University of Reading

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

Article

Accepted Version

Perez, A., Joseph, H. S. S. L. ORCID: https://orcid.org/0000-0003-4325-4628, Bajo, T. and Nation, K. (2016) Evaluation and revision of inferential comprehension in narrative texts: an eye movement study. Language, Cognition and Neuroscience, 31 (4). pp. 549-566. ISSN 2327-3801 doi: 10.1080/23273798.2015.1115883 Available at https://centaur.reading.ac.uk/45762/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1080/23273798.2015.1115883

Publisher: Taylor & Francis

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Evaluation and revision of inferential comprehension in narrative texts:

An eye movement study

Ana Pérez[1], Holly S. S. L. Joseph[2, 3], Teresa Bajo[1] and Kate Nation[2]

University of Oxford

University of Granada (Spain)[1]

University of Oxford (United Kingdom)[2]

University of Reading (United Kingdom)[3]

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: **anaipe@ugr.es**

Phone: +34660308256

Abstract

We investigated the processes of how adult readers evaluate and revise their situation model during reading by monitoring their eye movements as they read narrative texts and subsequent critical sentences. In each narrative text, a short introduction primed a knowledge-based inference, followed by a target concept that was either expected (e.g., "*oven*") or unexpected (e.g., "*grill*") in relation to the inferred concept. Eye movements showed that readers detected a mismatch between the new unexpected information and their prior interpretation, confirming their ability to evaluate inferential information. Just below the narrative text, a critical sentence included a target 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 target word, reflecting the facilitation of prior information. In addition, when the unexpected (but not expected) concept had been presented, participants with lower verbal (but not visuospatial) working memory span exhibited longer reading times and made more regressions (from the critical sentence to previous information) on encountering congruent information, indicating difficulty in inhibiting their initial incorrect interpretation and revising their situation model.

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

Introduction

To comprehend a text, readers integrate the relevant information in a 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 comprehension monitoring (e.g., Bohn-Gettler, Rapp, van den Broek, Kendeou, & White, 2011) and inference making (e.g., Cain & Oakhill, 1999).

Comprehension monitoring refers to the metacognitive processes by which comprehenders a) detect a mismatch between the current situation model and new information presented in a text (evaluation), and b) repair the mismatch by restructuring the situation model (regulation; see e.g., O'Brien, Rizella, Albrech, & Halleran, 1998). In addition, inference making is an indispensable skill in reading comprehension, because 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 e.g., Graesser & Zwaan, 1995 for review), inferences are the principal engine to establish consistency in text processing.

Interestingly, comprehension monitoring can be linked to inference making. For example, if text information primes an inference (e.g., "*It was 25th of December and Sophie's father was making a traditional Christmas dinner. The turkey was cooking in the…*"), readers would take longer to read a word mismatching the inference (e.g., "*grill*") than a word matching it (e.g., "*oven*"). Evidence consistent with this comes from Poynor and Morris' eye movement study (2003, Experiment 2), in which the presentation of an action inconsistent with a character's previous goal resulted in longer reading times than a consistent action, regardless of whether the goal had been explicitly or implicitly stated in the text. This finding shows that readers are able to *evaluate* inferential information by detecting a mismatch between their inferred situation model and new information. Moreover, when readers detect a

mismatch between their current situation model and incoming text information, the *revision* process forces the activation of the newly-encountered information as well as the reduction of activation from the no-longer-relevant information (this process has been termed *outdating*; see Kendeou, Smith, & O'Brien, 2013 for review). In relation to this, the Knowledge Revision Comprehension framework (KReC; Kendeou & O'Brien, 2014) proposes that the activation of the new information competes with the no-longer-relevant-information, drawing activation away from the outdated information. In the previous example, once presented with the mismatch (e.g., "*grill*"), readers that have revised their situation model would spend less time reading a sentence confirming this concept (e.g., "*The turkey needed to be barbecued"*) than a sentence confirming the prior incorrect inference (e.g., "*The turkey needed to be roasted"*). Empirical evidence has shown that readers commonly experience difficulty in revising new contradictory information (e.g., Guèraud, Harmon, & Peracchi, 2005; O'Brien et al., 1998)[1].

Despite evaluation and revision being closely related processes that are critical for comprehension, very few studies have focused on their interrelation (e.g., van der Schoot, Reijntjes, & van Lieshout, 2012). In addition, studies demonstrating a failure in revising the situation model show that outdated information affects subsequent comprehension (e.g., Kendeou, et al., 2013), but do not explain which cognitive processes underlie this failure. These studies have usually used a sentence-by-sentence procedure, measuring reading times for the whole sentence and/or answers to comprehension questions (e.g., Albrecht & O'Brien, 1993; Rapp & Kendeou, 2007; O'Brien et al., 1998; Zwaan & Madden, 2004). A recent study by Pérez, Cain, Castellanos and Bajo (2015) measured on-line processing in evaluation and revision by recording event-related potentials, but they did so only for a final disambiguating sentence, without providing the opportunity to reread the text after encountering the mismatch or before answering the comprehension sentence. To capture the interplay between mismatch

detection (evaluation) and mismatch regulation (revision) in the construction of the situation model, it is important to measure reading as it happens, across the whole text (not just an isolated sentence) together with the comprehension sentence. The first goal of the present work was to investigate this matter by using eye movements' technique.

In addition, the connection between poor reading comprehension and deficits in inhibitory control has been linked to low working memory capacity in a number of previous studies (Cain, 2006; Carretti, Cornoldi, de Beni, & Palladino, 2004; Pimperton & Nation, 2010). Working memory capacity is related to reading comprehension skills (e.g., Daneman & Merikle, 1996). This relationship has typically been probed by complex span measures, where readers are required to recall verbal information (e.g., digits or words) while completing an additional task (e.g., comprehending sentences). In general, it is concluded that low working memory readers are less able to maintain and process text information, and have difficulty integrating it with prior knowledge into a coherent situation model (e.g., Daneman & Carpenter, 1980; Hannon & Daneman, 2001). Working memory differences in reading comprehension are more closely associated with verbal than nonverbal working memory (Carretti, Borella, Cornoldi, & De Beni, 2009; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Pimperton & Nation, 2010; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000), and working memory capacity has also been related to higher-level aspects of reading comprehension such as inference making (e.g., George, Mannes, & Hoffman, 1997; Virtue, van den Broek, & Linderholm, 2006) and comprehension monitoring (e.g., Daneman & Carpenter, 1983; Schommer & Surber, 1986). Indeed, Pérez et al. (2015) found a relationship between the revision process and individual differences in working memory, which they explained by suggesting that low working memory readers have more difficulty inhibiting an initial incorrect interpretation. Then, although the revision process seems to depend on working memory capacity, it is unclear whether this relationship depends on a domain-specific mechanism. Therefore, our second goal was to assess whether individual differences in

working memory associated with the revision process are related to working memory differences. Given previous work, we anticipated a closer association with verbal working memory (Cain, 2006; Pérez et al., 2015; Pimperton & Nation, 2010).

Monitoring eye movements is an ecologically valid technique to study reading comprehension. It allows the reader to read a text at their own pace without the need for any secondary task; importantly, it captures reading as it happens, providing detailed information about the time course of text processing on-line. It also allows us to measure reading behaviour during and after the reading of a comprehension question. Accordingly, many studies have shown the importance of distinct eye movement measures in comprehension monitoring. Specifically, in relation to mismatch detection (evaluation), readers show longer gaze durations (the total duration of all fixations in a region before leaving it) for a word that is implausible rather than plausible in relation to a previous context, reflecting differences in lexical access (e.g., Connor, Radach, Vorstius, Day, McLean, & Morrison, 2015). Moreover, when readers encounter an ambiguous or implausible word in a sentence context, they make more regressions and show longer go-past times (the sum of all fixations before exiting the word from the right, including re-reading of previous information), indicating integration difficulty (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004). This integration problem is explained because the current (dominant) interpretation competes with the contextual (subordinate) meaning (Sereno, Pacht, & Rayner, 1992). Consistent with mismatch regulation (revision), regression-related measures (e.g., go-past times and regression probabilities) are associated with working memory differences in the semantic reanalysis of inferential sentence comprehension. For example, Calvo (2001; see also Estévez & Calvo, 2000) presented predictive contexts inducing an elaborative inference followed by a continuation sentence that contained an expected or unexpected concept. Compared to high span readers, those with a low spans showed longer regression-path time (i.e., go-past time), longer right-bounded time (sum of all fixations within a region before the eyes fixate any region to the right of it), and

more regressions from the unexpected concept, following the predicting compared to a control context. Interestingly, although Calvo's (2001) study did not account for the revision process, he suggests that these late processing measures may reflect differences in working memory influencing reanalysis or integration processes (p. 374). Finally, total time (the total duration of all fixations in a region) measure is considered a general index of processing difficulty (e.g., Liversedge, Paterson, & Pickering, 1998; Rayner, Chace, Slattery, & Ashby, 2006). For all these reasons, we recorded eye movements to investigate how readers evaluate and revise their inferential comprehension of a situation model during on-line reading.

The current study

Our paradigm was as follows (see Table 1). A brief two-sentence introduction (e.g., "*It was already 25th of December and Sophie was back home. As a special treat, her father was making her traditional Christmas dinner.*") primed a knowledge-based elaborative inference at the situation model level (McKoon & Ratcliff, 1980). A third sentence introduced a target concept that could be expected (e.g., "*oven*") or unexpected (e.g., "*grill*") according to the inference primed in the introduction. The comparison between the expected and unexpected condition provided an index of evaluation. Just below the main text, a critical sentence contained a target word that could be congruent (e.g., "*roasted*") with the expected but not with the unexpected concept, or incongruent (e.g., "*barbecued*") with the expected but not with the unexpected concept. The cross conditions of the main text (expected vs unexpected) and the critical sentence (congruent vs incongruent) allowed us to examine whether readers had revised their situation model. Eye movements were recorded as participants read both the main text and the critical sentence.

(Table 1 about here)

The general hypotheses were as follows. First, in relation to the evaluation process we predicted that if readers fully generate the inference as they read the first two sentences, they will exhibit longer gaze durations, more regressions and longer go-past times on encountering the unexpected (e.g., "*grill*") than the expected concept (e.g., "*oven*"). This result would reflect their ability to detect information that does not fit with the prior inferred situation model (mismatch detection). Specifically, a difference in gaze duration between the conditions would indicate a difference in time course in accessing previous information when encountering the unexpected versus the expected concept. Second, as a general effect of processing, we hypothesized that when reading the critical sentence, if only the expected concept has been encountered (e.g., "*oven*"), readers should benefit from their already activated memory representation and thus should show shorter total times on the congruent (e.g., "*roasted*") compared to the incongruent (e.g., "*barbecued*") target concept. Because no change has occurred in the story, this effect would reflect a processing facilitation when text information is coherent. On the contrary, if the unexpected concept has been encountered (e.g., "*grill*"), readers should exhibit longer total times to confirm this concept in the critical sentence (e.g., "*barbecued*"), than to confirm the expected concept (e.g., "*roasted*" after "*oven*"). This effect would signal a general processing difficulty when mismatch information has been presented in the text. Gaze duration was not expected to show significant differences in the critical sentence because, as this sentence required the reprocessing of the previously presented information, no early comprehension process (i.e. lexical access) could explain this effect. More importantly, if readers have problems integrating the unexpected concept, they will go back (regressing out of the critical sentence and/or into the text region) and then, they will spend time rereading previous information (go-past time). Therefore, we also expect longer go-past times and more regressions out of the critical sentence as well as more regressions into the target concept of the main text, coming from the unexpected than from the expected concept. Because these rereading indexes reflect integration problems, this

finding would point to the need to revise the situation model when a change in the story has taken place (mismatch regulation).

Interestingly, both go-past time and regression probabilities are associated with working memory differences in sentences requiring the revision of inferential information (Calvo, 2001). Therefore, we expect these regression-related measures to elucidate the relationship between revision and working memory capacity. We predicted no strong association between working memory differences and evaluation because the mismatch detection between the primed inference and the unexpected concept was very evident in the text (see e.g., Pérez et al., 2015). In contrast, we predicted an association between individual differences in verbal working memory and the revision process, with a less strong relationship between revision and visuospatial working memory. Specifically, we expected lower verbal span readers to exhibit longer go-past times and/or more regressions out (or into the main text) than higher verbal span readers on the critical sentence, coming from the unexpected than from the expected concept. This effect would signal the existence of individual differences in the ability to revise the situation model associated with the verbal (but not visuospatial) domain of working memory.

Finally, if the presentation of an unexpected concept exerts an influence on global comprehension, readers should make fewer correct responses to the critical sentence when the unexpected concept has been presented, compared to the expected concept. To our knowledge, this is the first study to assess the way that readers revise their situation model while reading a whole text and a comprehension sentence on-line, and to ask whether verbal and not visuospatial working memory may explain part of its variance.

Method

Participants

Forty undergraduate and postgraduate students (mean age 21.9 years, range 17–47 years[2]) were recruited by an internet advertisement placed on the webpage of 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 mismatch 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 mismatch task. We constructed 34 (4 practice, 30 experimental) four-sentence narrative texts (see Table 1; for full set of materials see Appendix A). The first two sentences of the story described a situation which biased a knowledge-based inference generated at the situation model level. The third sentence presented one of two conditions: a) the expected condition in which the concept was primed by the introduction (e.g., "*The turkey was cooking, and it needed another hour in the oven before it was done*"), and was therefore consistent with the previous information; or b) the unexpected condition in which a valid but unlikely concept (e.g., "*The turkey was cooking, and it needed another hour in the grill before it was done*") was presented, and was therefore inconsistent with the previous information and required participants to change their mental representation. A norming study was carried out to test the two target concepts (e.g., *oven*/*grill*). Participants read each text in one of the two versions of the third sentence. They were instructed to mark on a scale 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 experimental items, only texts for which the expected condition received a high score (more than 3; $M = 4.23$) and the unexpected condition a low score (less than 3; $M = 2.21$) were included. A t-test comparison confirmed the difference between the two target concepts, $t(29) = 15.80$, $p < .001$.

Subsequently, we ran a second norming study to provide confirmation of the inference activation primed by the main text. Participants read each text without the target concept of the third sentence, which was signaled by a gap ("___"). They were instructed to read each text and fill in the gap with the word that they thought fitted best in the context of the text. Fifty-five participants completed this study. χ^2 comparisons demonstrated that the expected concept (or a specific word of its category as e.g., "*spaghetti*" instead of "*pasta*") was significantly used in 27 of the 30 experimental texts (all $p < .012$)[3]. It is important to acknowledge that there was variability in the extent to which our 30 texts constrained the activation of the expected inference, making the activation of several potential related concepts possible in some cases (Lassonde & O'Brien, 2009). Finally, the two target concepts presented in the main text (e.g., *oven*/*grill*) were matched for both word frequency, using the SUBTLEX database (Brysbaert & New, 2009; *M* expected = 34.73; *M* unexpected = 36.24, $t(30) = 0.14$, $p = .89$) and length in number of characters (*M* expected = 5.58; *M* unexpected = 5.65, $t(30) = 0.30$, $p = .76$).

A return separated the main text from the critical sentence presented below, which were presented simultaneously. Participants were instructed not to read the critical sentence before they had read the main text. This sentence presented one of two possibilities: a) the congruent condition in which the information was related to the target concept read in the main text (e.g., "*The turkey needed to be <u>barbecued</u> for one more hour"* after reading "*grill*"); or b) the incongruent condition in which the information was always unrelated to the target concept read in the main text (e.g., "*The turkey needed to be <u>roasted</u> for one more hour*" after reading "*grill*"). 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, the two target concepts presented in the critical sentence (e.g., *barbecued*/*roasted*) were controlled for both SUBTLEX word frequency (*M* congruent = 256.17; *M* incongruent =

437.13, $t(44) = 1.52$, $p = .14$) and length in number of characters (M congruent = 5.78; M incongruent = 5.96, $t(44) = 0.69$, $p = .50$).

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 four out of the six trials. In contrast, when participants failed at least three trials of the same block, the task finished. The scores of each working memory task were the total number of trials correctly recalled. Errors were not included in the final score.

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.

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.

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.

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 were required to judge whether 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.

The standardized scores of the four working memory tasks were extracted from the AWMA battery (see Table 2). Because 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 listening recall and backward digits recall): *M* = 101.77 (*SD* = 12.61; range = 81–128), and the visuospatial domain (average of odd one out and spatial recall): *M* = 105.06 (*SD* = 13.27; range = 77–133).

(Table 2 about here)

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 Eyetrack software[4], and the extraction of eye movements measures were carried out using EyeDoctor and EyeDry. 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.

Procedure

The eye movement experiment (inferential mismatch 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 display was terminated 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, expected-incongruent, unexpected-congruent, or unexpected-incongruent) counterbalanced across participants, keeping a similar number of trials per condition (7 and 8 trials). 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 administered via a 15'' laptop computer screen (AWMA program) in the following order: listening recall, backward digit 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.

Data analysis

A linear mixed model is a powerful statistical analysis that accounts for both fixed and random effects, allowing the inclusion of participants and items in the same analysis. Thus, we constructed linear mixed models using the lmer function of the lme4 R package, version 1.1–7 (Bates, Maechler, Bolker, & Walker, 2013), with Participants and Items as the random

factors, and Expectation (expected vs unexpected) and Congruence (congruent vs incongruent) as the fixed factors. In addition, both domains of working memory were also included as fixed factors: Verbal and Visuospatial. Then, in each model, the fixed structure was composed by two three-way interactions (i.e., expectation x congruence x verbal + expectation x congruence x visuospatial). Separate models were run for each dependent variable (gaze duration, regressions out, go-past time, regressions in, total time and accuracy) across both regions of our inferential mismatch task (main text and critical sentence)[5]. All trials were included in the analyses for eye movements measures, while only correct trials were computed for accuracy. All data were checked to ensure that no participant read the critical sentence before the main text. Moreover, to check whether participants developed a strategy across trials improving their performance, we ran linear mixed models including trial order as a fixed factor: 15 initial vs 15 final trials. There was no fixed effect of trial order in any analysis, supporting the idea that the experimental design did not affect participants' responses. Nonetheless, it is still possible that the critical sentence may well have changed how our participants were reading the texts. Finally, the verbal and visuospatial factors were centred to improve the interpretability of the average (or intercept) of each factor (e.g., Schielzeth, 2010).

Because a maximal random effects structure caused severe convergence problems (Barr, Levy, Scheepers, & Tily, 2013), we followed a well-known procedure (taken from the field of ecology, see Zuur, Ieno, Walker, Saveliev, & Smith, 2009) to establish the optimal structure for the random and fixed components. First, keeping the full fixed structure, we looked for the best random structure using restricted maximum likelihood (REML). The two random factors of participants and items were kept in all models. We assumed different random intercepts because both of them could have a different baseline, and we found the optimal random slopes using data-driven model comparison (see Appendix B; see also Yu,

2015 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 2.0–11 (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 serves as a generalization of R^2 because 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).

Results

Our results are organised into three sections. We first examined the target concepts of the main text, addressing whether readers evaluated their comprehension by detecting unexpected information with respect to the previous inferred concept (see the introduction for predictions regarding evaluation and revision). Second, we analysed the target concepts of the critical sentence, addressing whether readers revised the inferred concept, establishing a new coherent situation model. Finally, we analysed the number of correct responses (accuracy) to the critical sentence, to examine whether the presentation of unexpected information impacted on global comprehension. Means and standard error values of the six dependent variables in both regions are provided in Table 3. Taking into account the large number of results presented in this study, we focused on the 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 C.

(Table 3 about here)

Text region analysis

Linear mixed models with expectation, congruence, and both domains of working memory (verbal and visuospatial) in the fixed structure, were run on the target concept of the text region (e.g., *oven*/*grill*) for all gaze duration, regressions out, go-past time, regressions in and total time eye movement measures[6]. The models performed on go-past time (Model 1) and total time (Model 2) demonstrated a main effect of expectation, with longer go-past times, $F(1) = 5.43$, $p = .02$, $dv = .60$, and longer total times, $F(1) = 6.38$, $p = .02$, $dv = .63$, in the unexpected than in the expected condition (see Table 4a and 4b). This cost associated with the unexpected concept confirmed readers' ability to evaluate their comprehension by detecting information that did not fit with their prior inferred situation model. In addition, the random structure of the total time measure manifested the random slope of expectation for the item factor. This meant that items varied within expectation variable, and its effects were controlled with its inclusion in the model.

Turning to individual differences in working memory, the linear mixed model conducted on regressions out (Model 3) reflected a main effect of verbal working memory, $F(1) = 4.65$, $p = .04$, $dv = .56$, where readers with lower verbal span made significantly more regressions out of the target concept than higher verbal span readers (see Table 4c). Because there was no difference between concepts (expected and unexpected), this finding suggests that lower verbal span readers were less able to integrate the target concept with prior text

information, rather than differences at the level of evaluation. There was no effect of visuospatial working memory. No other fixed or random effect was significant in this or other eye movement measure, in the text region.

Sentence region analysis

Linear mixed models with expectation and congruence along with verbal and visuospatial working memory in the fixed structure, were run on the target concept of the sentence region (e.g., *roasted*/*barbecued*), for all eye movement measures. The model performed on total time (Model 4) reflected a marginal main effect of congruence, $F(1) = 3.13$, $p = .09$, $dv = 0.16$, with longer total times on the incongruent than in the congruent concept; and a significant two-way interaction of expectation x congruence, $F(2) = 4.17$, $p = .02$, $dv = 0.21$ (see Table 4d). Post-hoc comparisons with Bonferroni correction on the two-way interaction showed that readers spent significantly less time reading the congruent than the incongruent concept only when they had previously seen the expected concept, $\chi^2 (1) = 8.53$, $p = .007$, but not when they have seen the unexpected concept, $\chi^2 (1) = 0.03$, $p = 1.00$, in the main text. This effect indicated that prior context facilitated reading only when no change occurred in the story. Additionally, readers took significantly longer to read the congruent concept in the critical sentence, when the main text had presented the unexpected compared to the expected concept, $\chi^2 (1) = 5.85$, $p = .03$; this difference was not significant for the incongruent concept, $\chi^2 (1) = 2.51$, $p = .23$. These effects signaled that while the presentation of the expected inference during the main text facilitated understanding of the congruent condition in the critical sentence, the alteration of the expected inference (with the presentation of the unexpected concept) affected the comprehension of the congruent information. The random structure produced the random slope of congruence for both participant and item, and its effects were controlled with its inclusion in the model.

In relation to individual differences in working memory, the linear mixed models performed on the go-past time (Model 5) and regressions into the target concept of the critical

sentence (Model 6 and 7), showed significant two-way interactions. First, there was a two-way interaction of expectation x verbal working memory for go-past time, $F(2) = 4.23$, $p = .02$, $dv = 0.31$, where 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 4e). This suggests that higher verbal span readers engaged in less re-reading of the main text to reduce activation of the expected inference, primed by the introductory sentences of the main text. In contrast, lower verbal span readers spent more time rereading the main text, probably to draw activation away from the already activated expected inference. Two two-way interactions were significant for regressions in. First, an expectation x verbal working memory interaction, $F(2) = 3.78$, $p = .03$, $dv = 0.26$, where higher verbal span readers made significantly fewer regressions from the last part of the critical sentence into the target concept than lower verbal span readers when it was expected, $t = -2.55$, $p = .01$, but not unexpected, $t = -0.39$, $p = .70$ (see Table 4f). This result indicated that higher verbal span readers had better integrated the expected information when reading the critical sentence than lower span readers; and second, a congruence x verbal working memory interaction, $F(2) = 3.58$, $p = .03$, $dv = 0.25$, where higher verbal span readers made significantly fewer regressions from the last part of the critical sentence into the target concept than lower verbal span readers when it was incongruent, $t = -2.50$, $p = .01$, but not congruent, $t = -0.42$, $p = .67$ (see Table 4g). This result suggested that higher verbal span readers returned less to the target concept than lower verbal span readers before rejecting the incongruent critical sentence. In the random structures, the item factor generated the random slope of expectation for the go-past time and congruence for the regression in measure, and their effects were controlled with their inclusion in each model. Once again, no effect including the visuospatial working memory domain was significant. No other fixed or random effect was significant in these or other eye movement measure, in the sentence region.

Comprehension accuracy analysis

Finally, a linear mixed model was performed on the number of correct responses (accuracy) to critical sentences. This model (Model 8) manifested a main effect of expectation, $F(1) = 18.05$, $p < .001$, $dv = 1.07$, with a smaller number of correct responses in the unexpected than the expected condition (see Table 4h). This result signalled that global comprehension was affected by the prior inferred concept, despite the fact that the unexpected concept was explicit in the main text. A complex random structure was generated and controlled, with the random slope of an interaction between expectation and congruence and both main effects for the item factor. No other random or fixed effect was significant in the accuracy measure.

(Table 4 about here)

Discussion

The aim of this study was to investigate how readers evaluate and revise inferential comprehension into a coherent situation model. We also assessed whether working memory was associated with individual differences in revision and whether this was restricted to verbal (but not visuospatial) working memory. Discussing our data, we distinguish between two aspects of comprehension monitoring, evaluation (mismatch detection) and revision (mismatch regulation), and consider their relationship with working memory capacity. We also consider on the consequences of encountering unexpected information for global comprehension, as tapped by answers to the comprehension questions.

Evaluation (mismatch detection)

In our inferential mismatch task, the first two sentences served as an introduction that facilitated an inference. The third sentence could be expected, if it contained the concept

primed by the introduction (e.g., "*oven*"), or unexpected, if it contained a different concept (e.g., "*grill*") that while possible was nevertheless improbable, given the previous context. In relation to our predictions, the results demonstrated longer reading times (go-past time and total time) in the unexpected compared to the expected target concept of the main text. This processing cost indicated that the introductory text allowed readers to generate the inference in the introduction, and then detect a mismatch between the unexpected concept and their inferred mental representation. That is, readers were able to evaluate inferential information while constructing the situation model of the text. These results are consistent with the literature on comprehension monitoring, where there is a processing cost after encountering inconsistent information, suggesting that situation coherence has been disrupted (e.g., O'Brien, et al., 1998; Poynor & Morris, 2003). Importantly, our materials did not require the detection of real inconsistent information, but a change between the interpretation of a probable inferred concept (given a specific context), to a less probable alternative. This process has been known as inference alteration, defined as "*the alteration of a person's understanding when led to believe one situation is true, but then finds that a different situation is operating*" (Radvansky & Copeland, 2004). Previous studies investigating inference alternation in reading have measured off-line measures of comprehension, such as accuracy of response to comprehension questions (e.g., Hamm & Hasher, 1992; Lorsbach, Katz, & Cupak, 1998; Radvansky & Copeland, 2004). Our results fit with these findings in signalling that an altered inference causes difficulties in comprehension and extend them to on-line processing, as reading happens. A compatible explanation is that unexpected information provokes a coherence break in the mental representation (e.g., Cain, Oakhill, & Bryant, 2004; Rinck & Weber, 2003), leading to longer reading times relative to when expected information is encountered.

We did not predict the evaluation component of our experimental task to be related to individual differences in working memory capacity because the identification of unexpected

information was very evident after a short (only two sentences) and highly biased context (see also Pérez et al., 2015). Interestingly however, we found that higher verbal span readers showed fewer returns to the introductory text (regressions out) than lower verbal span readers overall, evident after reading both the expected and unexpected target concepts of the main text. This shows that working memory differences were not specifically associated with the mismatch detection (evaluation process). Nonetheless, this general effect of verbal working memory suggests differences in the ability to integrate critical information with prior information: higher verbal span readers were better able to integrate the target concept with prior information than lower verbal span readers. The relative difficulty integrating critical information (expected and unexpected) with prior text information seen in lower verbal span readers suggests inefficient access to their memory representation; this also 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).

Revision (mismatch regulation)

Our second process of interest was revision. After reading the main text in the inferential mismatch task, participants read a critical sentence containing a target word that was either congruent (requiring a "*Yes*" answer) or incongruent ("*No*" answer) with the previous expected or unexpected concept. In general, readers spent less time (total time) reading the congruent than the incongruent target word of the critical sentence, demonstrating facilitation when the information was coherent with the prior concept. More specifically, as hypothesized, the presentation of the expected concept (e.g., "*oven*") resulted in shorter reading times when reading the congruent concept (e.g., "*roasted*") compared to the incongruent concept (e.g., "*barbecued*") in the critical sentence. Because no change had occurred with the expected concept, this result indicated that readers benefitted from the already activated inference when reading the congruent critical sentence and thus, were able

to construct a coherent situation model of the story. Moreover, readers spent more (total) time reading the congruent sentence after presentation of the unexpected concept (e.g., "*barbecued*" after "*grill*"), than the expected concept (e.g., "*roasted*" after "*oven*") within the main text. This processing cost confirms that the prior inferred concept was still active in memory when the unexpected concept was encountered, which in turn suggests a difficulty in revising the situation model. This result is consistent with studies demonstrating that the presentation of new contradictory information can lead to difficulties with revising, arguably because the prior encoded information continues to interfere with comprehension (e.g., Guèraud, et al., 2005; Kendeou, et al., 2013; O'Brien et al., 1998). Specifically, it supports the view of the KReC framework (Kendeou & O'Brien, 2014), by which activation of new information competes with the now-outdated information, drawing activation away from the previous incorrect interpretation. Furthermore, our result is also consistent with studies of inference alteration, which have found that participants continue to agree with the prior interpretation even when they are also able to appropriately accept the new interpretation (e.g., Hamm & Hasher, 1992).

We also investigated whether working memory capacity was associated with readers' ability to revise their 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 word of the critical sentence, but only when the unexpected concept appeared. That is, before responding to the critical sentence, lower verbal span readers reread previous information for a longer time, probably because they had difficulty dismissing the concept primed in the introduction. This result indicated that while all participants were able to detect a mismatch between the current mental representation and the alternative concept (evaluation process) and incorporate that concept into their situation model (updating process), those with lower verbal span were less efficient at drawing activation away the prior incorrect inference (revision process). This finding fits well with the Structure-Building model (Gernsbacher, 1990, 1997), which argues

that readers have difficulty in integrating new inconsistent information into the situation model because they are unable to suppress no longer relevant information, forming new substructures constructed out of the main mental representation. This effect has been also related to individual differences in working memory (e.g., Gernsbacher & Faust, 1991; Gernsbacher, Varner & Faust, 1990). Our findings complement a number of findings in the literature, including for example that younger children are less able to inhibit no longer relevant information from working memory than older children (Lorsbach, Katz, & Cupak, 1998); older adults show similar difficulties relative to younger adults (e.g., Hamm & Hasher, 1992), as do patients with right hemisphere brain damage (Brownell, Potter, Bihrle, & Gardner, 1986). Our results are also consistent with Pérez et al.'s (2015) electrophysiological study in which people low working memory were as efficient as those with high working memory at evaluating their comprehension (P3a subcomponent), but they had difficulty revising their situation model (P3b subcomponent). Taken together, this evidence points to the idea that verbal working memory differences associated with revision process in comprehension are partly explained by the inhibition process.

In our experiment, individual differences were specifically associated with the verbal domain of working memory with no influence of the visuospatial domain, highlighted once again the close association between verbal working memory and reading comprehension. Although the ability to revise 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, the ability to inhibit no longer relevant information in the situation model depends specifically on the verbal domain of working memory. This interpretation is consistent with the observation that children with poor reading comprehension are less able to suppress irrelevant information specifically in verbal working memory (Pimperton & Nation, 2010). Additionally, participants with higher verbal span made fewer regressions into the target concept of the

critical sentence than those with lower verbal span, when a) the expected concept was presented in the main text, which was interpreted as the ability of higher verbal span readers to easily integrate information into their mental representation of the story, when no change occurred; and b) the incongruent concept was presented in the critical sentence, which suggested that higher verbal span readers returned less to the target concept than lower verbal span readers before rejecting the incongruent critical sentence. In line with the observation that presence of comprehension questions encourages understanding in low span readers (e.g., Hannon & Daneman, 1998), the increased frequency of regressions to the critical sentence for lower verbal span readers might reflect a general effect of comprehension monitoring.

Overall comprehension, as indexed by the accuracy of responses, showed fewer correct responses when the concept presented was unexpected. This demonstrates that the context provided by the introductory text exerted a general influence on global comprehension, making it easier to respond to critical sentences after the presentation of expected information. Although many studies investigating revision (updating) of the situation model have only analysed comprehension accuracy after the presentation of inconsistent information (e.g., Kendeou, et al., 2013; O'Brien et al., 1998), our finding that general comprehension is affected by the presentation of unexpected information is again consistent with the literature on inference alteration (e.g., Lorsbach, et al., 1998; Radvansky & Copeland, 2004). We found no relationship between working memory and overall comprehension, suggesting that the presentation of the story together with the critical sentence helped all readers.

Conclusions

Our results showed that skilled adult readers evaluate their inferential comprehension, detecting a mismatch between their mental representation and prior text information. Lower span readers were less skilled at revising their situation model: prior interpretation competed with new information indicating difficulties in inhibiting no longer relevant information.

Furthermore, this was specifically associated with the verbal domain of working memory, where lower verbal span readers were less able to suppress a prior incorrect interpretation and required more rereading of the text to revise the situation model. Taking these findings into account, we believe this is the first study reporting verbal working memory differences in the process of revision using on-line measures of reading narrative texts and critical sentences.

Acknowledgements

This research was supported by the doctoral research grant from the Spanish Ministry of Education and Science (FPU AP2008–01893) to Ana Pérez, by a grant from the Economic and Social Research Council (RES-000-22-4652) to Holly Joseph and Kate Nation, and by grants of MINECO (PSI2012-33625) and Junta de Andalucía (Excelencia2012-CTS 2369) to Teresa Bajo.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1061−1070. doi:10.1037/0278-7393.19.5.1061

Alloway, T. P. (2007). Automated working memory assessment. Oxford: Harcourt.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255−278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Eigen, C., & Rcpp, L. (2013). Package "lme4".

Bohn-Gettler, C., Rapp, D. N., van, d. B., 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

Brownell, H. H., Potter, H. H., Bihrle, A. M., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. *Brain and Language*, *27*, 310−321. doi:10.1016/0093-934X(86)90022-2

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977−990. doi:10.3758/BRM.41.4.977

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

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*, 489−503. doi:10.1023/A:1008084120205

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31−42. doi:10.1037/0022-0663.96.-1.31

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, *9*, 365-381. doi:10.1080/09658210143000083

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19*, 245−251. doi:10.1016/j.lindif.2008.10.002

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

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology, 91*, 45−66. doi: 10.1016/j.jecp.2005.01.005

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual Differences in Fifth Graders' Literacy and Academic Language Predict Comprehension Monitoring Development: An Eye-Movement Study. *Scientific Studies of Reading, 19*, 114-134. doi: 10.1080/10888438.2014.943905

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450−466. Retrieved from http://search.proquest.com/docview/1297349869?accountid=14542

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561−584. doi:10.1037/0278-7393.9.4.561

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422−433. doi:10.3758/BF03214546

De Rosario-Martínez, H. (2012). phia: Post-hoc interaction analysis. *R package version 0.0-2*.

Estévez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*, 51-61. doi:10.1080/096582100387704

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*. (pp. 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

Graesser, A. C., & Zwaan, R. A. (1995). Inference generation and the construction of situation models. *Discourse comprehension: Essays in honor of Walter Kintsch*, 117−139. Retrieved from http://search.proquest.com/docview/618762663?accountid=14542

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

Hamm, V. P., & Hasher, L. (1992). Age and the availability of inferences. *Psychology and Aging, 7*, 56. doi:10.1037/0882-7974.7.1.56

Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology, 23*, 149−172. doi:10.1006/ceps.1997.0968

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103−128. doi:10.1037/0022-0663.93.1.103

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 (pp. 353–378). Cambridge: 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. doi:10.1037/a0029468

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363−394. doi:10.1037/0033-295X.85.5.363

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.

Lassonde, K. A., & O'Brien, E. J. (2009). Contextual specificity in the activation of predictive inferences. *Discourse Processes*, *46*, 426-438. doi:10.1080/01638530902959620

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. *Eye guidance in reading and scene perception.* (pp. 55-75) Elsevier Science Ltd, Oxford. doi:10.1016/B978-008043361-5/50004-3

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. doi:10.1016/S0022-5371(80)90355-2

Nation, K., Adams, J. W., Bowyer-Crane, C., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139−158. doi:10.1006/jecp.1999.2498

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

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29*, 344−354. doi:10.3758/BF03194929

Pérez, A., Cain, K., Castellanos, M. C., & Bajo, T. (2015). Inferential revision in narrative

texts: An erp study. *Memory & Cognition,* doi:10.3758/s13421-015-0528-0

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

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). Reasoning, integration, inference alteration, and text comprehension. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimentale*, *58,* 133. Retrieved from http://search.proquest.com/docview/200347353?accountid=14542

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

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10,* 241-255. doi: 10.1207/s1532799xssr1003_3

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1290-1301. doi:10.1037/0278-7393.30.6.1290

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition, 31,* 1284−1292. doi:10.3758/BF03195811

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1,* 103−113.

Schommer, M., & Surber, J. R. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology, 78,* 353. doi:10.1037/0022-0663.78.5.353

Seigneuric, A., Ehrlich, M., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13,* 81−103. doi:10.1023/A:1008088230941

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science, 3,* 296-300. Retrieved from http://search.proquest.com/docview/618264497?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

Tremblay, A., & Ransijn, J. (2013). Package "LMERConvenienceFunctions".

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing, 25,* 1665−1690. doi:10.1007/s11145-011-9337-4

Virtue, S., van, d. B., & Linderholm, T. (2006). Hemispheric processing of inferences: The effects of textual constraint and working memory capacity. *Memory & Cognition, 34,* 1341-1354. doi:10.3758/BF03193276

Yu, H. T. (2015). Applying Linear Mixed Effects Models with Crossed Random Effects to Psycholinguistic Data: Multilevel Specification and Model Selection. *The Quantitative Methods for Psychology, 11,* 78-88.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R.* Springer.

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30, 283–288. doi:10.1037/0278-7393.30.1.283

Footnotes

Footnote 1: It is important to note that instead of a contradiction, the previous example shows a change from the interpretation of a likely inference to a less likely inference, which has been known as the inference alteration process (e.g., Radvansky & Copeland, 2004).

Footnote 2: Only two participants differed from the mean group of age (32 and 47 years old), which explains the large age range. The linear mixed model analysis without these two participants demonstrated the same significant fixed and random effects found with all participants in both eye movements and accuracy measures (see Results).

Footnote 3: In the 3 texts remaining, the expected concept was also significantly preferred over most of mentioned concepts, but there was one concept competing with it: "*tiara/veil*", χ^2 (1) = 0.08, p = .78; "*trumpet*/*saxophone*", χ^2 (1) = 0.08, p = .77; and "*chili*/*salsa*", χ^2 (1) = 2.81, p = .09. The linear mixed model analysis without these 3 texts demonstrated the same significant fixed and random effects found with all texts in both eye movements and accuracy measures (see Results).

Footnote 4: Taken from http://www.psych.umass.edu/eyelab/software/

Footnote 5: Accuracy of response to the critical sentence was analysed independently of the region.

Footnote 6: The models performed with gaze duration, regressions out and go-past time measures did not include the congruence variable in the text region.