 2011). As noted, longer reading times do not enable us to establish if readers are able to replace the incorrect inference with the alternative inference when prompted by sentence 4 in the revise condition. The ERP data, registered for the disambiguating word of sentence 5, help us to understand whether readers not only detect a mismatch when reading sentence 4 in the revise condition, but also whether they successfully revise their situation model, as detailed below.

In line with Polich (2003, 2007), the P3a subcomponent shows if a word is evaluated as new or different with respect to the current mental representation. Thus, if readers activate the alternative inference when reading sentence 4 in the revise condition, then they will exhibit a reduction of the P3a to the disambiguating word (*'violin'*) in the revise condition (*'matching bow'*) compared with the neutral (*'national concert hall'*) and no revise (*'beautiful*

curved body') conditions. This pattern would signal not only the activation of the new inference when reading sentence 4 in the revise condition, but also the mismatch detection for sentence 5 in the neutral and no revise conditions. The P3b subcomponent indicates if a word prompts a revision of the situation model. Therefore, similar to the P3a, if readers are able to update the alternative inference and draw activation away from the previous incorrect inference when reading sentence 4 in the revise condition, they will show a reduction of the P3b to the disambiguating word in the revise condition compared with the neutral and no revise conditions. This effect would demonstrate that readers not only activate the new inference when reading sentence 4 in the revise condition, but are also able to revise their mental representation. Finally, the N400 component is an index of the ease with which information can be integrated into a reader's situation model. If readers integrate the alternative inference when reading sentence 4 in the revise condition, then they will demonstrate a reduction of the N400 to the disambiguating word in the revise condition compared with the neutral and no revise conditions. This result would indicate that readers are able to integrate the new inference into their situation model when reading sentence 4 in the revise condition.

Furthermore, if readers with low working memory are less able to evaluate their comprehension and revise their situation model than readers with high working memory, they will not manifest significant differences between conditions in the disambiguating word compared with high working memory readers who will. However, because to our knowledge this is the first ERP study investigating individual differences in the inferential revision process, we did not make specific hypotheses about the electrophysiological components for high and low working memory readers. Importantly, our paradigm enables us to combine behavioural and electrophysiological data to understand better how working memory relates to a reader's ability to evaluate and inferentially revise their situation model.

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 payment. 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), only participants with extreme working memory scores (see below) were invited to complete the situation model revision 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. The number of words in the lists increased from 2 to 12 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, according to their physical size. The number of words to be recalled was stated before each list and increased from 1 to 5, with a fixed presentation order. Participants were required to a) activate and maintain each new word in working memory to compare its size with previously presented words, b) maintain activation of the smallest objects in the specified set size, and c) inhibit any previously activated words that no longer meet the criteria (that is to inhibit a large-size object when they heard the name of a smaller object). Therefore, the recall set of words had to be constantly revised as new words were presented. All participants completed all trials.

Reading span task. We used a version of 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 to 6 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 score for both memory tasks was the total number of words correctly recalled minus the total number of words incorrectly recalled (intrusions). These scores were used to classify participants into the high and low working memory groups, using the criterion of being above or below (respectively) the mean of the total score in both working memory tasks (see scores below).

Situation model revision task. We constructed 93 (3 practice, 90 experimental) five-sentence narrative texts, some modified from texts used by Lorscheid, Katz, and Cupak (1998). An example is shown in Table 1 (see full set of materials in Appendix A). The first three sentences supported a specific inference to be made (e.g., ‘*guitar*’). There were three versions of sentence 4: the neutral condition, which did not refer back to either the supported or the alternative inference and, therefore, was neither consistent nor inconsistent with the introduction; the no revise condition, which was consistent with the inference primed by the introduction; and the revise condition, which prompted readers to revise their situation model so that only the alternative inference was encoded, rather than the inference supported by the introduction. Reading times were the dependent variable for this sentence. Sentence 5 concluded with a disambiguating word (e.g., ‘*violin*’), which was always incongruent with the inference supported by the introduction and congruent with the inference supported by sentence 4 in the revise condition. Consequently, the disambiguating word was unexpected in the neutral and no revise conditions, and expected in the revise condition. At the end of the text, a comprehension sentence requiring a true or false judgment was included to encourage participants to read for meaning.

A norming study provided empirical confirmation of concept preferences in our situation model revision task. Twenty-two participants ($M = 22.7$ years old; range: 18–55) read the introduction of each text (sentences 1–3) and were then presented with a single word.

Their task was to decide (yes/no) if the word fitted with the sense of the story. The word was either the target concept, which was most strongly supported by the introduction (e.g., ‘guitar’), the alternative concept (e.g., ‘violin’), or a nonstory concept (e.g., ‘poker’). Results of a one-way ANOVA performed on the percentage of accuracy showed a main effect of concept type: $F(2,42) = 92.92, p < .001, p\eta^2 = .82$, because participants were more likely to correctly accept the target concept ($M = 83.95, SD = 7.71$) and to correctly reject the nonstory concept ($M = 88.95, SD = 7.24$) than to accept the alternative concept ($M = 50.43, SD = 12.43$). Further, when participants did accept the alternative concept, they took longer to do so ($M = 2079$ ms) compared with response times to the target concept ($M = 1612$ ms): $t(21) = 3.72, p < .001^2$. This difference suggests that, after reading the introduction, the target concept was significantly more likely to be activated than the alternative concept, as intended. It is important to acknowledge that there was variability in the extent to which our 90 texts constrained the activation of the target concept in the introduction. A second norming study with a two-alternative forced choice task confirmed that the two critical concepts were both supported by sentence 4 (e.g., ‘guitar’ for the no revise, and ‘violin’ for the revise). Fourteen participants ($M = 20.9$ years old; range: 18–26) read the introduction, followed by one of the two versions of 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 sample used in the norming studies did not take part in the main study.

The word frequency for each of the two critical concepts was examined using the Word Frequency Guide database (Zeno, Ivens, Millard, & Duvvuri, 1995) and did not differ ($M_s = 56.58$, and 47.18 , for the no revise and revise concepts respectively: $t(89) = 0.27, p = .79$). The word length of sentence 4 did not differ between conditions ($M_s = 11.70, 11.46$, and

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 concept in the response time analysis.

11.81, for the neutral, no revise and revise conditions respectively: $F(2, 178) = 1.74, p = .18$). Finally, although we tried to minimise non-manipulated differences between our conditions, the structure for sentence 4 varied across conditions.

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 on a computer screen 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: 18 low and 18 high working memory readers. The mean number of words recalled for the low 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; and for the high memory 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. T-tests confirmed significant group differences in both memory tasks: 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 revision task. This session took approximately 90 minutes and included only participants with low and high working memory. First, 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. Sentences 1–4 were presented one sentence at a time, and participants were instructed to read each sentence at their own pace, pressing the space bar to display the next sentence. The reading time of sentence 4 (neutral vs. no revise vs. revise) was recorded. Immediately after, sentence 5 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 disambiguating word to ensure that the electrophysiological activities of the ERPs were registered. To prevent excessive noise in the electrophysiological data, we asked participants to try not to blink during the presentation of sentence 5. Finally, participants were presented with a true/false comprehension sentence. This sentence always referred to information in the introduction (equally distributed across sentences 1–3).

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 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 revision task, scalp voltages were recorded from a SynAmps2 64 channels Quik-Cap, plugged into 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 5k Ω . 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 mastoids. Six regions of interest (ROI) were extracted from the 64 channels (see Figure 1), 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).

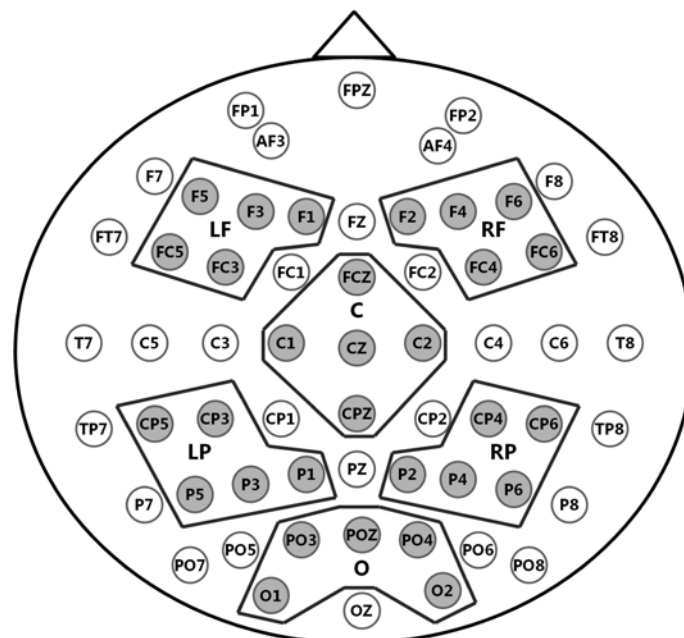


Figure 1. The six regions of interest (ROI): left frontal (LF); right frontal (RF); central (C); left parietal (LP); right parietal (RP); and occipital (O).

³ Because the magnitude of the components involved in the inferential revision 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 36 participants for all trials⁴. Working memory group was a between-subjects factor in all analyses. The behavioural analysis of the situation model revision task was conducted on reading times (milliseconds) per sentence. In the ERP analyses, the critical time-windows were predefined by visual inspection. In this way, 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 2). Outlier amplitude data per condition, 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 Comprehension sentences always referred to sentences 1–3, so did not affect either sentence 4 (reading times) or the disambiguating word (ERPs). T-test comparison on reading times confirmed no differences between the sample without incorrect responses and the whole sample, $t(35) = 0.07$, $p = .95$.

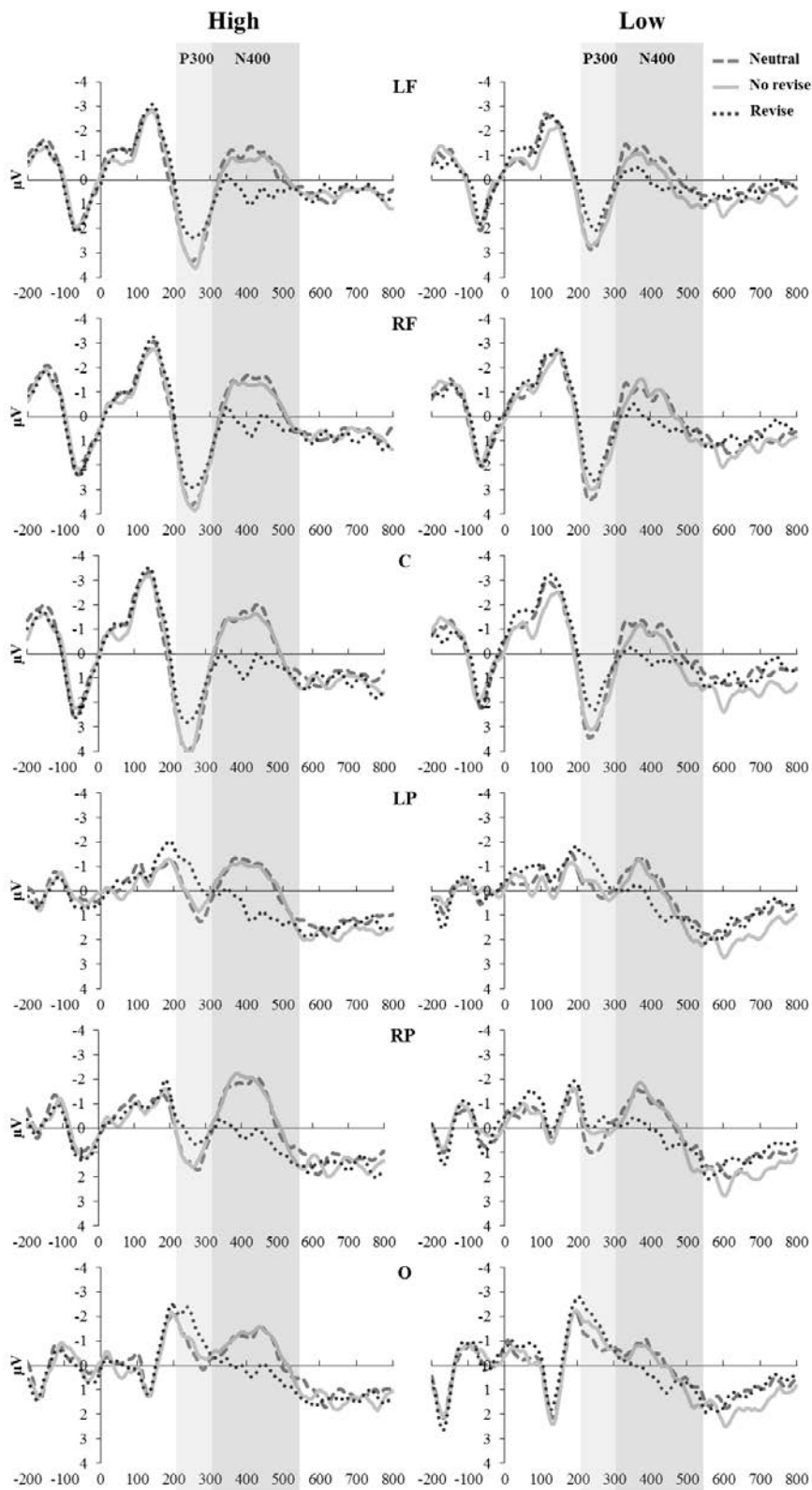


Figure 2. Graphical representation of the mean amplitude (in microvolts) for the P300 (pale grey column) and the N400 (dark grey column) components, divided by working memory group, condition and region of interest.

Results

Behavioural analysis (sentence 4)

Reading times. To understand if readers generated the inference in the introduction and then, evaluated their comprehension by detecting a mismatch in the revise condition, we performed a mixed model ANOVA with working memory group (high vs. low) and condition (neutral vs. no revise vs. revise) on the reading time of sentence 4. There was only a main effect of condition, $F(2, 68) = 11.27, p < .001, \eta^2 = .25$, where the revise condition resulted in longer reading times ($M = 3076$ ms) than the other two conditions: neutral ($M = 2801$ ms), and no revise ($M = 2714$ ms). T-tests revealed that the revise condition significantly differed from the neutral, $t(35) = 3.01, p = .005$, and the no revise, $t(35) = 4.21, p < .001$, conditions. The comparison between the neutral and the no revise condition was not significant, $t(35) = 1.43, p = .17$. The memory group effect, $F(1, 34) = 1.82, p = .19$, and the memory group x condition interaction, $F(2, 68) = 1.82, p = .17$, were not significant (see Table 2 for means)⁵.

Table 2. Reading time means for sentence 4 of the situation model revision task, divided by working memory group and condition.

	Condition	<i>M</i>	<i>SD</i>	Range
High	Neutral	2675	891	1535–4954
	No revise	2631	763	1588–4149
	Revise	2845	533	2003–3837
Low	Neutral	2926	651	1751–4406
	No revise	2798	575	1685–3963
	Revise	3306	762	1741–4647

ERP analysis (disambiguating word)

⁵ To investigate whether the behavioural result was significant across texts, we conducted a linear mixed effect model with reading times, using condition as a fixed effect and participants and texts as random effects. In addition, to control the variability caused by the number of words, we also included sentence length as a random slope in the text factor. The result demonstrated the same significant main effect of condition found in the ANOVA, $F(2) = 52.54, p < .001$. The p value and df of the fixed effect were extracted using the `anova` function of the `lmerTest` R package, version 2.0–11 (Kuznetsova, Brockhoff, Christensen, 2012).

First, in order to see if both subcomponents of the P300 (P3a and P3b) could be distinguished in our data, we carried out a mixed model ANOVA with working memory group, condition and ROI 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 tendency towards a larger positivity in the high memory group compared to the low memory group, $F(1, 34) = 3.19, p = .08, p\eta^2 = .09$. There was a significant main effect of condition, $F(2, 68) = 6.17, p < .01, p\eta^2 = .15$, with more positive amplitude in the neutral and no revise conditions compared to the revise condition. There was also a main effect of ROI, $F(1, 34) = 174.67, p < .001, p\eta^2 = .84$, because the central-frontal regions were significantly more positive than the posterior regions. Critically, although no two-way interaction reached significance (all $p > .28$), the three-way interaction was significant, $F(2, 68) = 7.26, p < .01, p\eta^2 = .18$. Therefore, we conducted separate analyses for the P3a (C, LF and RF) and the P3b (LP, RP and O) subcomponents.

P3a analysis. Our aim was to see if readers generated the alternative inference on reading sentence 4 in the revise condition and then, evaluated the disambiguating word as already activated. To do this, we performed a mixed model ANOVA with working memory group, condition, and the three ROI⁶ associated with the P3a (LF, RF, and C) on the mean amplitude data (for the disambiguating word) in the time window of 220–300 ms. As before, 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 disambiguating word following the neutral and no revise conditions of sentence 4 was larger ($M = 2.51, SD = 1.25$ and $M = 2.54, SD = 1.17$, respectively) than that found in the revise condition ($M = 2.02, SD = 1.58$). There was also a main effect of ROI, $F(2, 68) = 5.83, p =$

⁶ ROI p -values are based in a Greenhouse–Geisser correction.

.006, $p\eta^2 = .15$, with larger positivity in the C and RF regions than in the LF region. No interactions were significant (all $p > .35$).

P3b analysis. To determine if readers revised their mental representation reducing activation from the previous incorrect inference on reading sentence 4 in the revise condition, we performed a third mixed model ANOVA with working memory group, condition, and the three ROI related to the P3b (RP, LP, and O) on the mean amplitude data (for the disambiguating word) in the same temporal window. The main effect of memory 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 revise conditions was more positive than in the revise condition. There was also a main effect of ROI, $F(2, 68) = 72.11, p < .001, p\eta^2 = .68$, 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 memory group and condition, $F(2, 68) = 3.79, p = .03, p\eta^2 = .10$. No other interactions reached significance (all $ps > .10$).

To identify the locus of the interaction between memory group and condition (see Figure 3), planned comparisons between conditions were carried out for each group separately, with a Bonferroni correction setting the alpha at .008. For the high memory group, significant differences between the revise condition and both the neutral and the no revise conditions were apparent: $t(17) = 4.02, p < .001$, and $t(17) = 3.13, p = .007$, respectively; whereas the neutral and no revise 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 $ps > .44$).

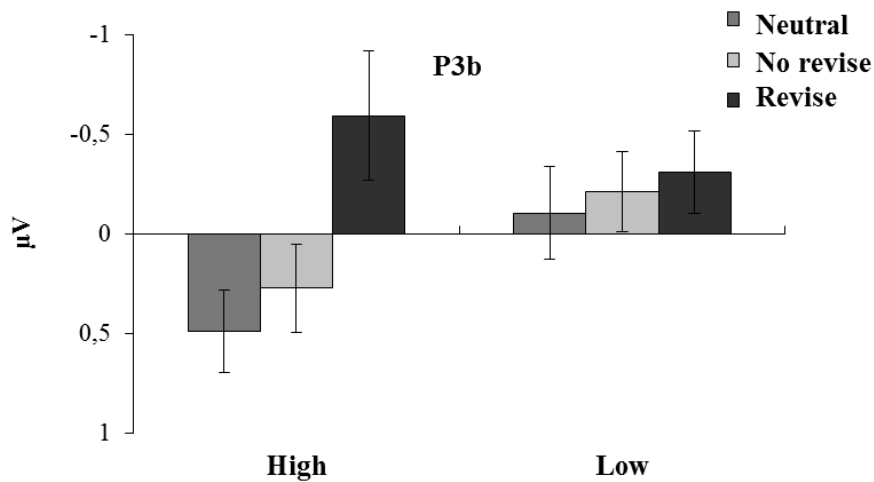


Figure 3. Mean amplitude of the P3b subcomponent for the disambiguating word of the situation model revision task, divided by working memory group and condition.

N400 analysis. To see if readers integrated the alternative inference on reading sentence 4 in the revise condition, we performed a final mixed model ANOVA with working memory group, condition and the six ROI on the mean amplitude data (for the disambiguating word) in the time window of 300–550 ms. The main effect of memory 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 revise conditions was more negative than in the revise condition. There was also a tendency toward a main effect of ROI, $F(5, 170) = 2.65, p = .07, p\eta^2 = .07$, with less negativity in the LP region. In addition, there were two significant interactions. The first two-way interaction between condition and ROI, $F(10, 340) = 3.94, p = .001, p\eta^2 = .10$, arose because the neutral and the no revise conditions were more negative than the revise condition, particularly in the RP region, $t(35) = 7.23, p < .001$, and $t(35) = 6.60, p < .001$, respectively. The second two-way interaction between memory group and condition, $F(2, 68) = 3.85, p = .03, p\eta^2 = .10^7$, is of

⁷ To explore whether the effect of working memory in the P3b and N400 was linear, we conducted separate linear mixed effect models for each component, using working memory (continuous scores) and condition as fixed effects, and participants and texts as random effects. These models shown the same significant two-way interaction between working memory and condition as found in the ANOVAs for both the P3b, $F(3) = 5.67, p = .001$, and the N400, $F(3) = 3.79, p = .01$. However, because we previously selected participants with high or low

specific interest to our understanding of the integration process and is discussed in detail below. A further two-way interaction between memory group and ROI showed a tendency toward significance, $F(5, 170) = 2.08, p = .07, p\eta^2 = .06$, with larger negativity for the high compared to the low working memory group in the posterior regions (LP, RP and O). Finally, the three-way interaction was not significant, $F(10, 340) = 1.29, p = .23$.

The two-way interaction between memory group and condition (see Figure 4) was explored further to understand the integration process. Planned comparisons between conditions for each memory group separately were used to identify the locus of this interaction, again with a Bonferroni correction setting the alpha at .008. Only the high memory group showed more negative amplitude in the no revise condition compared to the revise condition. Specifically, this group showed larger negativity in the neutral condition, $t(17) = 6.80, p < .001$, and the no revise condition, $t(17) = 6.02, p < .001$, compared to the revise condition. The neutral and the no revise conditions did not differ, $t(17) = 0.19, p = .85$. In contrast, the low memory group showed larger negativity in the neutral compared with the revise condition, $t(17) = 3.44, p = .003$, but there was no difference between the no revise and the revise condition, $t(17) = 1.46, p = .16$; nor between the neutral and the no revise condition, $t(17) = 1.49, p = .15$.

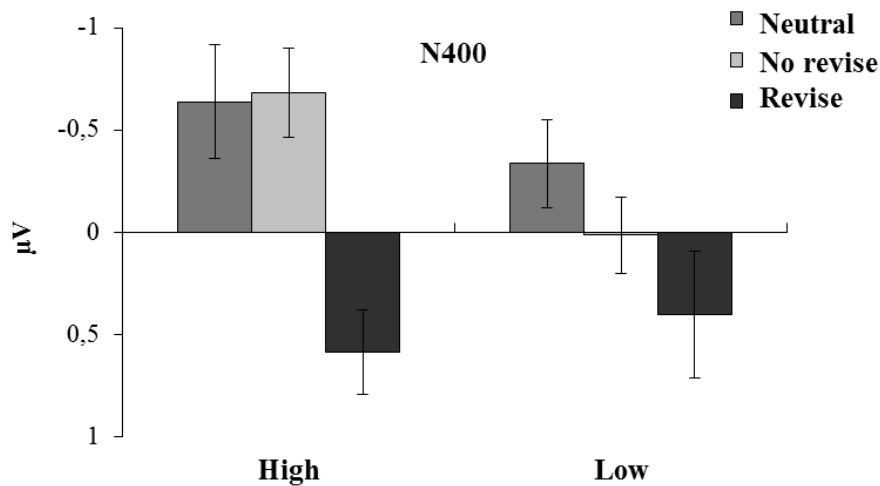


Figure 4. Mean amplitude of the N400 component for the disambiguating word of the situation model revision task, divided by working memory group and condition.

Discussion

The goal of this study was to investigate the dynamics of inferential revision in adults' reading comprehension, using both behavioral and electrophysiological measures. To do so, we created a bespoke reading comprehension paradigm: the situation model revision task. In this, the introduction (sentences 1–3) provided a general context that facilitated at least two plausible inferences, one of which was more likely than the other. Sentence 4 was either neutral, did not require a revision (no revise), because it was inferentially consistent with the most likely concept, or did prompt a revision (revise), because the description prompted an inference that was consistent with only the less likely concept. Our behavioral results indicated that all participants took longer to read this sentence in the revise compared to the neutral and no revise conditions.

The final sentence ended with the disambiguating word, which was always inconsistent with the most likely concept, but consistent with less likely concept supported in the revise condition. Here, our electrophysiological results differed by the specific ERP component. There were no working memory differences in the amplitude of the P3a: both

memory groups presented larger positivity in the neutral and no revise conditions compared to the revise condition. In contrast, 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 revise conditions compared to the revise condition, while the low memory group did not differ between conditions. Similarly, there were working memory differences in the N400 component: the high memory group demonstrated larger negativity in the neutral and no revise conditions than in the revise condition; however, the low memory group did not show a difference between the no revise and the revise condition, although a difference was apparent between the neutral and the revise condition.

Evaluation at the inferential level

In our texts, sentence 4 of the revise condition always mismatched the interpretation supported by the introduction. The question was then whether readers were able to detect mismatches with their current situation model even though this information was processed at the inferential level. The reading time results demonstrated a large cost for both working memory groups, suggesting that all readers detected a mismatch between the new inferable information and the inference that was supported by the introduction of the text. In addition, the difference in reading times found between the neutral and revise conditions confirmed that the initial interpretation (e.g., *'guitar'*) was inferred by readers and incorporated into the situation model. Therefore, our behavioural results signal that, when a highly constrained semantic context is provided, both high and low working memory readers are equally able to infer a knowledge-based elaborative inference and subsequently detect inferential information that is incompatible with that elaboration. This finding indicates that the evaluation process of monitoring may occur at the inferential level, which is congruent with other studies demonstrating that adults are able to evaluate their inferential comprehension (e.g., Poynor & Morris, 2003). In addition, it is also consistent with the minimalist hypothesis (McKoon & Ratcliff, 1980), which has claimed that elaborative inferences are automatically encoded

during reading when 1) information is quickly and easily available in memory, or 2) they are necessary to provide coherence by text information or prior knowledge. Nevertheless, as we previously mentioned, our sentences were constructed so that the mismatch was easily detected. Thus, it is possible that working memory differences could arise if more subtle mismatches were introduced in these sentences.

A less clear matter is whether readers successfully revised their situation model after reading the revise condition in sentence 4. From our point of view, two things could be happening here. One possibility is that readers activated and encoded the alternative interpretation (e.g., '*violin*') and reduced activation of the initial interpretation. This would reflect revision of the memory representation. Alternatively, they may have activated and encoded the alternative interpretation without reducing activation of the initial interpretation. This would reflect a lack of revision of the memory representation. The reading times by themselves only speak to the evaluation process of comprehension monitoring, and do not clarify if the revision process took place when reading sentence 4. The electrophysiological data recorded in the disambiguating word address this critical issue.

Evaluation and revision processes: P3a and P3b

According to the context-updating theory (Polich, 2003, 2007), the P3a occurs when incoming information demands attentional control because it is evaluated as 'new' or 'different' with respect to the current memory representation; in contrast, the P3b appears when that incoming information forces subsequent attentional resources to favour context updating by memory operations. Although this theoretical framework has been developed using a traditional attentional task (oddball paradigm), our situation model revision task produced results that are consistent with this framework.

First, we found larger positivity associated with the P3a in the neutral and no revise conditions compared to the revise condition. This effect indicates that readers required greater attentional control on reading the disambiguating word (e.g., '*violin*') when earlier

information had not prompted a revision (e.g., *'beautiful curved body'* - no revise; *'national concert hall'* - neutral), but not when it had prompted one (e.g., *'matching bow'* - revise). Thus, in general, readers were able to detect a difference or mismatch between the new (disambiguating) word and their current situation model. Interestingly, the lower positivity found in the revise condition also signaled that readers activated the alternative interpretation when reading sentence 4. Second, we also found larger positivity associated with the P3b in the neutral and no revise conditions compared to the revise condition. However, the P3b effect was qualified by an interaction with group. Critically, the high memory group manifested smaller positivity related to the P3b for the revise condition relative to the other two conditions, while the low memory group did not show significant differences between conditions. These working memory differences found for the P3b suggest that the two groups had engaged in different processing when they read sentence 4 in the revise condition.

On the one hand, the P3a findings indicate that both high and low working memory readers perceived the disambiguating word as 'new' when prior text information had not prompted a revision. That is, all readers detected the mismatch between that word and their current situation model. Of note, this is convergent with the reading time data: all readers detected the mismatch between sentence 4 and the introduction in the revise condition. Thus, both the behavioural and P3a results suggest that adult readers are able to evaluate their comprehension during reading. On the other hand, the P3b findings indicated that high and low working memory readers differed in their way to revise the situation model. The smaller positivity for the revise condition relative to the other two conditions manifested by the high memory group signaled that this group did not require additional memory processes to update their situation model on reading the disambiguating word, because they had already revised their situation model on reading sentence 4. Therefore, the high memory group had not only evaluated their comprehension detecting a mismatch (as indicated the longer reading times and the P3a) on reading sentence 4, but also had revised their situation model, updating the

final interpretation and significantly reducing activation of the previous interpretation. In contrast, the low memory group did not show significant differences between conditions for the P3b. This lack of differences indicated that the low memory group had not successfully revised their situation model on reading sentence 4 in the revise condition because they had difficulties drawing activation away from the initial interpretation.

This interpretation is congruent with studies demonstrating that poor comprehenders with poor working memory capacity have problems in inhibiting irrelevant information (Cain, 2006; Carretti, et al., 2004; Pimperton & Nation, 2010). It is also consistent with the Structure-Building framework (Gernsbacher, 1990, 1997), which argues that low working memory readers may experience problems with comprehension, because they fail to suppress no longer relevant information due to the generation of new substructures that reduce coherence of the situation model (e.g., Gernsbacher & Faust, 1991). In relation to this, a more specific proposal in the field of revision suggests that the information that is no longer relevant or outdated may exert an influence disrupting comprehension (e.g., Kendeou, et al., 2013; O'Brien, et al., 2010; Rapp & Kendeou, 2007, 2009). Taking all this evidence into account, our P3b results indicate that, in contrast to high working memory readers, low working memory readers have problems revising their situation model because they fail to inhibit the initial wrong interpretation. In addition, the presence of inferred information could make the revision of the situation model more difficult for low working memory readers, who may construct a more 'imprecise' or 'inaccurate' mental representation of the story. This is consistent with the literature of inference alteration, which suggests that memory processes are involved when an inference that has been previously activated must be replaced by a new one (e.g., Radvansky & Copeland, 2004). The process of integrating text information into a coherent situation model sheds light on this issue.

Integration process: N400

Similar to the P3b, the analysis of the N400 demonstrated that working memory capacity underpinned the ability to integrate information into the situation model. The high memory group showed larger negativity in the neutral and no revise conditions compared to the revise condition. This result indicated that they experienced difficulties with integrating the disambiguating word into their situation model when earlier text information had not required a revision; in contrast, they did not experience difficulties when a revision had been prompted by the text. Thus, high working memory readers were able to integrate the alternative interpretation into their situation model on reading sentence 4 in the revise condition (supported by the reading time results). A different pattern was evidence for the low memory group, broadly similar to that found for the P3b: they showed no significant difference between the no revise and the revise condition, although larger negativity was found in the neutral condition compared to the revise condition.

The lack of difference between the no revise and the revise condition for the low working memory group, strongly suggests that they had not successfully integrated the alternative interpretation into their memory representation on reading sentence 4 in the revise condition. Both critical concepts (e.g., '*guitar/violin*') shared similar semantic properties, which could potentially interfere and disrupt the construction of an accurate situation model. Moreover, the difference found between the neutral and the revise condition suggests that the low working memory group was able to integrate the alternative concept reading the disambiguating word only when the neutral condition was presented in sentence 4, because the nature of the neutral information (not related with the critical concept) did not cause semantic interference. These results are congruent with those studies showing the pervasive effect of semantic interference in relation to updating the contents of working memory (e.g., Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). Therefore, low working memory readers seem to be able to accurately integrate new information into the situation model under

some circumstances, but they experience difficulties if that information semantically interferes with other information that is already encoded.

A compatible hypothesis that cannot be ruled out in this study is the possibility that the activation of the N400, and therefore the level of integration into the situation model, could be related to the degree of awareness with which readers detected the unexpected information. Unfortunately, we did not systematically ask participants if they had noticed anything ‘odd’ in the texts. However, there was a tendency for some participants to report this. Future research should include checks for awareness of inconsistencies in the debriefing and analyses of these to determine if differences in awareness exist between memory groups.

A comprehensive view

Our findings can be understood within Kendeou and O’Brien’s (2014) KReC framework. This proposes five key principles that are required for knowledge revision within a situation model during reading: encoding, passive activation, co-activation, integration, and competing activation. In relation to our paradigm, it means that once the previous interpretation (e.g., ‘*guitar*’) is encoded and passively activated from long-term memory, the presentation of information supporting the revised interpretation (e.g., ‘*matching bow*’) will cause the co-activation of both interpretations (e.g., ‘*guitar/violin*’). Furthermore, the revise information of sentence 4 will lead to the integration of the alternative interpretation (e.g., ‘*violin*’) within the situation model, drawing activation away from the previous but now-incorrect interpretation and reducing the interference between the two (competing activation).

Applying the KReC logic to our results, the evaluation of mismatched information (prompted by sentence 4), requires both the activation and integration of the two interpretations, whereas the revision of the situation model (prompted by the disambiguating word) involves the better integration and increase in activation of the new concept. Therefore, on the one hand, a failure in the co-activation of both interpretations and in the integration of new information will result in problems in evaluation, because no incompatibility will be

detected. On the other hand, a failure to integrate, together with the competing activation of the final concept, will cause difficulties in revision, because interference between both interpretations means that activation will spread to the now-incorrect interpretation, hampering stronger activation of the new concept. Our data suggest that low working memory readers have problems revising their situation model, because the activation of the previous interpretation continues to compete (by semantic interference) with the new concept. That is, these readers have problems strengthening the activation of the final concept because they fail to inhibit the wrong interpretation. We believe that this is a promising framework for the future study of the inferential revision process.

Finally, to better understand our results it is important to consider the processes that were evaluated in the two working memory tasks and how they relate to the reading comprehension processes studied here. The reading span task was used as an index of the ability to actively maintain information, making it readily retrievable (Daneman & Carpenter, 1980), while the memory updating task was selected to measure the suppression of no longer relevant information (e.g., Carretti, Belacchi, & Cornoldi, 2010). Our situation model revision task tapped both of these aspects of working memory: it required the activation and maintenance of the previous interpretation, as well as the inhibition of that interpretation when a more plausible inference was apparent. This viewpoint is not incompatible with Engle's (2002) perspective that working memory reflects the ability to allocate attentional resources when the task involves interference control to maintain or suppress information. We believe that the ability to actively maintain and suppress no longer relevant information, combined with the need to control interference, are both crucial in explaining how individual differences in working memory are related to the inferential revision process.

Conclusions

To our knowledge this is the first study to report ERP data associated with the inferential revision of readers' situation models. We have proposed that the context-updating

theory is a promising framework to dissociate the comprehension monitoring processes of evaluation and revision. Through this, we have identified where in the reading process readers with poor working memory have difficulties revising their situation model. Our ERP data provide evidence that strongly suggest that low working memory readers are able to evaluate their comprehension and detect coherence breaks in the text (evidenced by the P3a), but that they have difficulties revising their memory representation because they fail to inhibit an initial wrong interpretation (P3b) and integrate new information to ensure an accurate situation model (N400).

Acknowledgements

This study has been supported by a grant of the Spanish Ministry of Education and Science (FPU) to A.P., by grants of MINECO–PSI2012-33625 and JA-2012–CTS 2369 (Proyecto Excelencia) to T.B., and by the Department of Psychology of Lancaster University (UK). In addition, we would like to thank Thomas Lorschach for providing his materials, Kate Nation and Holly Joseph for their contribution in the construction of some of the additional materials, and Sasha Smith for help with our norming studies.

References

- Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. *Metacognition, Cognition, and Human Performance, 1*, 155-205. Academic Press, New York, NY.
- Bohn-Gettler, C. M., Rapp, D. N., Van den Broek, P., Kendeou, P., & White, M. J. (2011). Adults' and children's monitoring of story events in the service of comprehension. *Memory & Cognition, 39*, 992-1011. doi:10.3758/s13421-011-0085-0
- Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory, 14*, 553-569. doi:10.1080/09658210600624481
- Carretti, B., Belacchi, C., & Cornoldi, C. (2010). Difficulties in working memory updating in individuals with intellectual disability. *Journal of Intellectual Disability Research, 54*, 337-345. doi:10.1111/j.1365-2788.2010.01267.x
- Carretti, B., Cornoldi, C., De Beni, R., & Palladino, P. (2004). What happens to information to be suppressed in working-memory tasks? Short and long term effects. *The Quarterly Journal of Experimental Psychology Section A, 57*, 1059-1084. doi:10.1080/02724980343000684
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466. Retrieved from <http://search.proquest.com/docview/616528436?accountid=14542>
- Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*, 309-321. doi:10.1016/j.cogbrainres.2004.09.006
- Dutke, S., Baadte, C., Hähnel, A., von Hecker, U., & Rinck, M. (2010). Using diagnostic text information to constrain situation models. *Discourse Processes, 47*, 510-544.

doi:10.1080/01638530903416257

- Dutke, S., & Von Hecker, U. (2011). Comprehending ambiguous texts: A high reading span helps to constrain the situation model. *Journal of Cognitive Psychology, 23*, 227–242. doi:10.1080/20445911.2011.485127
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. doi:10.1111/1467-8721.00160
- Evans, J. L., Selinger, C., & Pollak, S. D. (2011). P300 as a measure of processing capacity in auditory and visual domains in specific language impairment. *Brain Research, 1389*, 93-102. doi:10.1016/j.brainres.2011.02.010
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25*, 355-373. doi:10.1016/S0149-7634(01)00019-7
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ. Retrieved from <http://search.proquest.com/docview/617869883?accountid=14542>
- Gernsbacher, M. A. (1997). Attenuating interference during comprehension: The role of suppression. *Psychology of Learning and Motivation, 37*, 85-104. Academic Press, San Diego, CA. Retrieved from <http://search.proquest.com/docview/620012574?accountid=14542>
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 245-262. doi:10.1037/0278-7393.17.2.245
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 430-445. doi:10.1037/0278-7393.16.3.430
- Getzmann, S., & Falkenstein, M. (2011). Understanding of spoken language under

- challenging listening conditions in younger and older listeners: a combined behavioral and electrophysiological study. *Brain Research*, 1415, 8-22.
doi:10.1016/j.brainres.2011.08.001
- Guéraud, S., Harmon, M. E., & Peracchi, K. A. (2005). Updating situation models: The memory-based contribution. *Discourse Processes*, 39, 243-263.
doi:10.1207/s15326950dp3902&3_8
- Hacker, D. J. (1998). Self-regulated comprehension during normal reading. *Metacognition in Educational Theory and Practice*, 165-191. Lawrence Erlbaum Associates Publishers, Mahwah, NJ. Retrieved from
<http://search.proquest.com/docview/619360628?accountid=14542>
- Hartikainen, K. M., & Knight, R. T. (2003). Lateral and orbital prefrontal cortex contributions to attention. *Detection of Change*, 99-116. Kluwer Academic Publishers, Dordrecht. Retrieved from <http://search.proquest.com/docview/619975371?accountid=14542>
- Kendeou, P. (2014). Validation and Comprehension: An Integrated Overview. *Discourse Processes*, 51, 189-200. doi:10.1080/0163853X.2013.855874
- Kendeou, P., & O'Brien, E. J. (2014). *The Knowledge Revision Components (KReC) framework: Processes and mechanisms*. In D. N. Rapp & J. Braasch (Eds.), *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences*, Cambridge, MA: MIT Press.
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 854-865. doi:10.1037/a0029468
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press, New York, NY. Retrieved from
<http://search.proquest.com/docview/619359467?accountid=14542>
- Kintsch, W., & Dijk, T. A. (1978). Toward a model of comprehension and production.

- Psychological Review*, 85, 363-394. doi:10.1037/0033-295X.85.5.363
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 31, 557-577. doi:10.1017/S0048577201990559
- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23, 1230-1246. doi:10.1162/jocn.2010.21452
- Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia*, 4, 7790.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647. doi:10.1146/annurev.psych.093008.131123
- Kuznetsova, A., Christensen, R. H. B., & Brockhoff, P. B. (2012). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, 1-0.
- Lorsbach, T. C., Katz, G. A., & Cupak, A. J. (1998). Developmental differences in the ability to inhibit the initial misinterpretation of garden path passages. *Journal of Experimental Child Psychology*, 71, 275-296. doi:10.1006/jecp.1998.2462
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668-682. Retrieved from <http://search.proquest.com/docview/616539054?accountid=14542>
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466. doi:10.1037/0033-295X.99.3.440
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, 51, 297-384. doi:10.1016/S0079-7421(09)51009-2
- O'Brien, E. J., Cook, A. E., & Guéraud, S. (2010). Accessibility of outdated information.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 979.

doi:10.1037/a0019763

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: a memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1200-1210. doi:10.1037/0278-7393.24.5.1200

Perfetti, C. A., Stafura, J. Z., & Adlof, S. M. (2013). Reading comprehension and reading comprehension problems: a word-to-text integration perspective. In B. Miller, L. E. Cutting & P. McCardle (Eds.). *Unravelling reading comprehension: behavioral, neurobiological, and genetic components*, 22-32. Baltimore: Paul Brookes Publishing.

Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: Evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*, 62, 380-391. doi:10.1016/j.jml.2010.02.005

Polich, J. (2003). Theoretical overview of P3a and P3b. *Detection of Change*, 83-98. Kluwer Academic Publishers, Dordrecht. Retrieved from <http://search.proquest.com/docview/619964755?accountid=14542>

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128-2148. doi:10.1016/j.clinph.2007.04.019

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128-2148. doi:10.1016/j.clinph.2007.04.019

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 3-9. doi:10.1037/0278-7393.29.1.3

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology*, 117, 191-213. doi:10.2307/4149022

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, 35, 2019-2032. doi:10.3758/BF03192934

Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold.

Discourse Processes, 46, 1–24. doi:10.1080/01638530802629141

- Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 988-1001. doi: 10.1037/0278-7393.30.5.988
- Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension an integrative view of reading. *Current Directions in Psychological Science*, 14, 276–279.
doi:10.1111/j.0963-7214.2005.00380.x
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime User's Guide, version Pittsburgh: Psychology Software Tools Inc.
- Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science*, 22, 361–366. Retrieved from
<http://search.proquest.com/docview/1442383666?accountid=14542>
- St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9, 776–787.
Retrieved from <http://search.proquest.com/docview/619171213?accountid=14542>
- Szmaliec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 137. doi:10.1037/a0020365
- Vauras, M., Kinnunen, R., Salonen, P., & Lehtinen, E. (2008). Emotions in comprehension monitoring and regulation. *International Journal of Psychology*, 43, 17.
- Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain Research*, 1084, 104-114.
doi:10.1016/j.brainres.2006.02.053
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18, 328-346. Retrieved from
<http://search.proquest.com/docview/616749135?accountid=14542>

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). The educator's word frequency guide/Touchstone Applied Science Associates. *Inc, Brewster, NY.*