Evaluation and revision of inferential comprehension in narrative texts: An eye movement study

Pérez, A.[1], Joseph, H. [2], Bajo, T. [1], & Nation, K. [2]

University of Granada (Spain)[1]

University of Oxford (UK)[2]

Acknowledgements

This research was supported by the doctoral research grant from the Spanish Ministry of Education and Science (FPU AP2008–01893) to Ana Pérez, and by a grant from the Economic and Social Research Council (RES-000-22-4652) to Holly Joseph and Kate Nation.

Abstract

We investigated how adult readers evaluate and revise their situation model, by monitoring their eye movements as they read narrative texts and critical sentences. In each text, a short introduction primed an inference, followed by a concept that was either expected (e.g. “oven”) or unexpected (e.g. “grill”). Eye movements showed that readers detected a mismatch between the unexpected information and their prior interpretation, confirming their ability to evaluate inferential information. Subsequently, a critical sentence included a word that was either congruent (e.g. “roasted”) or incongruent (e.g. “barbecued”) with the expected but not the unexpected concept. Readers spent less time reading the congruent than the incongruent word, reflecting the facilitation of prior information. In addition, when the unexpected concept had been presented, participants with lower verbal (but not visuospatial) working memory span exhibited longer reading times and made more regressions on encountering congruent information, indicating difficulty in revising their situation model.

Keywords: comprehension monitoring; revision; inference alteration; verbal working memory; linear mixed models.

1. Introduction

In order to comprehend a text, readers integrate the relevant information in the passage with prior knowledge, building up a coherent and accurate mental representation of the text typically known as a situation model (Kintsch & van Dijk, 1978). The construction of a situation model entails the involvement of distinct high–level cognitive processes such as inference making (e.g., Cain & Oakhill, 1999), comprehension monitoring (e.g., Bohn–Gettler, Rapp, van den Broek, Kendeou, & White, 2011), and/or updating information (e.g., Rapp & Kendeou, 2007).

Comprehension monitoring refers to the metacognitive processes by which comprehenders supervise and evaluate their own understanding of a text. According to Wagoner (1983), '*it is an executive function, essential for competent reading, which directs the reader's cognitive processes as he/she strives to make sense of incoming textual information.*' Therefore, the ability to monitor comprehension is essential to track new information presented in a text as well as any changes that force a restructuring of the situation model. The classic paradigm to study comprehension monitoring has been the contradiction or inconsistency detection task (e.g., Albrecht & O'Brien, 1993; Huitema, Dopkins, Klin, Myers, 1993; O'Brien, Rizella, Albrech, & Halleran, 1998; Orrantia, Múñez & Tarín, 2013). In this paradigm readers are presented with information regarding a specific character followed by an action that the character performs that may be consistent or inconsistent with the previous information. The typical effect is a time cost in the inconsistent compared to the consistent sentence, suggesting an increase in information processing when the coherence of the situation model is disrupted.

Interestingly, comprehension monitoring can be linked to inference making. Inference making is one of the most fascinating and indispensable abilities in reading comprehension. Its main function is to provide coherence by joining together text information with a reader's prior knowledge. Thus, beyond all taxonomies and names that have been proposed (see Graesser, Singer, & Trabasso, 1994; and also Graesser & Zwaan, 1995 for review), inferences are the principal engine to establish consistency in text processing. Consequently, if the reader encounters information that is inconsistent with any inferred concept, there should be a cost to the ease of the reading process. Evidence consistent with this comes from a study using the contradiction paradigm while measuring readers' eye movements. Poynor and Morris (2003, Experiment 2) presented a character's goal that was explicit (e.g., '*[Dick] wanted to go somewhere warm and sunny*') or implicit (e.g., '*[Dick] had always been a real sun–worshipper*'), followed by an action that was consistent (e.g., '*[He] ... asked for a plane ticket to Florida*') or inconsistent (e.g., *[He] ... asked for a plane ticket to Alaska*') with the character's goal. They found longer reading times (first and second pass times) in the inconsistent than in the consistent condition regardless of whether the goal had been explicitly or implicitly stated in the text, and more frequent rereading (regressions in) of the goal information in the inconsistent condition when the goal has been explicitly mentioned. In addition, Poynor and Morris (2003, Experiment 1) also evaluated the recall of the passages using an off–line recall test and they found better recall (for both goal and action) in the inconsistent than in the consistent passages. The better recall of the inconsistent passages was interpreted as 1) the successful reinstatement of the prior unsatisfied goal (explicit: '*somewhere warm and sunny*', or implicit: '*a real sun–worshipper*') and its integration into the memory representation, and 2) the strengthening of the inconsistent information because the reinstatement of the prior goal drew attention to the relationship between the goal and the action. From our point of view, these findings are interesting since they indicate readers' ability to monitor inferential information. However, there are several reasons why it is difficult to state that the new information was incorporated into the mental representation at the moment of the action. First, longer reading times in the inconsistent information per se do not necessarily reflect integration, but a disruption in comprehension because incoming information does not fit with the current representation. Second, immediately after reading each passage, readers were presented with a comprehension question (e.g., '*Did Dick originally plan to go to a cold climate?*') that was focused on the resolution of the inconsistency. Arguably, this might bias the subsequent recall. Finally, readers with more than 15% of incorrect responses to the comprehension questions (14% of participants) were excluded from the analysis of the recall data, and no explanation was provided for such a noticeable amount of incorrect responses. Hence, although these results support the hypothesis that comprehension monitoring includes the monitoring of information not explicitly present in the text, it is not clear if readers integrate the new inconsistent information into their situation model while reading.

In addition to monitoring, when readers detect an inconsistency between text information and their inferred mental representation, they are forced to revise the current memory representation, replacing it with the newly uncovered information. This process is known as updating and has been extensively studied in connection with situation models (e.g., Albrech & O'Brien, 1993; O'Brien, et al., 1998; Rapp & Kendeou, 2007; de Vega, 1995; Zwaan & Madden, 2004; Radvansky & Copeland, 2010). Accordingly, it has been observed that readers commonly experience difficulty in updating the new contradictory information, because the prior encoded information continues to interfere with comprehension (e.g., Guèraud, Harmon, & Peracchi, 2005; O'Brien et al., 1998; van Oostendorp & Bonnebakker, 1999). Several models of comprehension have tried to provide an explanation of why this occurs. For example, the Structure–Building model (Gernsbacher, 1990, 1997) proposes that when inconsistent information cannot be integrated into the mental representation, readers try to suppress the information that is not longer relevant. However, when readers are unable to suppress the irrelevant information, they may form new substructures constructed out of the main mental representation. After several substructures are established, the accessibility of information in memory is reduced and readers thus fail to update their situation model (e.g., Gernsbacher & Faust, 1991; Gernsbacher, Varner & Faust, 1990). Therefore, monitoring and updating are closely related processes that are critical for comprehension. However, very few studies have studied their interrelation (e.g., van der Schoot, Reijntjes, & van Lieshout, 2012). In addition, studies investigating the updating of situation model have usually used a sentence–by–sentence procedure and thus, they have only evaluated reading times and/or comprehension questions in an off–line way (e.g., Albrecht & O'Brien, 1993; Rapp & Kendeou, 2007; O'Brien et al., 1998; Zwaan & Madden, 2004). Since monitoring and updating should occur on–line as the arguments in text unfold, it is important to use on–line measures to capture the interrelation between these processes. Therefore, the first goal of the present work was to study the cognitive processes of comprehension monitoring and updating information during on–line reading.

Monitoring eye movements is an ecologically valid technique to study reading comprehension since 1) it allows the reader to read text at their own pace without the need for any secondary task; 2) it provides information about the time course of text processing on–line, as reading happens; and 3) it also provides information about reading behavior during and after reading of a comprehension question. For these reasons, we used eye movements to measure inferential monitoring and updating of a situation model on–line, as participants read text. Specifically, we presented a prior context which primed an inferential concept and subsequent explicit information that either confirmed or disconfirmed the inferred concept. This provided a manipulation of inferential monitoring. In addition, below the main text a critical sentence was presented that was either congruent or incongruent with the explicit concept introduced in the main text. This sentence allowed us to examine if readers had incorporated the new concept into their situation model. Eye movements were recorded as participants read both the main text and the subsequent critical sentence.

An additional goal of the present study was to explore the role of working memory in monitoring and updating. The findings that some people show significantly more incorrect responses after monitoring inconsistencies (Poynor & Morris, 2003) and that poor readers have difficulty suppressing irrelevant information (e.g., Gernsbacher & Faust, 1991) suggest that there are individual differences in the way that readers update information. Consequently, we also aimed to explore whether individual differences in working memory capacity are associated with this process.

Working memory capacity has been related to high–level language skills such as listening or reading comprehension (e.g., Daneman & Merikle, 1996). The relationship between reading comprehension and working memory has been typically found in complex span measures such as the reading and listening span tasks (Daneman & Carpenter, 1980), where readers are required to recall verbal information (e.g., digits or words), while completing an additional task (e.g., comprehending sentences). In general, these studies conclude that readers with low working memory capacity are less able to maintain and process text information, having difficulties integrating it with prior knowledge into a coherent situation model (e.g., Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Hannon & Daneman, 2001). Subsequent studies have demonstrated the influence that working memory exerts in high–level processes such as inference making (e.g., George, Mannes, & Hoffman, 1997; Virtue, van den Broek, & Linderholm, 2006). In relation to comprehension monitoring, although several studies have found a relationship between working memory and the ability to detect inconsistencies in children (e.g., Cain, Oakhill, & Bryant, 2004; Oakhill, Hartt & Samols, 2005), the relationship between working memory and comprehension monitoring in young adults has shown inconclusive results (e.g., Bohn–Gettler, et al., 2011; Daneman & Carpenter, 1983; De Beni, Borella, & Carretti, 2007). Furthermore, it is still unclear whether the process of updating information into the situation model depends on working memory, and if so, if it is a domain–specific mechanism. Individual differences in reading comprehension are more closely associated with verbal working memory (Carretti, Borella, Cornoldi, & De Beni, 2009; Nation, Adams, Bowyer–Crane, & Snowling, 1999; Pimperton & Nation, 2010; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). Therefore, an additional goal of this study was to investigate if updating was specifically associated with verbal and not visuospatial working memory.

The current study

We recorded eye movements during natural reading to investigate how readers monitor and update their inferential comprehension. Our paradigm was as follows (see Table 2; see Appendix C for full materials, p. 187). A brief introduction of two–sentences (e.g., '*...the last two players were concentrating hard at each end of the table... taking a long time to make a decision*') primed a specific concept at the situation model level. The third sentence introduced a target concept that could be expected (e.g., '*chess*') or unexpected (e.g., '*snap*') according to the prior information. Just below, a critical sentence contained target words that could be congruent (e.g., '*moving pieces*') with the expected but not with the unexpected concept, or incongruent (e.g., '*playing cards*') with the expected but not with the unexpected concept.

Table 2. Example of a text used in the inferential monitoring task.

At the end of the tournament, the last two players were concentrating hard at each end of the table. Both of them were taking a long time to make a decision. After an hour and a half the *chess/snap* game was finished.	**Text**
- The players were *moving pieces* to win the tournament. - The players were *playing cards* to win the tournament.	**Sentence**

Note. Participants read either the expected ('*chess*') or unexpected ('*snap*') concept in the main text, and the congruent or incongruent words with the prior concept in the critical sentence.

Our general proposal is that if readers properly generate the inference as they read the first two sentences, they will exhibit longer reading times on the unexpected concept (e.g., '*snap*') than expected concept (e.g., '*chess*') and/or more regressions out of this concept to previous parts of the text. This result would reflect their ability to monitor information that does not fit with the prior inferred situation model. Subsequently, when reading the critical sentence, if only expected information has been encountered, readers should benefit from their already activated memory representation (e.g., '*chess*') showing shorter reading times, and/or fewer regressions out of the congruent (e.g., '*moving pieces*') compared to the incongruent (e.g., '*playing cards*') target words. This effect would reflect a general ability to construct a coherent situation model, since no change has occurred in the story. Correspondingly, the target words of the critical sentence will require longer reading times and/or a larger number of regressions when the unexpected concept is congruent (e.g., '*playing cards*' after '*snap*') than when the expected concept is congruent (e.g., '*moving pieces*' after '*chess*'). This processing cost would point to the need to suppress the concept introduced at the beginning of the text, updating the new unexpected concept into the situation model.

We also examined whether the effect of inferential monitoring and updating processes were related to readers' verbal (measured by the listening recall and backward digits recall tasks) and/or visuospatial (measured by the odd one out and spatial recall tasks) working memory capacity. Although we do not yet know if young adults monitor inconsistencies during normal reading (e.g., Bohn–Gettler, et al., 2011; De Beni, et al., 2007), we predicted no working memory differences in the inferential monitoring process because the inconsistency was very evident in the text. In contrast, we predicted more efficiency in the updating of higher compared to lower verbal span readers, with few or no influence of the visuospatial domain. Specifically, we expected higher verbal span readers to exhibit shorter reading times and/or fewer regressions than lower verbal span readers on the target words in the critical sentence. Finally, regarding accuracy, if the presentation of the unexpected concept exerts an influence on the product of comprehension, we should find a smaller number of correct responses to the critical sentence when the unexpected compared to the expected concept has been presented. Working memory differences were not predicted in accuracy, because readers had the opportunity to check text information before answering to the critical sentence. To our knowledge, this is the first study assessing the process of updating the situation model during on–line reading. To do this, we measured eye–movements during reading of both the main text and the critical sentence. In addition, we wanted to see whether updating process was associated with verbal and not visuospatial working memory differences.

2 Method

Participants

Forty people (mostly undergraduates but also postgraduate students) with a mean age of 21.9 years old (range: 17–47 years[1]) were recruited by an internet advertisement placed in the webpage of the University of Oxford (UK). All were native English speakers with no known reading disabilities and normal or corrected to normal vision. They participated for either course credits or money (£10).

Materials

Our experiment was the inferential monitoring task, where eye movements were measured as participants read a main text and a critical sentence. In addition, participants completed four working memory span tasks.

Inferential monitoring task. We constructed 34 (4 practice, 30 experimental) four–sentence narrative texts (see Table 2). The first two sentences biased an inference

[1] Most of participants had an age close to 20 years old. Only two participants differed from the mean group of age (32 and 47 years old), which explains the large age range.

generated at the situation model of the story. The third sentence presented one of two conditions: a) in the expected condition appeared the concept primed by the introduction (e.g., '*After an hour and a half the chess game was finished*'), which was consistent with the previous information; b) in the unexpected condition was presented a valid but unlikely concept (e.g., '*After an hour and a half the snap game was finished*'), which was inconsistent with the previous information and forced participants to change their mental representation. A preliminary pilot study was carried out in order to test the two target concepts (e.g., *chess*/*snap*). Participants read each text in one of the two versions of the third sentence. They were instructed to mark from 1 (worst) to 5 (best) how well one of the two target concepts fitted with the ideas in the text. Seven participants completed each version of each text. In the final study, only texts for which the expected condition (M = 4.23) had a better fitting than the unexpected condition (M = 2.21) were included. A t–test comparison confirmed the difference, $t(29) = 15.80$, $p < .001$. Then, the word frequency of the two target concepts was matched using the SUBTLEX database (Brysbaert & New, 2009): M expected = 34.73; M unexpected = 36.24, $t(30) = 0.14$, p = .89. In addition, the number of characters of each word was also controlled between conditions, Ms = 5.58, and 5.65 for the expected and unexpected concepts respectively: $t(30) = 0.30$, $p = .76$. Only an empty line (gap) separated the main text from the critical sentence presented below, which were presented at the same time. Participants were instructed not to read the critical sentence before they had read the main text. This sentence presented one of two possibilities: a) in the congruent condition the information was related to the target concept read in the main text (e.g., after reading '*snap*', they encountered '*The players were playing cards to win the tournament*' as the critical sentence); or b) in the incongruent condition the information was always unrelated to the target concept read in the main text (e.g., after '*snap*', they encountered '*The players were moving pieces to win the tournament*' as the critical sentence). Participants were instructed to press one of two buttons to answer '*Yes*', if they thought the critical sentence was correct, or '*No*' if they thought it was incorrect. Once more, we used SUBTLEX to control the word frequency of the target words (e.g., *playing cards*/*moving pieces*) presented in the critical sentence: M congruent = 256.17; M incongruent = 437.13, $t(44)$ = 1.52, $p = .14$. The number of characters did not differ between conditions, Ms = 5.78, and 5.96 for the congruent and incongruent target words respectively: $t(44) = 0.69$, p = .50.

Working memory measures

Working memory capacity was measured by standardized Automated Working Memory Assessment battery (AWMA; Alloway, 2007). Four span tasks were used: two verbal and two visuospatial. Each task was administered according to the manual instructions with difficulty increasing progressively over blocks by the number of items to be remembered. Participants continued with the next block if they recalled 4 out of the 6 trials. In contrast, when participants failed at least 3 trials of the same block, the task finished. The scores of each working memory task were the total number of trials correctly recall. Errors were not included in the final score.

(i) *Listening recall*. Participants listened to sets of spoken sentences presented one–by–one and were instructed to verify if each sentence was 'true' or 'false'. In addition, at the end of each set of sentences, they were required to recall the last word of each sentence in the order of presentation.

(ii) *Backward digits recall*. Immediately after the presentation of a spoken list of digits, participants had to recall the sequence in the reverse order. There were six levels increasing in difficulty from 2 (Level 1) to 7 (Level 6) digits.

(iii) *Odd one out*. Participants were presented with a three square matrix containing a shape in each space, and they had to point at the figure that was different to the other two. At the end of each set, participants were also required to indicate on the screen the location of each shape in order of presentation.

(iv) *Spatial recall*. Two shapes were presented at the same time. The shape of the right side could be rotated and contained a red dot that changed position (over three compass points). Participants must judge if both shapes were the 'same' when they followed the same direction or 'opposite' when they had a different direction. At the end of the set, a figure with the three compass points appeared on the screen and participants had to recall the location of the dot in order of presentation.

Since we were more interested in the working memory domain rather than in the specific tasks, we used the average of the standardized scores distinguishing between the verbal domain (average of the listening recall and backward digits recall) and the visuospatial domain, (average of the odd one out and spatial recall). The standardized scores were extracted from the AWMA battery.

Apparatus

Eye movements were monitored using an Eyelink 1000 (SR Research; Mississauga, Canada) eye–tracker. The sampling rate was 1000 Hz. A chinrest and forehead rest were used to minimise head movements and to maintain a constant viewing distance of approximately 60 cm. Viewing was binocular but only the right eye was tracked during the experiment. A nine–point calibration procedure was performed to ensure that tracking accuracy was within 1º of visual angle. Re–calibration was carried out between trials as needed. The stimuli were presented on a 19" CRT video monitor (refresh rate = 75 Hz), using the Eyetrack software[7], and the extraction of eye movements measures were carried out using EyeDoctor and EyeDry[2]. The eye movements were: *gaze duration*, the total duration of all fixations in a region before leaving it from left or right side; *regressions out*, the probability of making a leftward eye movement from a region to read previously encountered text; *go–past time*, the sum of all fixations from the first entering a region from the left to exiting it from the right, including re–reading of previous parts of the text; *regressions in*, the probability of making a leftward eye movement into a specific region; and *total time*, the total duration of all fixations in a region, including first and second–past times. The administration of the working memory tasks (AWMA program) was via a 15" laptop computer screen.

Procedure

The eye movement experiment (inferential monitoring task) was completed first, taking approximately 30 minutes. Participants triggered the onset of each trial by fixating a box on the left of the screen. Both the main text and the critical sentence appeared and readers read at their own pace, starting with the text. The information disappeared from the screen when participants pressed the designated true or false key to respond to the critical sentence. Each of 30 experimental trials was presented to each participant only once in one of the four cross conditions (expected–congruent, unexpected–congruent, expected–incongruent, or unexpected–incongruent) counterbalanced across participants. The same number of participants completed each condition, and the presentation of trials was randomized. Four practice trials presented at the beginning of the experiment ensured that instructions were understood, and a small break (about 1 min.) halfway through the task prevented fatigue. Following the experiment, the four working memory tasks were presented in the following order: listening recall, backward digits span, odd one out and spatial recall. In all of them instructions appeared as a sound file with a blank screen, followed by the practice trials. In the experimental trials, responses were recorded discreetly by the experimenter using the right arrow key on the keyboard (⇨) for a correct response and the left arrow key (⇦) for an incorrect response.

Data analysis

[2] Taken from http://www.psych.umass.edu/eyelab/software/

We constructed linear mixed models using the lmer function of the lme4 R package, version 1.0–5 (Bates, Maechler, Bolker, & Walker, 2013). These models are very powerful since they account for both fixed and random effects, allowing the analysis of participant and item at the same time. Separate models were run for each dependent variable (gaze duration, regressions out, go–past time, regressions in and total time) across both regions of our inferential monitoring task (main text and critical sentence). All data were checked to ensure that no participant read the critical sentence before the main text. In addition, accuracy of response to the critical sentence was also analyzed. Participants and items were the random factors of the model.

Expectation (expected vs unexpected) and Congruence (congruent vs incongruent) were always fixed factors. In addition, one of the two domains of working memory was also included as a fixed factor: Verbal working memory, M = 101.77 (SD = 12.61; range = 81–128); or Visuospatial working memory, M = 105.06 (SD = 13.27; range = 77–133). To improve interpretability, the verbal or visuospatial factors were centred in order to understand the average (or intercept) of each factor (e.g., Schielzeth, 2010). Thus, in both cases, the fixed structure was composed by a three–way interaction (e.g., expectation x congruence x verbal working memory). In order to establish the optimal structure for the random and fixed components, we followed a well–known procedure in the field of ecology (see Zuur, Ieno, Walker, Saveliev, & Smith, 2009). First, keeping the full fixed structure, we looked for the best random structure using restricted maximum likelihood (REML). We assumed different random intercepts since both of them could have a different baseline, and we found the justified–by–the–design optimal random slopes using model comparison (see Appendix D, p. 193; see also Barr, Levy, Scheepers, & Tily, 2013 for a review). Second, keeping the already known random structure, we found the best fixed structure using stepwise model comparison from the most complex model (the three–way interaction) to the simplest (a main effect) model, and selecting the one with lower AIC and BIC, and significant χ^2 test for the Log–likelihood, using the maximum likelihood (ML). Finally, for those models with significant fixed effects, the p values were provided by the anova function of the lmerTest R package, version 2.0–3 (Kuznetsova, Brockhoff, Christensen, 2012), using the REML. To assess the overall goodness of fit we calculated the explained deviance by the pamer.fnc function of the LMERConvenienceFunctions R package, version 2.5 (Tremblay & Ransijn, 2013). This statistic lies between 0 and 1 and serves as a generalization of R^2 since it measures the marginal improvement or reduction in unexplained variability in the fixed component after accounting for a given predictor effect. In the case that post–hoc comparisons were necessary, we used the testInteractions function of the phia R package, version 0.1–5 (de Rosario–Martínez, 2012).

3 Results and discussion

Our results are organised into four sections. Each section assesses a specific theoretical hypothesis and provides a short interpretation of the key findings. We first examined the target concepts of the main text, addressing whether readers generate the critical inference in the introductory sentences and monitor their comprehension by the detection of unexpected information. Second, we analysed the target words of the critical sentence, considering whether readers have updated the concept of the text establishing a coherent situation model. Third, in relation to individual differences in working memory, we analysed two regions 1) the target concept of the main text, where no working memory differences were expected in the detection of the unexpected information; and 2) the target words of the critical sentence, where working memory differences should be associated with updating of the unexpected concept. Additionally we observed whether these differences are related to specific verbal capacity or to more general capacities including visuospatial working memory. Finally, we analysed readers' accuracy, explaining whether the product of comprehension depends on comprehension monitoring.

Taking into account the large number of results presented in this study, we focused on the significant fixed effects of each linear mixed model, and only reported the factors comprising the random structure. The summary details (lmerTest package) of each model are provided in Appendix E (p. 195).

Did readers monitor inferentially unexpected information?

To investigate if readers generated an inference in the introductory sentences and thus, were able to detect the inferentially unexpected information, we ran linear mixed models on the target concepts of the text region (e.g., *chess*/*snap*) for all gaze duration, regressions out, go–past time, regressions in and total time eye movement measures.

The linear mixed models performed on go–past time (Model 1), the number of regressions into the target words (Model 2), and total time (Model 3) demonstrated a main effect of expectation, with longer go–past times, $F(1) = 5.82$, $p = .02$, $dv = .61$, larger number of regressions, $F(1) = 4.15$, $p = .05$, $dv = 0.46$, and longer total times, $F(1) = 6.38$, $p = .02$, $dv = .63$, in the unexpected than in the expected condition (see Table 3a, 3b, and 3c). In addition, the random structure of the go–past time showed the random slope of congruence for the participants factor, while the regression in and total time measures manifested the random slope of expectation for the item factor. This meant that the participants varied within congruence variable and items did it within the expectation variable. These effects were controlled with their inclusion in the model. No other random or fixed effect was significant in any of the three models, and not other eye movement measure was significant in this region.

Therefore, once the text information biased a context, readers were able to infer the target concept and incorporate it into their mental representation. Moreover, the cost associated with processing the unexpected relative to the expected concept, confirmed readers' ability to monitor their comprehension by the detection of information that did not fit with their prior inferred concept.

Did readers update the unexpected information into their situation model?

In order to understand if readers updated the target concepts of the text into a coherent situation model, we carried out linear mixed models on the target words of the critical sentence region (e.g., *moving pieces*/*playing cards*), once more for all eye movement measures.

The linear mixed model performed on total time (Model 4) reflected a significant main effect of congruence, $F(1) = 4.53$, $p = .04$, $dv = 0.25$, with longer total times on the target words of the incongruent than in the congruent condition; and a significant two–way interaction of expectation x congruence, $F(2) = 5.68$, $p = .003$, $dv = 0.62$ (see Table 3d). Post–hoc comparisons with Bonferroni correction on the two–way interaction showed that readers spent significantly less time reading the congruent than the incongruent target words only when they had previously seen the expected concept, χ^2 (1) = 14.19, $p < .001$, not when they have seen the unexpected concept, χ^2 (1) = 0.25, p = 1.00, in the main text. Additionally, readers took significantly longer to read the target words in congruent critical sentences, χ^2 (1) = 7.48, $p = .01$, when the main text had presented the unexpected concept, compared to the expected concept. In contrast, this effect was not significant for the target words encountered in incongruent critical sentences, χ^2 (1) = 3.90, $p = .10$. The random structure produced the random slope of congruence for both participant and item. No other random or fixed effect was significant in these or others eye movement measures.

Plausibly, at least two different cognitive processes might underlie this interaction. On one hand, after the presentation of the expected concept, the benefit for reading the target words in the congruent relative to the incongruent comprehension sentence, verified that readers had activated the concept in their memory representation. Nevertheless, no benefit effect was found when the unexpected concept was presented, suggesting that besides the unexpected concept, the expected concept was still active in memory. The pattern of means is also consistent with this possibility, since the presentation of the unexpected concept entailed shorter reading times in the critical sentence (in both the congruent and incongruent conditions) than reading the incongruent condition after the presentation of the expected concept. On the other hand, the processing cost that readers showed on the target words of the congruent critical sentence after the unexpected concept compared to the expected concept indicated that readers experienced difficulty discarding the expected concept from their situation model.

Was there any evidence of individual differences in working memory capacity associated with updating the unexpected information?

To address whether individual differences in the way in which readers update their mental representation after the presentation of inconsistent information was associated with differences in working memory, we ran separate linear mixed models for the verbal and visuospatial domain of working memory. We analysed two regions: 1) the target concept of the main text to clarify if there were individual differences in the monitoring process, and 2) the target words of the critical sentence to understand if the updating process was explained by the verbal domain of working memory. Again, we did this for all eye movements measures.

The linear mixed model conducted on the regressions out of the target concept of the main text (Model 5) reflected a main effect of verbal working memory, $F(1) = 4.64$, $p = .04$, $dv = .56$, where readers with higher verbal span made significantly fewer regressions out of the target concepts, than lower verbal span readers (see Table 3e). No other fixed or random effect was significant in this or others eye movement measures. Therefore, no model including the visuospatial working memory domain was significant.

The linear mixed models performed on the go–past time (Model 6) and regressions into (Model 7) the target words of the critical sentence, showed a two–way interaction of expectation x verbal working memory: for go–past time, $F(2) = 4.23$, $p = .02$, $dv = 0.61$, higher verbal span readers showed significantly shorter go–past times than lower verbal span readers when the target concept was unexpected, $t = -2.86$, $p = .005$, but not when it was expected, $t = 0.03$, $p = .97$, (see Table 3f); for regressions in, $F(2) = 4.14$, $p = .02$, $dv = 0.59$, higher verbal span readers made significantly fewer regressions into the target words of the critical sentence than lower verbal span readers when the target concept was expected, $t = -2.57$, $p = .01$, but not unexpected, $t = -0.21$, $p = .83$ (see Table 3g). In the random structures, the item factor generated the variables of expectation for the go–past time and congruence for the regression in measure. These effects were controlled with the inclusion of their random slopes. No other random or fixed effect was significant in any of the two models, or in any of the other eye movement measures. Once again, no model including the visuospatial working memory domain was significant.

Our data revealed only a main effect of verbal working memory in the monitoring process: higher verbal (but not visuospatial) span readers regressed out of the two target concepts fewer times than lower verbal span readers. Thus, rather than a more specific problem detecting the inconsistency, these data suggest that lower verbal span readers are less able to integrate the target concept (expected and unexpected) with prior text information. Moreover, individual differences were apparent in the process to update inconsistent information. First, readers with higher verbal span showed shorter go–past times than lower verbal span readers but only when the unexpected concept has been presented. This suggests that higher verbal span readers did not require rereading the main text to discard the expected concept primed in the introduction. In contrast, lower verbal span readers spent more time rereading the main text, probably to discard the expected interpretation. Second, readers with higher verbal span made a smaller number of regressions into the target words of the critical sentence than lower verbal span readers, when the expected concept has been previously presented. This indicated that higher verbal span readers had better integrated the expected information when reading the critical sentence than lower span readers.

Did the unexpected information affect the product of comprehension?

Finally, to examine whether the presentation of information that did not fit with the context of the story resulted in problems of comprehension, we performed a linear mixed model on the number of correct responses (accuracy) answering critical sentences. This (Model 8) manifested a main effect of expectation, $F(1) = 15.62$, $p < .001$, $dv = 0.89$, with a smaller number of correct responses in the unexpected than the expected condition (see Table 3h). In addition, a complex random structure was generated and controlled, with the random slope of expectation for the participant factor and an interaction between expectation and congruence for the item factor. No other random or fixed effect was significant in this eye movement measure.

Therefore, the reduction in the number of correct responses in the unexpected compared to the expected condition indicated that the product of comprehension was affected by the prior inferred concept, despite the fact that the unexpected concept was explicit in the main text. Finally, there were no working memory differences in accuracy, suggesting that lower verbal span readers benefitted from the availability of the story to answer the critical sentences.

Table 3. Mean and standard deviation values for each level of the significant fixed effects that resulted in each linear mixed model.

	Measure	Fixed effects	*M*	*SE*
Text	a) *Go-past time*	Expectation: - expected - unexpected	 271 309	 21 37
	b) *Regressions in*	Expectation: - expected - unexpected	 0.11 0.16	 0.02 0.05
	c) *Total time*	Expectation: - expected - unexpected	 268 309	 16 32
Sentence	d) *Total time*	Congruence - congruent - incongruent Expectation:Congruence - expected:congruent - expected:incongruent - unexpected:congruent - unexpected:incongruent	 517 570 476 600 557 541	 63 75 49 81 78 69
	e) *Regressions out*	Verbal working memory - verbal working memory* - higher verbal span	0.152 0.149	 0.03 0.03
	f) *Go-past time*	Expectation:Verbal working memory - expected:verbal working memory* - expected:higher verbal span - unexpected:higher verbal span	 445 445 438	 51 54 51

g) *Regressions in*	Expectation: Verbal working memory - expected:verbal working memory * - expected:higher verbal span - unexpected:higher verbal span	 0.340 0.334 0.339	 0.04 0.03 0.03
h) *Accuracy*	Expectation: - expected - unexpected	 0.87 0.67	 0.02 0.07

* The intercept of the working memory factors represents their average value.

4 General discussion

The aims of this experiment were to investigate how readers monitor and update inferential information into a coherent situation model. In addition, we evaluated whether working memory was associated with individual differences in updating and whether this was equally so for verbal vs. visuospatial working memory.

In our inferential monitoring task, the first two sentences worked as an introduction that facilitated a concept. The third sentence could be expected, if it contained the concept primed by the introduction (e.g., '*chess*'), or unexpected, if it contained a different concept (e.g., '*snap*') that while possible was nevertheless improbable, given the context established by the two–sentence introduction. In relation to our predictions, our results demonstrated longer reading times (go–past time and total time) and a larger number of regressions (regressions in) in the unexpected compared to the expected target concept of the main text. This processing cost demonstrated that readers generated the inference in the introduction, and then detected an inconsistency between the unexpected concept and their memory representation. Thus, readers were able to monitor inferential information. These results are consistent with the literature on comprehension monitoring, where a processing cost after encountering inconsistent information indicate that situation coherence has been disrupted (e.g., Albrecht & O'Brien, 1993; O'Brien, et al., 1998; Orrantia, et al., 2013; Poynor & Morris, 2003). In contrast to previous experiments that used texts focusing on a character's goal and subsequent action, our task tapped a wide range of inferences based on reader's world knowledge (e.g., the knowledge that chess is a two–player board game, which requires concentration and frequently a long time to decide on the next move). This distinction could be interesting to theories of comprehension, since it demonstrates that more general knowledge–based inferences are subjected to the process of comprehension monitoring.

After reading the main text, participants read a critical sentence containing either congruent (requiring a '*Yes*' answer) or incongruent ('*No*' answer) target words with the expected or unexpected concepts. In general, readers took less time (total time) to read the congruent than the incongruent target words of the critical sentence, demonstrating a facilitation when the information was coherent with the prior concept. More specifically, as hypothesized, the presentation of the expected concept (e.g., '*chess*') resulted in shorter reading times (total time) when reading congruent target words (e.g., '*moving pieces*') compared to incongruent target words (e.g., '*playing cards*') in the critical sentence. Since no change had occurred with the expected concept, this result indicated that readers benefitted from the already activated memory representation and thus were able to construct a coherent situation model of the story. Consistent with this, reading times (total time) were longer for target words in critical sentences that were congruent with the unexpected concept (e.g., '*playing cards*' after '*snap*'), than with the expected concept (e.g., '*moving pieces*' after '*chess*'). This time cost might reflect difficulty in updating the situation model because the prior incorrect interpretation was still active in memory. This result is consistent with those studies demonstrating that the presentation of new contradictory information can lead to difficulties with updating because the prior encoded information continues to interfere with comprehension (e.g., Guèraud, et al., 2005; O'Brien et al., 1998; Oostendorp & Bonnebakker, 1999). Moreover, it is also consistent with the Structure–Building model (Gernsbacher, 1990, 1997), which argues that readers experience problems in integrating the mental representation because they are unable to suppress no longer relevant information. Importantly, the information to be suppressed in our task was inferential (e.g., idea of chess), which could complicate the removal of that elaborated interpretation.

A second aim of our study was to explore the association between inferential monitoring and working memory. Since in our inferential monitoring task the inconsistent information was very evident we did not predict working memory differences to be specifically related to the presentation of unexpected information. In our experiment, higher verbal span readers showed fewer returns to the introductory sentences (regressions out) than lower verbal span readers after reading both the expected and unexpected target concepts of the main text. Then, as expected, working memory differences were not specifically associated with the detection of inconsistent information. Nonetheless, higher verbal span readers were better able to integrate the target concept with prior information, suggesting better accessibility to the memory representation. In contrast, lower verbal span readers experienced difficulty integrating the target concept (expected and unexpected) with prior text information, suggesting problems accessing their memory representation. This finding chimes with the assumption that less efficient readers are less able to maintain previous relevant information in working memory because they need those resources to process incoming information (Daneman & Carpenter, 1983).

Furthermore, we predicted individual differences in verbal working memory to be associated to the updating of the situation model. In line with this, higher verbal span readers exhibited shorter reading times (go–past time) than lower verbal span readers on the target words of the critical sentence, but only when the unexpected concept appeared. Therefore, the difference was exclusively found when inconsistent information was presented. This result demonstrated that higher verbal span readers were able to incorporate the unexpected concept into their mental representation and more importantly, discard the prior incorrect interpretation, thereby updating their situation model. In contrast, lower verbal span readers were able to activate the unexpected concept but they had problems discarding the inference generated in the introduction. Thus, lower verbal span readers failed to update the situation model because they did not suppress the no longer relevant information. This finding also fits well with studies showing individual differences in the ability to suppress irrelevant information (e.g., Gernsbacher & Faust, 1991). In addition, these differences were specifically associated with the verbal domain of working memory with no influence of the visuospatial domain. Although the ability to update the contents of working memory has been commonly defined as an executive function (Carretti, Cornoldi, De Beni, & Romanò, 2005; Palladino, Cornoldi, De Beni, & Pazzaglia, 2001), our results reflect that rather than having a general relationship with working memory, the ability to suppress no longer relevant information in the situation model depends specifically on the verbal domain of working memory. This interpretation is consistent with findings demonstrating that the impaired ability of children with poor reading comprehension to suppress irrelevant information is restricted to verbal working memory (Pimperton & Nation, 2010). Moreover, participants with higher verbal span made fewer regressions into the target words of the critical sentence than those with lower verbal span, when the expected concept was presented. This result was interpreted as the ability of higher verbal span readers to easily understand the critical sentence according to their mental representation of the story. Once more, this highlights the relationship between reading comprehension and the verbal domain of working memory. Additionally, the increased frequency of regressions to the critical sentence for lower verbal span readers could be reflecting comprehension monitoring, since it has been seen that the presentation of comprehension questions encourage understanding in low span readers (e.g., Hannon & Daneman, 1998).

Lastly, the product of comprehension (i.e., accuracy of responses) showed fewer correct responses when the concept presented was unexpected. This suggests that the context provided by the introduction exerted a general influence in the global comprehension of the story, making it easier to respond to critical sentences after the presentation of expected than unexpected information. In addition, working memory was not associated with this, suggesting that the presentation of the story together with the critical sentence helped all readers to improve their reading comprehension.

We believe this is the first study reporting verbal working memory differences in the process of updating information using on–line measures of reading narrative texts and critical sentences. Our results showed that readers are able to monitor their inferential comprehension detecting inconsistencies between their mental representation and text information. However, lower span readers failed to successfully update their situation model, because their prior interpretation interferes with the new information, pointing to difficulties with suppressing no longer relevant information. Furthermore, this difficulty in suppressing irrelevant information is specifically connected with the verbal (not visuospatial) domain of working memory capacity, where lower verbal span readers are less able to suppress a prior incorrect interpretation requiring more rereading of the text to update the situation model.