Research Report

The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs

Ruth Filik, Hartmut Leuthold

School of Psychology, University of Nottingham, Nottinghamshire NG7 2RD, UK
Department of Psychology, University of Tübingen, 72074 Tübingen, Germany

Article history:
Accepted 11 February 2013
Available online 16 February 2013

Keywords:
Language comprehension
Eye-tracking
ERPs
Narrative comprehension
Situation model

Abstract

Little is known about the on-line evaluation of information relating to well-known story characters during text comprehension. For example, it is not clear in how much detail readers represent character-based information, and the time course over which this information is utilized during on-line language comprehension. We describe an event-related potential (ERP) study (Experiment 1) and an eye-tracking study (Experiment 2) investigating whether, and when, readers utilize their prior knowledge of a character in processing event information. Participants read materials in which an event was described that either did or did not fit with the character’s typical behavior. ERPs elicited by the critical word revealed an N400 effect when the action described did not fit with the character’s typical behavior. Results from early eye movement measures supported these findings, and later measures suggested that such violations were more easily accommodated for well-known fictional characters than real-world characters.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

To successfully comprehend a piece of text, readers must form a mental representation of the people, objects, and events that are being described, as well as the relationships between them. To this end, information provided by the content of the text must be integrated with information already possessed by the reader (background or general world knowledge, including schemas and stereotypes), in order to construct a mental model (Johnson-Laird, 1983) or situation model (Kintsch, 1988). Situation models have been proposed to encode various dimensions of texts such as time, space, causation, motivation, and information about story characters (e.g., Zwaan et al., 1995a, 1995b; Zwaan and Radvansky, 1998). This allows readers to generate inferences (e.g., McKoon and Ratcliff, 1992) and anticipate future outcomes, for instance regarding the likely behavior of certain characters. It is an open question, however, how readers utilize their background knowledge in on-line processing of information relating to character behavior. In order to investigate this issue, and to make more fine-grained inferences about the time course and underlying mechanisms involved, we will use methodologies with high temporal acuity, specifically, event-related brain potentials (ERPs) and measures of eye movement behavior during reading.

Previous research relating to the character-based dimension of situation models (e.g., Rapp et al., 2001) showed narrative inconsistencies to have a disruptive influence on language processing. For example, behavioral studies have reported longer sentence reading times when story completions (e.g., descriptions of a character eating a cheeseburger) are inconsistent with
character traits that are either explicitly mentioned (e.g., “X has been a strict vegetarian for 10 years”) (Albrecht and O’Brien, 1993), or can be inferred on the basis of the preceding story context (Peracchi and O’Brien, 2004, Rapp et al., 2001; Rapp and Kendeou, 2007, see also Mensink and Rapp, 2011; Rapp and Gerrig, 2002). Such findings indicate that, based on information about the specific characters involved in the story, readers create (trait-based) character models that enable them to anticipate and track the characters’ future behavior (e.g., Russelle and Bilandzic, 2008; Gaessler et al., 2002; Rapp et al., 2001). It is important to note, however, that research on narrative comprehension has been exclusively concerned with inferences based on information about characters provided in the text. It is therefore critical to investigate whether similar inferential and predictive processes take place if the text introduces a fictional character that is already well known to the reader, such as Dr. Gregory House or Harry Potter.

At present, not much is known about how character-based knowledge stored in long-term memory (LTM) is processed during text comprehension. However, from previous research concerned with the processing of implausible events we know that violations of world knowledge result in disruption to the eye movement record during reading (e.g., Braze et al., 2002; Ferguson and Sanford, 2008; Filik, 2008; Ni et al., 1998; Rayner et al., 2004; Stewart et al., 2004; Warren et al., 2008) and in an enhanced centroparietally distributed, negative-going deflection at about 400 ms (N400) after the critical word in ERP studies (e.g., Filik and Leuthold, 2008; Hagoort et al., 2004; Hald et al., 2007; Leuthold et al., 2012; Nieuwland and Van Berkum, 2006; Van Berkum et al., 2008).

Following on from work carried out by Nieuwland and van Berkum (2006), who provided some initial evidence for the importance of the story context in evaluating the plausibility of events in unfamiliar cartoon-like scenarios, Filik (2008) studied fictional scenarios in which the characters were already well known to participants (e.g., The Incredible Hulk). Specifically, Filik monitored readers’ eye movements while they read sentences containing a locally implausible event that involved a violation of world knowledge, such as “The man picked up the lorry and carried on down the road” which either appeared in a ‘real-world’ context describing an angry man stuck in a traffic jam or an established fictional context such as The Incredible Hulk. A further baseline was included in which an action was described that would be plausible in both contexts, such as, “...glared at the lorry....” Results showed longer reading times on the region of text following the critical word lorry (in this case the words and carried) when the man picks up the lorry in the nonfictional context than in all other conditions. Filik concluded that a fictional context that is already established in the mind of the reader can immediately eliminate the processing difficulty normally found when readers encounter an implausible event (but see Warren et al., 2008). In support of these findings, in an ERP study using similar materials, Filik and Leuthold (2008) reported that the N400 effect typically elicited by locally implausible events is eliminated when a well-known fictional context makes the anomalous word globally coherent. Thus, readers appear to make very rapid use of contextual information, bypassing real-world expectations and instead evaluating information on the basis of their stored knowledge about the fictional scenario.

However, it is important to note that the studies using well-known fictional contexts (Filik, 2008; Filik and Leuthold, 2008) do not yet provide clear-cut evidence for the use of character-based information stored in LTM during text comprehension, because they contrasted fictional with real-world context conditions. Thus, it is conceivable that when presented with information in a fictional context, readers generally expect the unexpected, in that the fantasy context to some degree licenses a “willing suspension of disbelief” (see Gerrig, 1998, for a discussion), and “the impossible becomes possible” (e.g., Warren et al., 2008). Specifically, readers may be more accepting of information that is at odds with their experience of how a certain fictional character might behave, compared to reading about a ‘real world’ man picking up a lorry. Similarly, Huang and Gordon (2011) pointed out that readers might notice the difference between real world and fictional contexts and, hence, process these materials differentially by applying a more lenient interpretation strategy in the fictional scenarios.

Certainly, investigations of the role of character-based LTM information during comprehension could be advanced by comparing experimental conditions within fictional (narrative) contexts. One exciting possibility then is to manipulate the plausibility of characters’ behavior all within fictional scenarios. For example, is a phrase such as “The mouse picked up the dynamite still acceptable when one is referring to Mickey Mouse instead of Jerry (from Tom and Jerry, who is more well known for his violent nature)? This avenue towards a better understanding of the online mechanisms serving narrative comprehension is important in its own right, because the specificity of character-based information as activated by the narrative context remains an open issue. Thus, what is currently not known is the effect of placing information in a well-known fictional context, and manipulating whether or not the described event fits with the reader’s background knowledge of a particular character. Comprehension studies concerned with text-based inferences showed them to be quite specific for certain traits (Rapp et al., 2001) and emotions (Gernsbacher et al., 1992, 1998; but see Gygax et al., 2004). Still, it is conceivable that in a fictional context readers may be more lenient regarding their character-based inferences and predictions.

Indirect support for a more unspecific effect of contextual information stems from a study reported by Rapp (2008), in which he provided evidence for the assumption that the global context can quickly reduce the impact of knowledge-based information during text processing. Rapp examined the boundary conditions for readers’ failure to notice false information that contradicts even well-known facts that are stored in semantic memory (e.g., Gerrig, 1998; Gerrig and Prentice, 1991; Marsh and Fazio, 2006) by measuring reading times to historically accurate vs. inaccurate outcomes (e.g., “The Statue of Liberty was [not] delivered from France to the United States”). Reading times were faster for historically accurate than inaccurate outcomes (e.g., were faster for “The Statue of Liberty was delivered from France to the United States”) when a control context preceded the false statements, indicating detection of the discrepancy. Crucially, however, this reading time effect was attenuated when a preceding context supported the inaccurate outcome,
suggesting that readers did not fully evaluate all incoming information. From a more theoretical perspective, this result together with other findings demonstrating people’s failure to detect violations of world knowledge (e.g., Erickson and Mattson, 1981; Barton and Sanford, 1993; Sanford et al., 2011) agrees with recent language processing views which assume that readers perform, at least initially, only a superficial analysis of incoming words in a sentence (see Sanford and Sturt, 2002; Ferreira et al., 2002; Ferreira and Patson, 2007). For example, according to the Scenario Mapping and Focus Theory (Sanford and Garrod, 1998), the processing system might simply perform a coarse check for the plausibility of incoming information on the basis of the current context.

In sum, studies examining well-known fictional stories (Filik, 2008; Filik and Leuthold, 2008) have provided preliminary evidence for the assumption that readers rapidly infer from stored knowledge the likely behavior of fictional characters introduced in the context. However, in order to strengthen and advance this assumption, it is important to (a) control for possible differences between fictional and real world contexts with regard to the processing strategy readers adopt (e.g., Ferreira and Patson, 2007; Huang and Gordon, 2011; Sanford and Sturt, 2002; Warren et al., 2008) and (b) to investigate the specificity of the character-based information that readers utilize during text comprehension (e.g., Rapp et al., 2001). Thus, the critical questions we ask here are, when reading about a well-known fictional character, (a) does the mere presence of a fictional context have a general influence on processing, and (b) do readers carefully evaluate incoming information in terms of their prior experience of this character (e.g., what Harry Potter typically would or would not do)?

To address these questions, the present research will use fictional scenarios that are already well known to participants (e.g., Harry Potter, Mickey Mouse, Tom and Jerry). For these scenarios, readers might be expected to already have relatively well-formed models about the story and the associated characters, which is in contrast to previous work in which story and character information relating to an unfamiliar protagonist was made explicit only in the text (e.g., Albrecht and O’Brien, 1993; Rapp et al., 2001). Texts will contain the description of an action that either does or does not match with the typical behavior or abilities of the well-known fictional character featured in the story, the processing of which will be compared to descriptions of implausible actions being performed by ‘real world’ characters. The focus is on when and how readers utilize their background knowledge in order to evaluate incoming information during text comprehension. Thus, as will be outlined in more detail below, we would expect to find evidence for on-line detection of information that is inconsistent with the reader’s real world knowledge using both ERPs (Experiment 1) and eye-tracking (Experiment 2).

1It is of course the case that the ‘real world’ conditions in the current study are also in fact fictional scenarios. We use the terminology ‘real world’ as we assume that, in the absence of other information about the characters in these conditions, participants will apply ‘real world’ expectations to what the characters can and cannot do.

2. Experiment 1

2.1. Introduction

In ERP studies, the centroparietal N400 that is generally observed in response to content words (for a recent review, see Kutas and Federmeier, 2011), is regarded as an index of semantic processing, with a word that is a poor fit within the semantic context eliciting a larger N400 than one that is a good fit. This N400 effect is obtained whether the semantic context is at a local (sentence) or a global (discourse) level (e.g., Van Berkum et al., 2003) and it is sensitive to even subtle differences in semantic fit (e.g., Van Berkum et al., 2005). Importantly for the current study, as mentioned earlier, the N400 is also sensitive to real world knowledge violations (e.g., Filik and Leuthold, 2008; Hagoort et al., 2004; Hald et al., 2007; Leuthold et al., 2012; Nieuwland and van Berkum, 2006, Sanford et al., 2011; Van Berkum et al., 2008).

In order to address the issues described above, we compared the ERPs elicited by three experimental conditions (see Table 1). The ‘real world inconsistent’ condition described an action (e.g., a man picking up a lorry) that is inconsistent with the reader’s knowledge of the world. In this condition, we expected the word lorry to elicit a large N400, because our real world knowledge tells us that it is not plausible for a man to pick up a lorry. For comparison, we included a “narrative consistent” condition, which represented a coherent narrative (i.e., a man picking up a lorry is globally feasible in the context of a well-known fictional scenario in which the protagonist, the Incredible Hulk, has superhuman strength). Indeed, the materials for these two conditions were based on those used by Filik and Leuthold (2008), in which the real world inconsistent condition showed a reliably larger N400 than the consistent fictional condition, thus, we expected to replicate this finding.

Of primary interest in the current study is the pattern of effects observed for the “narrative inconsistent” condition, which described an event that was inconsistent with information stored in the reader’s model for that story character (e.g., the information stored for the Scooby Doo story and the character Shaggy would specify that he is not capable of picking up a lorry). If it were the case that in a narrative context readers generally anticipate violations of their real-world expectations, or adopt a shallower processing strategy, or simply do not activate character-based behavior in much detail, then we would expect this condition to pattern with the narrative consistent condition. In contrast, if readers track a character’s behavior based on stored information in LTM, and if this character-specific information is rapidly utilized, we would predict the narrative inconsistent condition to elicit a significantly larger N400 than the consistent condition, and possibly even to pattern with the real world inconsistent condition.

2.2. Results and discussion

Fig. 1 shows the superimposed grand average ERP waveforms for the three trial type conditions over midline electrodes (left panel) and the corresponding difference waveforms together
with their scalp topographies (right panel). Following a positive deflection peaking at about 230 ms at the Cz electrode, a slowly growing negative-going deflection developed that appeared to be more pronounced for the two inconsistent conditions than for the narrative consistent condition.

To assess the time course of this trial type effect, we first examined possible influences of experimental condition on N1 and P2 amplitudes. The lateral ROI analysis of 150–200-ms (N1) mean amplitude data revealed the typical ventrolateral-posterior N1 topography, \( F(2,48) = 18.11, p < .001 \). Trial type did not influence N1 amplitude, all \( F_s < 1.7, p > .16 \). The subsequent P2 (200–250 ms) was maximal over central midline electrodes, \( F(7,168) = 10.35, p < .001 \), and was also unaffected by trial type, all \( F_s < 1 \).

Most importantly, the amplitude of the following negative-going deflection in the 300–500 ms interval (N400), which was most pronounced over central midline electrodes, \( F(7,168) = 4.76, p = .01 \), was reliably influenced by condition as indicated by the significant main effect, \( F(2,48) = 6.58, p < .01 \). Planned comparisons showed a reliably more negative N400 for the real world conditions.

### Table 1 – Example material for Experiment 1 (Critical word in bold).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative consistent</td>
<td>THE INCREDIBLE HULK</td>
</tr>
<tr>
<td></td>
<td>The Incredible Hulk was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.</td>
</tr>
<tr>
<td>Narrative inconsistent</td>
<td>SCOOBY DOO AND SHAGGY</td>
</tr>
<tr>
<td></td>
<td>Shaggy was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.</td>
</tr>
<tr>
<td>Real-world inconsistent</td>
<td>TRAFFIC JAM</td>
</tr>
<tr>
<td></td>
<td>Terry was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.</td>
</tr>
</tbody>
</table>

---

Fig. 1 – (A) Horizontal and vertical EOG activity and grand average event-related brain potentials (ERPs) at electrodes Fz, Cz, and Pz following critical word onset (S) as a function of trial type. Shaded areas indicate the time interval for which mean N400 amplitude was measured and statistically analyzed. (B) Scalp distribution of difference waves in the N400 time interval (300–500 ms). (C) ERP difference waveforms reflecting the N400 effect at electrodes Fz, Cz, and Pz.
inconsistent condition \((M = -0.10 \mu V)\) as compared to the narrative consistent condition \((M = 1.38 \mu V)\), \(F(1,24) = 8.92, p < .01\), replicating the finding of Flik and Leuthold (2008). The narrative inconsistent condition was also significantly more negative than the narrative consistent condition \((-0.05 \mu V \text{ vs. } 1.38 \mu V)\), \(F(1,24) = 9.09, p < .01\). To our knowledge, this is the first evidence to suggest that readers are able to rapidly utilize detailed prior knowledge about a character’s likely behavior based on information stored in LTM, with a violation of this knowledge resulting in an N400 effect. Indeed, the narrative and real world inconsistent condition did not differ significantly from each other \((-0.05 \mu V \text{ vs. } -0.10 \mu V)\), \(F < 1\), suggesting that readers do not immediately accommodate violations of their expectations just because an event occurs in a well-known fictional context. None of the Condition \times Electrode interactions were significant, all Fs < 1.5, \(p s > .22\). Analysis of ERP amplitudes over lateral ROIs confirmed the main effect of condition, \(F(2,48) = 6.6, p < .01\). Planned comparisons revealed the same N400 effect pattern as in the midline analysis.

In summary, the ERP results suggest that when readers encounter information that is inconsistent with their prior knowledge regarding a character’s behavior, then this has an early influence on on-line language processing. Specifically, the N400 elicited by the narrative inconsistent condition was significantly larger than for the consistent condition, and instead patterned with the real world inconsistent condition.

### 3. Experiment 2

#### 3.1. Introduction

While results from Experiment 1 would suggest that readers respond to implausible events in a well-known fictional context in a similar manner as in the real world (i.e., there was an equivalent N400 effect in each case of inconsistency, compared to the narrative consistent condition), this result may be somewhat surprising, particularly in the light of previous studies which suggest that a fantasy context may help readers to integrate ‘impossible’ information (e.g., Nieuwland and van Berkum, 2006; Warren et al., 2008). Thus, we decided to further investigate this issue by monitoring participants’ eye movements while they were reading. During normal reading, readers have the opportunity to go back and re-inspect earlier portions of the text if they are experiencing difficulty in integrating the current text with the context in which it appears and/or their background knowledge. For example, in Warren et al’s (2008) eye-tracking study, effects of context on later integrative processing were found in measures of reading time which take into account time spent re-reading earlier portions of the text. Such behavior is not possible during ERP studies of text comprehension in which the stimuli are visually presented word-by-word. Therefore, in order to potentially get a clearer picture of any later, integrative processes which may be occurring, we carried out an eye-tracking study.

As in Experiment 1, there was a real world inconsistent condition, a narrative inconsistent condition, and a narrative consistent condition. In Experiment 2, a further control condition was included (labeled ‘real world consistent’ in Table 2), in which the critical word lorry was rendered non-anomalous in the real world by changing the preceding verb phrase (e.g., picked up was changed to glared at). See Table 2 for an example material with analysis regions.

Following results from Experiment 1, if readers do immediately utilize character-based knowledge during on-line language comprehension, then we would expect to see more disruption to eye movements in early measures of reading time (e.g., first fixation duration, first-pass reading time, first-pass regressions out) on the critical word for both narrative inconsistent and real-world inconsistent conditions, compared to their consistent controls. If it is the case that a general fantasy context can aid the later integration of implausible information, then we may expect to see a reduction in difficulty for narrative inconsistent conditions in later measures, for example, total reading times on the critical word, or effects downstream in the post-critical region, in the face of continuing difficulty for real-world inconsistent conditions compared to their consistent controls.

#### 3.2. Results and discussion

**First fixation duration**: There were no significant effects in any region \((Fs < 2.40, ps > .13)\).

**First-pass reading times (or gaze duration, for the critical word)**: There were no significant effects in any region \((Fs < 2.27, ps > .14)\).

**First-pass regressions out**: There were no effects in the pre-critical region \((Fs < 1.1)\). In the critical region there was a significant main effect of consistency, \(F_{1}(1,32) = 19.48, p < .01; F_{2}(1,23) = 9.23, p < .05\), with more regressions out for inconsistent than consistent conditions. This suggests that readers had immediately experienced difficulty on encountering the...
critical word in both the real world inconsistent and the narrative inconsistent conditions, which is in line with the ERP findings reported above. There was no main effect of context, and no interaction ($F_{s < 1.4}$). There were no significant effects in the post-critical region ($F_{s < 3.1}$, $p > .09$).

Regressions in: At the pre-critical region, there was also a significant main effect of consistency, $F_{1(1,32)} = 34.78$, $p < .01$; $F_{1(2,3)} = 10.60$, $p < .01$. There were more regressions in for inconsistent than consistent materials, suggesting that on encountering the critical word, participants had gone back to re-inspect the previous region, possibly in order to attempt to integrate the critical word with the preceding verb, or simply to 'buy time' to overcome processing difficulty before moving on in the text. There were no other significant effects in the pre-critical ($F_{s < 1}$), critical ($F_{s < 3.05}$, $p > .09$) or post-critical regions ($F_{s < 1}$).

Regression path (go-past) reading time: There were no significant effects in the pre-critical region ($F_{s < 2.1}$, $p > .16$). At the critical region, there was again a main effect of consistency, $F_{1(3,31)} = 4.04$, $p = .05$; $F_{1(2,3)} = 6.19$, $p < .05$, indicating that when readers had encountered information that was inconsistent with their real world knowledge, or their knowledge about well-known fictional characters, they had gone back to re-inspect earlier portions of the text. There were no other effects in the critical ($F_{s < 1}$) or post-critical region ($F_{s < 2.67}$, $p > .11$).

Total reading time: There were no effects in the pre-critical region ($F_{s < 1.7}$). At the critical region there was a main effect of consistency, $F_{1(3,31)} = 5.44$, $p < .05$; $F_{1(2,3}) = 7.27$, $p < .05$. However, this was qualified by a Consistency x Context interaction, $F_{1(3,31)} = 4.24$, $p < .05$; $F_{1(2,3)} = 4.33$, $p < .05$. Analysis of simple main effects indicated that for real world contexts, readers had longer total reading times when the target word was inconsistent with their real world knowledge than when it was not, $F_{1(3,31)} = 12.49$, $p < .01$; $F_{1(2,3)} = 5.99$, $p < .05$. However, for fictional contexts, there were no differences between inconsistent and consistent conditions ($F_{s < 1}$). This suggests that whilst readers were still finding it difficult to integrate violations of their real world knowledge, experiencing information that is inconsistent with their model about the story and the associated characters led to no equivalent difficulty with later integrative processes. There were no other significant effects in the critical ($F_{s < 1.86}$, $p > .18$) or post-critical ($F_{s < 1.4}$) regions.

In summary, early measures of reading behavior showed that readers made more regressive eye movements, and had longer regression path (go-past) reading times when they encountered the critical region of sentences which contained real world and narrative inconsistencies. This is consistent with the ERP findings from Experiment 1, and suggests that when readers encounter information that is inconsistent with their prior knowledge regarding a well-known character’s behavior, or inconsistent with their knowledge of the real world, then this has an early influence on on-line language processing. Later eye movement measures, specifically, total reading times on the critical word showed that by this point, readers no longer experienced difficulty with narrative inconsistencies, relative to conditions in which the text is consistent with the content of the reader’s character-based knowledge of the story. In contrast, sentences which were inconsistent with the reader’s real world knowledge still resulted in elevated reading times, compared to the consistent real world condition. This suggests that readers may be able to more readily accommodate information which does not fit with their model for the current well-known fictional scenario than information that does not fit with their knowledge of the real world.

4. General discussion

With this study we aimed to investigate the utilization of character-based information stored in LTM during on-line language processing. More specifically, we examined whether the mere presence of a fictional context generally influences online processes underlying text comprehension, and whether readers carefully track the behavior of well-known characters. Findings from ERPs (Experiment 1) and eye-tracking (Experiment 2) suggest that when readers encounter information that is inconsistent with their knowledge of the real world, or with information stored in their model about the story and the associated characters, that this has an early influence on on-line language processing. Specifically, the N400 elicited by the narrative inconsistent condition was significantly larger than for the consistent control condition, and in fact did not differ from the N400 elicited by a violation of real world knowledge. This finding was supported by early measures of reading behavior in the eye-tracking study, which showed that readers made more regressive eye movements, and had longer regression path reading times when they encountered the critical region of sentences containing both real world and narrative inconsistencies.

It is worth noting first that the findings from relatively early measures of language processing suggest that readers had rapidly detected a real world knowledge violation, resulting in an N400 effect and early disruption to eye movements during reading. This replicates previous findings reported in Filik (2008) and Filik and Leuthold (2008) and also agrees with previous eye-tracking and ERP research showing that violations of real world knowledge are readily detected in the eye movement record (e.g., Braze et al., 2002, Ferguson and Sanford, 2008; Ni et al., 1998; Rayner et al., 2004; Warren et al., 2008) and in the N400 elicited by the critical word (e.g., Hagoort et al., 2004; Hald et al., 2007; Leuthold et al., 2012; Nieuwland and van Berkum, 2006; Van Berkum et al., 2008). Thus, the present study demonstrates again that higher-level contextual factors can have an early influence on on-line language processing.

The key novel contribution of the current results is the finding that when readers encounter information that is inconsistent with their model about the story and the associated characters, this also leads to disruption in relatively early measures of eye movement behavior, and to N400 effects in the ERPs elicited by the critical word. There are two main implications of this result. First, it indicates that it is not the presence of a general fictional context per se, which generally influences how readers initially process information relating to character behavior, but rather, specific expectations that readers are able to generate based on their prior knowledge about the likely behavior of the characters.
involved. Therefore, we can rule out the possibility that a fictional context changes the initial processing of incoming information in an unspecific manner, for example, by readers adopting a more lenient or superficial interpretation strategy (e.g., Huang and Gordon, 2011).

Second, and perhaps more importantly, present findings clearly demonstrate that readers rapidly utilized character-based information from LTM to evaluate the reported behavior of a fictional character that is well known to them. Thus, it appears that readers are able to generate specific expectations based on their prior knowledge about the likely behavior of the characters involved. The present findings hence further strengthen the view that (text-based) inferences about characters’ behavior with regard to quite specific traits (Rapp et al., 2001) and emotions (Gernsbacher et al., 1992; Gernsbacher et al., 1998; but see Gygax et al., 2004). Importantly, and in contrast to theories that assume only superficial processing of incoming information (e.g., see Sanford and Sturt, 2002; Ferreira et al., 2002; Ferreira and Patson, 2007), it appears that character-based information, in contrast to our knowledge about facts (such as historical events, e.g., Rapp, 2008), is more fully processed. Ultimately, this might have to do with the importance of the social domain for us as human beings, as in every day life we strive to understand other people’s behavior with regard to their intentions and motives (cf. Heider, 1958).

More generally, our study extends previous work specifically relating to the character-based dimension of situation models (e.g., Albrecht and O’Brien, 1993; Rapp et al., 2001; Rapp and Kendeou, 2007). These studies have indicated that character-based information that is inconsistent with information that has been explicitly presented in the prior context, or can be readily inferred, can result in longer sentence reading times, suggesting that readers do detect such inconsistencies. The current work extends these findings by showing that readers can utilize character-based information stored in LTM to evaluate on-line the behavior of a well-known story character. In addition, it is informative regarding the time course of this process, in that violations of character-based knowledge have an early influence on online reading behavior, detectable in relatively early measures of eye movement behavior and electrical brain activity.

Finally, it is worth noting that later eye movement measures revealed an interesting pattern of consistency effects. Specifically, total reading times on the critical word showed that by this point, readers no longer experienced difficulty with the narrative inconsistent condition, relative to the narrative consistent condition. In contrast, the real world inconsistent condition still resulted in elevated reading times, relative to a control condition. This would suggest that, in a fictional context, readers may be able to more readily accommodate information which does not fit with their situation model for the current scenario. These later effects are consistent with previous studies showing that a general ‘fantasy’ (Nieuwland and van Berkum, 2006; Warren et al., 2008), or counterfactual (Ferguson and Sanford, 2008), context can aid the integration of impossible information into the readers’ situation model (see also Stewart et al., 2004, for related findings on the influence of prior knowledge of brands on eye movements during reading). Interestingly, then, the current results indicate a similar time course and ERP signature for the initial detection of both real world inconsistencies and narrative inconsistencies, which may suggest that there are similar underlying mechanisms at play in the detection of inconsistencies in text, regardless of the source of the inconsistency. Where the findings diverge is in how the processor then attempts to integrate, or accommodate, the violation into the current situation model in each case.

In conclusion, the current results suggest that character-based information stored in LTM allows for the rapid evaluation of a character’s described behavior in relation to the reader’s prior knowledge. In addition, it appears that violations of this fictional character-based knowledge are more easily accommodated than violations of knowledge relating to what is and what is not possible in the real world.

5. Experimental procedures

5.1. Experiment 1

5.1.1. Participants
Twenty-five right-handed native English-speaking Glasgow University students (11 males, 14 females) were given an inconvenience allowance to take part.

5.1.2. Materials and design
There were three experimental conditions (see Table 1 for an example material showing all three conditions, and Appendix for a larger selection). As discussed above, in the ‘real world inconsistent’ condition, the target sentence described an event which violates our knowledge of what is possible in the real world. In the ‘narrative inconsistent’ condition, the target sentence describes an event which does not match information stored in LTM regarding the character’s typical behavior or abilities (i.e., the cartoon character Shaggy is not known for his superhuman strength). Finally, the consistent control condition represented a coherent narrative. The materials were arranged in counterbalanced stimulus files, such that each item appeared in only one of its three versions in a given file, but appeared in all versions across files. A given file comprised 120 experimental materials, with 40 materials in each of the three conditions. Each file also included 200 filler items and six practice items. Fillers were of a similar length to the test materials. Some included referential ambiguities and others contained mild anomalies, but none mentioned well-known fictional characters. Participants were tested in an electrically shielded booth with ambient light kept at a low level. Word stimuli were presented in white 16-point Helvetica font on a black background at the center of a 21-inch computer monitor at a viewing distance of 80 cm. Approximately three characters subtended 1° of visual angle.

Overall, 40 different well-known fictional scenarios (e.g., Tom and Jerry) were used. After taking part in the ERP study, participants rated on a scale of 1–9 how familiar they were with these 40 scenarios, with 1 indicating ‘Not at all familiar’, and 9 indicating ‘Highly familiar’. Overall, scenarios received
an average score of 6.3, indicating that participants were familiar with the scenarios used, which is further supported by the present N400 results.

5.1.3. Procedure
Participants were informed about the EEG procedure and experimental task. After electrode application they were seated in a booth where they read the materials from a computer screen. There were six practice trials to familiarize them with the procedure, after which the experimenter answered any questions. Practice trials were followed by 10 experimental blocks, each consisting of 32 trials. Blocks were separated by a break, the duration of which was determined by the participant. Each trial started with the presentation of the title and first sentence of each sentence pair. Participants pressed the spacebar on a computer keyboard when they had finished reading. A blank interval of 500 ms followed, after which a fixation cross was presented in the center of the screen for 1000 ms. Then the word-by-word presentation of the second sentence started, during which participants were asked to maintain fixation at the center of the screen. Each word was displayed centrally for 300 ms, with 200-ms blank intervals between successive word presentations. A break of 500–1500 ms separated each experimental trial. A comprehension question was displayed following one in ten trials. The mean correct response rate was 86.0%, indicating that participants were reading for comprehension.

5.1.4. EEG recording
A BIOSEMI Active-Two amplifier system was used for continuous recording of electroencephalographic (EEG) activity from 72 Ag/AgCl electrodes, including electrodes for recording horizontal and vertical eye movements (cf. Filik and Leuthold, 2008). EEG and EOG recordings were sampled at 256 Hz. Offline, EEG and EOG recordings were high-pass filtered (0.1 Hz, 6 dB/oct), EEG channels were recalculated to a linked mastoid reference, and trials containing blinks were corrected using the adaptive artifact correction method of Brain Electromagnetic Source Analysis (BESA) software (Ille et al., 2002). The analysis epoch started 200 ms prior to the onset of the critical word and lasted for a total duration of 1700 ms.

5.1.5. Data analysis
Trials containing non-ocular artifacts (drifts, channel blockings, EEG activity exceeding ±60 μV) were discarded. This resulted in a loss of 2.88 to 3.96 trials (SD 3.02–4.45) on average across experimental conditions. The signal at each electrode site was averaged separately for each experimental condition time-locked to the onset of the critical word. Before the measurement of ERP parameters, EEG and EOG activity was low-pass filtered (10 Hz, 6 dB/oct) and aligned to a 100-ms preword baseline. Mean ERP amplitudes were measured in average waveforms within time intervals during which specific ERP deflections and experimental effects were found to be most pronounced, similar to our previous study (Filik and Leuthold, 2008): 150–200 ms (N1), 200–250 ms (P2), and 300–500 ms (N400) relative to critical word onset. N1 amplitudes were examined to reveal possible influences on early word processing (e.g., Sereno et al., 1998) and P2 amplitudes to determine whether processes before those eliciting the N400 were influenced.

ERP amplitude data at midline electrodes were analyzed separately from data recorded over lateral electrode sites. Lateral electrode sites were pooled to form 12 regions of interest (ROIs). That is, electrodes were divided along left–right, anterior-to-posterior, and dorsal–ventral dimensions as in previous studies conducted in our lab (e.g., Filik et al., 2008). Statistical analyses were performed by means of Huynh–Feldt corrected repeated measures analyses of variance (ANOVA).

ERP amplitudes measured at midline electrodes were subjected to an ANOVA with the variables trial type (real world inconsistent, narrative inconsistent, consistent control) and electrode (AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz). ERP data acquired from lateral electrode sites (ROIs) were submitted to an ANOVA with variables trial type, hemisphere (left, right), ant–pos (anterior, central, posterior), and verticality (ventral, dorsal). Bonferroni-adjusted planned comparisons were performed to decompose the effect of trial type.

5.2. Experiment 2

5.2.1. Participants
Thirty-two native English-speaking Glasgow University students (13 males, 19 females) were given an inconvenience allowance to take part.

5.2.2. Materials and design
Twenty-four materials were constructed; based on those used in Experiment 1 (see Table 2 for an example showing analysis regions). As in Experiment 1, there was a real world inconsistent condition, a narrative inconsistent condition, and a narrative consistent condition. In Experiment 2, a further control condition was included (labeled ‘real world consistent’ in Table 2), in which the critical word lorry was rendered non-anomalous in the real world by changing the preceding verb phrase (e.g., picked up was changed to glanced at). Items were divided into four lists, each including one version of each item, with equal numbers of items in each of the four conditions. Each list also contained 80 unrelated filler items. Materials were presented across two or three lines, with two blank lines between each line of text. The critical region always appears towards the middle of a line.

Overall, 23 different well-known fictional scenarios, a subset of the 40 scenarios from the ERP study, were used. After taking part in the eye-tracking study, participants rated on a scale of 1–9 how familiar they were with these 23 scenarios, with 1 indicating ‘Not at all familiar’, and 9 indicating ‘Highly familiar’. Overall, scenarios received an average score of 7.3, indicating that participants were familiar with the scenarios used.

Materials were further normed to ensure that participants interpreted the character description given at the beginning of the target sentence (e.g., The angry man, in the example given in Table 2) as referring to a character already mentioned in the context (i.e., The Incredible Hulk in the narrative consistent condition, Shaggy in the narrative inconsistent condition, or Terry in the real world conditions), rather than a new character, and that this was equally likely across conditions. Twelve new participants were given the 24 context sentences, counterbalanced across three stimulus files, so that each
participant saw eight sentences from the real world condition, eight from the narrative consistent condition, and eight from the narrative inconsistent condition. Context sentences were followed by the beginning of the target sentence which referred to one of the characters (e.g., The angry man...). Participants were asked to indicate whether the description was most likely to refer to a character that was mentioned in the previous sentence, or to a new character, on a scale from 1 to 8 with 1 indicating “mentioned” and 8 indicating “new”. An additional six filler items were included in which the character mentioned in the second sentence was intended to be new (e.g., Shrek was heading towards the castle with Princess Fiona. The donkey...). All participants gave a score of ‘8’ to all filler items, which would indicate that they were engaged in the task. Results for the experimental materials, in which the target sentence was intended to refer to a previously mentioned character, revealed no significant differences in ratings between conditions (Fs <1). The means were as follows; real world=2.50 (SE = .34), narrative inconsistent=2.66 (SE = .43), narrative consistent=2.71 (SE = .42). This would suggest that participants were equally likely to interpret the target sentence as referring to the appropriate character across all experimental conditions.

5.2.3. Procedure
Eye movements were recorded via an SR Research Eyelink 1000 eye-tracker, which sampled eye position every millisecond. Viewing was binocular, but only the right eye was recorded. Materials were displayed on a computer monitor approximately 72 cm from participants’ eyes. Before the start of the experiment, the procedure was explained and participants were instructed to read normally and for comprehension. Participants were seated at the eye-tracker and placed their head on a chin and forehead rest to minimize head movements. Participants then completed a calibration procedure. Before the start of each trial, a fixation box appeared in the upper left quadrant of the screen. Once the participant fixated this box the stimulus computer displayed the target text. If the participant’s apparent point of fixation did not match with the fixation box then the experimenter recalibrated the eye-tracker. Once the participant had finished reading each item, they pressed a key. A comprehension question was displayed following one third of trials. A correct response rate of 97% indicated that participants were reading for comprehension.

5.2.4. Data analysis

5.2.4.1. Regions. Materials were divided into regions for analysis (see Table 2). The pre-critical region comprised the beginning of the target sentence up to the target word, the critical region comprised the target word (e.g., lorry), and the post-critical region contained the following two words.

5.2.4.2. Analysis. An automatic procedure pooled short contiguous fixations. Fixations under 80 ms were incorporated into larger adjacent fixations within one character. Fixations of less than 40 ms and not within three characters of another fixation were deleted, as were fixations over 1200 ms. Prior to analyzing the data we removed trials where two or more adjacent regions had zero first-pass reading times (indicating track-loss), which accounted for 3.6% of the data.
Six measures of reading behavior are reported. First fixation duration is the duration of the first fixation in a region before going past that region (either a single fixation, or the first of multiple fixations). First-pass reading time (also known as gaze duration when the analysis region comprises a single word) is the sum of all the fixations made in a region until the point of

<table>
<thead>
<tr>
<th>Measure (ms)</th>
<th>Region</th>
<th>Narrative consistent</th>
<th>Narrative inconsistent</th>
<th>Real-world consistent</th>
<th>Real-world inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SE</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>First fixation duration</td>
<td>Critical</td>
<td>203</td>
<td>5.6</td>
<td>204</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>217</td>
<td>9.7</td>
<td>208</td>
<td>7.2</td>
</tr>
<tr>
<td>First-pass reading time</td>
<td>Pre-critical</td>
<td>743</td>
<td>47.7</td>
<td>746</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>234</td>
<td>9.4</td>
<td>244</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>277</td>
<td>13.2</td>
<td>271</td>
<td>14.4</td>
</tr>
<tr>
<td>First-pass regressions out</td>
<td>Pre-critical</td>
<td>3.1</td>
<td>1.2</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>9.7</td>
<td>2.4</td>
<td>19.6</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>4.3</td>
<td>1.3</td>
<td>6.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Regressions in</td>
<td>Pre-critical</td>
<td>18.3</td>
<td>3.5</td>
<td>33.9</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>6.6</td>
<td>2.0</td>
<td>5.5</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>20.0</td>
<td>3.3</td>
<td>17.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Regression path</td>
<td>Pre-critical</td>
<td>799</td>
<td>65.3</td>
<td>767</td>
<td>33.8</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>280</td>
<td>12.8</td>
<td>314</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>303</td>
<td>16.7</td>
<td>313</td>
<td>20.3</td>
</tr>
<tr>
<td>Total reading time</td>
<td>Pre-critical</td>
<td>846</td>
<td>59.5</td>
<td>864</td>
<td>46.1</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>271</td>
<td>10.1</td>
<td>278</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>Post-critical</td>
<td>347</td>
<td>19.1</td>
<td>331</td>
<td>19.1</td>
</tr>
</tbody>
</table>
fixation exits the region either to the left or to the right. First-pass regressions out is the proportion of trials where readers looked back from the region to an earlier piece of text between the time when the region was first entered from the left to the time when the region was first exited to the right. Regression path (or go past) reading time is the sum of fixations from the time that a region is first entered until a saccade transgresses the right region boundary. This measure includes fixations made to re-inspect earlier portions of text and is usually taken to reflect early processing difficulty along with (at least some) time spent re-inspecting the text in order to recover from such difficulty. Regressions in reflects the proportion of trials in which a reader made a regressive eye movement into the region, and provides an indication of the probability of re-inspecting a particular portion of text. Finally, total reading time sums the duration of all fixations made within a region, and provides a measure of overall processing of a region of text. In cases where the region was skipped in first-pass, regression path, and total reading times the relevant point was excluded from the analysis, and means were calculated from the remaining data points in the design cell. This procedure resulted in data losses of 0.3% (pre-critical region), 13.8% (critical region), and 14.5% (post-critical region) for first-pass and regression path reading times, and 0% (pre-critical region), 11.8% (critical region), and 10% (post-critical region) for total reading times.

Data for each region were subjected to two 2 context (real world vs. narrative) × 2 consistency (consistent vs. inconsistent) repeated measures ANOVAs, treating participants (F1) and items (F2) as random variables (see Table 3 for descriptive statistics).

Acknowledgments

Thanks to Ian Mackenzie and Kirsty Murphy for assistance with ERP data collection. Correspondence concerning this article should be addressed to Ruth Filik, School of Psychology, The University of Nottingham, University Park, Nottingham NG7 2RD, UK. Email: ruth.filik@nottingham.ac.uk.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.brainres.2013.02.017.

References


