Citation for published version

Ferguson, Heather J. and Sanford, Anthony J. (2008) Anomalies in real and counterfactual worlds: An eye-movement investigation. Journal of Memory and Language, 58 (3). pp. 609-626. ISSN 0749-596X.

DOI

https://doi.org/10.1016/j.jml.2007.06.007

Anomalies in Real and Counterfactual Worlds: An eye-movement investigation

Heather J Ferguson

Anthony J Sanford

University of Glasgow, Scotland, UK

Correspondence to:
Heather Ferguson
email: h.ferguson@psy.gla.ac.uk
Tel: +44 141 330 5034

Department of Psychology
University of Glasgow
Glasgow, G12 8QQ
Scotland, UK

Acknowledgements.

HF acknowledges the support of ESRC in the form of a postgraduate studentship. AJS's contribution was supported by AHRC grant MRG-APN19525. Thanks are due to Robin Morris, Alan Kennedy, Sara Sereno, Christopher Hand, Simon Garrod, Fernanda Ferreira, Simon Liversedge, two anonymous reviewers and members of the Glasgow Language Group, for helpful comments and suggestions on this work.

Abstract

Counterfactual reasoning is valid reasoning arising from premises that are true in a hypothetical model, but false in actuality. Investigations of counterfactuals have concentrated on reasoning and production, but psycholinguistic research has been more limited. We report three eye-movement studies investigating the comprehension of counterfactual information. Prior context depicted a counterfactual-world (CW), or real-world (RW), while a second sentence was manipulated to create RW anomalous continuations, where events included a violation of RW knowledge, and RW congruent continuations, where the events described were congruent with RW knowledge. Results showed that RW violations can be 'neutralised' within an appropriate pre-specified CW context, and RW-congruent items can lead to the experience of an anomaly following an inconsistent CW context. Importantly, there was also evidence in all three studies for early processing difficulty with RW violations regardless of prior context, indicating that a proposition is rapidly evaluated against real-world knowledge, just prior to the accommodation of a proposition into a counterfactual world representation. We discuss the results in terms of a variety of accounts of the nature of counterfactual worlds.

Key words: Counterfactual reasoning, eye-movements, discourse processing, anomalies

Counterfactual reasoning, an understanding of events that are counter to reality, or false, is an essential ingredient of our everyday cognition. Counterfactual situations are frequently depicted through language, yet surprisingly little is known of how they are processed during reading or listening. In this paper, we attempt an exploration of counterfactual processing during reading. Counterfactuals are cases of possibly valid reasoning from premises that are false in actuality (Fauconnier & Turner, 2003), and require the comparison of reality to a model-based alternative. People understand a counterfactual statement, such as, *If money grew on trees then we'd all be millionaires* by keeping in mind two possibilities from the outset: the conjecture, *money grows on trees* and *we are all millionaires*, and the presupposed facts, *money does not grow on trees* and *we are not all millionaires* (Byrne & Tasso, 1999). The counterfactual thus requires that a person represent false information that is temporarily supposed to be true. Linguistic analyses have catalogued a number of ways in which counterfactual worlds may be triggered, including modal terms such as *could*, and *might*, and *if- then* constructions. It is also known that tense influences the plausibility of counterfactual interpretation (e.g. Cowper, 1999; Kratzer, 1991). In the present paper, we rely on *If-then* constructions that clearly can signal a counterfactual world for consideration.

There has been a very large amount of research on reasoning with counterfactuals (c.f., Byrne, 2002), and on what sort of constraints there are on the kinds of counterfactual thoughts people are likely to generate in a variety of circumstances (e.g., Kahneman & Miller, 1986; Byrne, 1997; Markman & Tetlock, 2000). Counterfactuals are ubiquitous in cognitive activities, ranging from simple imagination beyond reality, and fantasy (e.g., Sternberg & Gastel, 1989) to the exploration of possibilities in deductive reasoning (e.g. Byrne & Tasso, 1999; Johnson-Laird & Byrne, 2002). They serve important social functions, for instance in reflecting on past events with negative outcomes [the "if-only.." effect; of Kahenman and Tversky (1982); see also Byrne, 2007; Johnson-Laird & Byrne, 1991; Kahneman, 1995].

In contrast to research within the framework of reasoning and its social concomitants, there has been very little research on how counterfactuals are understood during language comprehension, for instance of what kinds of representations they set up. One approach is that of mental spaces, described by Fauconnier (1985; 1997). Mental spaces are defined as structured, incremental sets that include elements and relationships between them, with availability for new elements to be added and new interactions between the elements to be created. Mental spaces, and the relationships between them, are a way of specifying an interpretation of a discourse. According to Fauconnier, two mental spaces are produced in the case of counterfactual conditionals; one is the reality space and the other is the counterfactual hypothetical space. He sees counterfactuality as a case of forced incompatibility between these two spaces, since what is true in the counterfactual space is false in the reality space. Although Fauconnier presents some very interesting analyses of what is entailed with counterfactual worlds, his analyses do not really provide any basis for predicting how propositions are processed with respect to real world and counterfactual world spaces.

A similar psychological account of reasoning, the mental model theory, has been proposed (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). This theory has a "core" *extensional* account of conditionals, making a conditional 'if p then q' logically equivalent to 'not-p or q'. Consequently, in the case of counterfactual conditionals, it is proposed that both factual and counterfactual possibilities are represented by the reader. An alternative view that has gained increasing interest was initiated by Ramsey (1931), who proposed that when comprehending a conditional statement, people "hypothetically add p to their stock of knowledge and argue on that basis about q". This practice is commonly known as the *Ramsey test*. Recent literature has challenged the mental model theory (Evans & Over, 2004; Evans, Over & Handley, 2005). As an alternative, authors suggest a Suppositional theory where a conditional of the form "if p then q" directs attention to possibilities following from p, and *not* to "not-p or q" possibilities. Therefore, counterfactual statements should be evaluated with respect to suppositional or hypothetical possibilities first.

The present paper is an attempt to examine the role played by real-world (factual) knowledge, and inferences from counterfactual-worlds during on-line comprehension of simple statements. We illustrate the problem with a simple example. In the real world, it is anomalous to say (1):

(1) If the cat is hungry, the owner could feed the cat carrots and it would happily gobble them down.

If a counterfactual world is set up through a statement like (2), then statement (1) is not anomalous with respect to that counterfactual world, although it remains so with respect to the real world.

(2) It would be great if cats were vegetarian.

According to the mental model theory, people have to keep in mind both the conjecture *If cats were vegetarians then (1)*, and the presupposed facts that *cats are not vegetarian and do not like carrots* (e.g., Byrne & Tasso, 1999). Similarly, according to Fauconnier (1985; 1994), two spaces reflecting the real and the counterfactual world are set up. However, according to the Suppositional theory, people would hypothetically suppose that *cats are vegetarians* and then judge their degree of confidence in *feeding cats a bowl of carrots* given that supposition. If the conditional probability was high, they would confidently believe the statement and accept it. Conversely, if the conditional probability was low, they would have doubts about the statement and either reject it or initiate further inferences in order to determine whether it could be consistent with the counterfactual scenario. Although it is undoubtedly true that ultimately a proper appreciation of counterfactuals requires knowledge about both real and counterfactual worlds, it is unclear whether the two would both be present simultaneously in a representation of the discourse model associated with the introduction of a counterfactual situation, or whether there would be a sequential process, in which the counterfactual was temporarily accepted as the true world, and sometime later the consequences of this are tested against the true world for inference. This immediately gives rise to a processing question: can something that is anomalous given our real-world knowledge be "neutralised" as an anomaly if it is consistent within a pre-specified counterfactual world context? According to the model theories of reasoning, this contextual integration process may be delayed so that it initially leads to typical anomaly detection responses, and later becomes accommodated by the counterfactual world representation. Alternatively, the counterfactual world may be the only discourse representation against which a following statement is evaluated, and so if an inferential statement follows from the counterfactual world, it would not show as anomalous immediately (though there might be later consequences). This is the basic question of the present paper. Therefore, for the remainder of this paper, the term 'anomaly' will be used with reference to real-world expectancies, while 'consistency' will be used to refer to the level of consistency with the prior context.

We present three eye-tracking experiments in which the materials utilize propositions that are anomalous with respect to the real-world, but not with respect to some counterfactual world. The question was whether a counterfactual setting, making the anomaly acceptable, would result in the removal of all difficulties associated with the anomaly occurring in the absence of a counterfactual setting, or whether there would still be a difficulty observed because the proposition is compared with real-world information despite the counterfactual setting.

Clear semantic and pragmatic anomalies have effects on the eye-tracking record. Thus, pragmatic anomalies like (3) have been found to induce longer reading times prior to a gradual increase in regressive eye-movements, reaching a maximum at the sentence conclusion (Ni, Fodor, Crain & Shankweiler, 1998; Braze, Shankweiler, Ni, & Palumbo, 2002).

(3) This exotic spice might possibly *seek* the subtle flavour she craves.

In another study, Rayner, Warren, Juhasz and Liversedge (2004) compared anomalous, implausible, and plausible sentences using eye-tracking. The results showed evidence of differential processing with anomalous target words leading to immediate disruption in gaze duration on critical words, while implausible target words showed considerably delayed effects. So, eye-tracking is a useful tool for investigating the time-course of disruption effects due to anomalies, and provides evidence for the early detection of anomalies. Event-related potential (ERP) studies likewise show that pragmatically anomalous words are detected as soon as they occur. The processing of semantic information influences the amplitude of a negative-going ERP component between roughly 250 and 550 ms, and with maximal amplitude at about 400 ms (e.g., Kutas & Hillyard, 1980), the well-documented N400 effect. The N400 has been observed with both word-by-word written presentation, and in listening to continuous speech (e.g., Conolly & Phillips, 1994; Holcomb & Neville, 1990; Kutas & Kluender, 1994). In fact, a large N400 is the default for words that are unpredicatable, or do not fit a context well. Various degrees of contextual support reduce this effect (Kutas, Van Petten, &

Kluender, 2006). An important effect was demonstrated by Van Berkum, Hagoort and Brown (1999), who used ERPs to investigate semantic integration of information in text. Participants were presented with short stories, some of which contained a critical word that, although acceptable in the local sentence context, was semantically anomalous with respect to the wider discourse (4).

(4) Jane told her brother that he was exceptionally *slow*.

While the word *slow* is not anomalous in this sentence, it becomes anomalous when (4) follows a discourse context highlighting the idea that Jane's brother had been very quick, in fact. Relative to coherent control words, the discourse-dependent context anomalies elicited a large N400 effect, similar in surface form to pragmatic anomalies, showing that immediate anomaly detection is not confined to the constraints of the local sentence.

In the remainder of the paper, we present an eye-tracking study in which we examine the pattern of disruption caused by the use of words that are anomalous in real world contexts (RW anomalous), but are not anomalous with respect to some counterfactual world (CW congruent), compared the use of a word that is not anomalous in the real world (RW congruent). The two further eye-tracking studies expand on the initial findings.

Experiment 1

Method

Participants

Thirty-six participants from the undergraduate population of Glasgow University were paid to take part in the study. All participants were native speakers of English, who did not have dyslexia and with vision that they reported to be normal or corrected to normal using soft contact lenses. Participants were naïve to the purpose of the study and had no previous exposure to the test materials.

Materials and Design

Twenty-four experimental items were constructed as in Table 1. In each condition, the first sentence acts as a context. In the RW-inconsistent condition, the first sentence introduces a setting that fits with the real world. The second sentence is then inconsistent, in that the critical word, *carrots*, does not fit RW expectations. In the RW-consistent case, the first sentence sets up a proposition that fits with the real world, and the critical word of the second sentence is changed to one that fits with RW (*fish*, instead of *carrots*). In the case of the CW material, this first sentence introduced a counterfactual world. The second sentence then contained a statement that is consistent with the CW (…*could feed their cat a bowl of* ***carrots***….). The basic design is aimed at comparing the eye- tracking record in response to *carrots* under the inconsistent and consistent conditions with each other, and with *fish* in the consistent condition.

----------Table 1 about here------------

The critical nouns were matched across conditions for length, and for frequency using the British National Corpus and no significant differences were found. The nouns in CW-consistent and RW-inconsistent conditions averaged 5.96 (min.= 3, max.= 9) characters, while the nouns in RW-consistent conditions averaged 6.0 (min.= 3, max.= 11) characters. The nouns in CW-consistent and RW-inconsistent conditions averaged 81.1 appearances per million words, whereas the nouns in RW-consistent conditions averaged 63.0 appearances per million words. Hence, any difference in reading times between conditions will not to be due to discrepancies in length or frequency of the nouns.

One version of each item was assigned to one of three presentation lists, with each list containing twenty-four experimental items, eight in each of the three conditions, blocked to ensure that they were evenly distributed. By rotation over the three lists, all materials appeared in all conditions, but in only one condition within a list. In addition, eighty-two filler materials of different types were added to each list. Twelve participants were randomly assigned to read each list. The twenty-four experimental items in each list were interspersed randomly among the eighty-two filler sentences to create a single random order and each subject only saw each target sentence once, in one of the three conditions. At least one filler item intervened between each experimental item. Comprehension questions followed half of the experimental (i.e., 12) and 41 of the filler trials. Participants did not receive feedback for their responses to these questions. Only participants scoring at or above 90% accuracy on the comprehension questions were used in the data analysis.

Eye tracking

Participants' eye movements were recorded during reading using a Forward Technology Dual Purkinje generation 5.5 eye-tracker. The eye- tracker recorded participants' gaze location and movement from the right eye, although viewing was binocular. Recordings were taken every millisecond. A forehead rest and head strap, along with a bite-bar, were used to stabilize participants' head position and to minimise interference to the signal caused by head movements. All sentences were presented in size 10 Courier New font style through a PC, on a VDU screen, 60cm from the participants' eyes.

Procedure

The eye tracker was calibrated using a series of nine fixed targets distributed across the display, during which the participant was asked to fixate on each point on the computer screen as they appeared in order to establish the correlation between *x/ y* voltages and screen position. Calibration was monitored and adjustments to the calibration were made throughout the experiment as necessary.

Prior to the presentation of each sentence, a pattern of fixation points appeared on the screen, spanning the position to be occupied by the sentences. Participants were instructed to fixate on a sequence of fixation points ending at the top left side, where the first character of the text would be displayed. Once this calibration check was completed accurately, the experimenter pressed a button to advance the screen to display the next item. This procedure ensured that participants were consistently tracking well, and that eye-movement records began uniformly with the initial words in each sentence. Adjustments to the calibration were made at this stage if necessary. Materials appeared individually on the screen, spanning a maximum of four lines of text, with each line separated by two blank lines and a maximum of 75 characters per line. Participants were instructed to read at their normal rate and to read to comprehend the sentences, in order to answer the questions. After reading each sentence, they pressed a button that led to the presentation of a comprehension question or the next trial. Comprehension questions followed 50% of target and filler trials in a pseudo-random order. Participants responded to the questions by clicking either the button in their left hand or the button in their right hand when given a choice of two answers on either left or right side of the screen. Adjustments to the calibration were made during the experiment when the experimenter deemed it necessary.

Results and Discussion

Regions of analysis

The critical second sentence was divided into four regions for the purpose of aggregating reading times and classifying eye-movements, as shown in (5). For each sentence frame, corresponding regions contained the same number of words in all three versions.

(5) If cats were vegetarians they would be cheaper for owners to look after. |Families could$_1$| feed their cat a bowl of$_2$| carrots and$_3$| it would gobble it down happily.$_4$| Cats are loving pets when you look after them well.

The first sentence created a RW or CW context for the critical sentence. Region 1 introduced an agent followed by the modal verb *could*. Region 2 (pre-critical) contained material that that led to the critical region. Region 3 (critical) always consisted of the inconsistent or consistent noun, plus the connective *and*, and thus this region was always two words long. Region 4 (post-critical) comprised the critical sentence wrap-up.

The following measures were used to analyse the tracking patterns. **First-pass reading time**, the sum of the duration of fixations made on first entering a region of text until an eye-movement exits the region to either the left or right. First-pass reading times can provide an indication of the difficulty experienced when participants initially process a region of text. **First-pass regressions out** is the sum of regressive saccades made from the current most rightward fixation with a region of text, indicating the degree to which left to right movement was disrupted during the first sweep of the eyes through a region of text. By examining the location of this disruption it is possible to establish particular regions in a text that cause difficulty for semantic processing. **Regression path times** is the sum of fixations from the first entry into that region from the left to the time the region was first exited to the right. This measure includes fixations made to re-inspect earlier portions of text and therefore provides an indication of the time spent re-reading after the initial detection of a problem. **Total reading time** is the sum of the of all fixations made within a region and provides an indication of the overall amount of time spent processing text in that region. Finally, **regressions-in** convey information on the percent of regressive movements from the right into each region and consequently present researchers with details of the regions of text that readers need to revisit in order to make sense of a piece of text.

An automatic procedure pooled short contiguous fixations. Fixations shorter than 80ms were integrated with larger adjacent fixations within one character and fixations shorter than 40ms that were not within three characters of another fixation were excluded. Fixations longer than 1200ms were truncated. Trials where participants failed to read the sentence or there had been tracker loss were eliminated prior to analysis. Specifically, trials where two or more adjacent regions had zero first-pass reading times were removed, which accounted for less than 1% of the data for any of the experiments reported here.

Table 2 displays mean values for each measure in each condition and region.

----------------------TABLE 2----------------------

First-pass reading times In critical region 3, first-pass reading times showed a significant difference over conditions [$F_1(2, 35) = 5.2$, $p < 0.01$; $F_2(2, 23) = 3.7$, $p =$

0.03]. Further analysis using Bonferroni comparisons showed that RW-inconsistent sentences resulted in significantly longer first-pass reading times than RW-consistent sentences ($t_1(35) = 2.8$, $p < 0.05$; $t_2(23) = 2.4$, $p < 0.05$). Thus the effect of anomaly in the RW context is as predicted, and appears in an early measure. However sentences in the CW-consistent conditions led to longer first-pass reading times than those in the RW-consistent condition ($t_1(35) = 2.8$, $p < 0.05$; $t_2(23) = 2.3$, $p < 0.05$). Furthermore, there were no differences between CW-consistent and RW-inconsistent conditions (ts < 0.2). This early measure thus provides evidence that violations of real-world knowledge are important whether they appear as a direct violation (in the RW-inconsistent condition) or in the context of a counterfactual (the CW-consistent condition, which is of course RW – violating). Thus there is evidence that real-world violations were not neutralised by a counterfactual context. These data are illustrated in Figure 1. Even within a counterfactual context, participants appear to still automatically process information in terms of their real-world knowledge.

In the post-critical region there were no effects of condition on first pass reading times [Fs < 1.2].

-------------------FIGURE 1-----------------------

First-pass regressions out Figure 2 illustrates how each condition affected the mean first-pass regressions out of each region as the sentence progressed.

----------------FIGURE 2-------------------

In the critical region 3, there was no evidence of any effect of condition on regressions out [Fs < 1.8]. In the post-critical region 4, clear effects emerged amongst conditions [$F_1(2, 35) = 24.0$, $p < 0.001$; $F_2(2, 23) = 24.7$, $p < 0.001$]. Planned contrasts showed more first-pass regressions for the RW-inconsistent condition than the RW-consistent condition ($t_1(35) = 5.5$, $p < 0.001$; $t_2(23) = 5.6$, $p < 0.001$), and for the RW-inconsistent condition than the CW-consistent condition ($t_1(35) = 6.4$, $p < 0.001$; $t_2(23) = 6.5$, $p < 0.001$). However, no significant difference was found between CW-consistent and RW-consistent conditions with this measure (ts < 1.3). This data suggests that by the time readers encounter the post-critical region, they are already using the CW context to make a real-world anomaly acceptable.

Regression path times Regression path times showed a difference between conditions in the critical region 3 [$F_1(2, 35) = 4.3$, $p = 0.02$; $F_2 (2, 23) = 3.7$, $p = 0.03$]. The RW-inconsistent condition led to longer regression path times than the RW-consistent condition ($t_1(35) = 2.9$, $p < 0.05$; $t_2(23) = 2.7$, $p < 0.05$), as expected. However, there was no reliable difference between the RW-inconsistent and CW-consistent conditions (ts < 1.8), or between the RW-consistent and the CW-consistent conditions (ts < 1.1) on this measure. This may be because the CW-consistent condition is still being checked against real-world knowledge, though this did not show up in the number of regressions out of region 3. The fact that readers are regressing equally often, but spending more time re-reading in the RW-inconsistent condition suggests differential recovery and integration strategies for RW-inconsistent and CW-consistent conditions following a real-world anomaly.

Differences over conditions in the post-critical region 4 reflect a clearer effect of processing on the basis of prior context. The three conditions differ from one another, [$F_1(2, 35) = 25.1$, $p < 0.001$; $F_2 (2, 23) = 12.6$, $p < 0.001$], and times were longer following the RW-inconsistent than RW-consistent condition ($t_1(35) = 6.2$, $p < 0.001$; $t_2(23) = 4.4$, $p < 0.001$) and longer for RW-inconsistent than the CW-consistent condition ($t_1(35) = 6.1$, p < 0.001; $t_2(23) = 6.1$, p < 0.001). No differences were found between RW-consistent and CW-consistent conditions (ts < 0.2). As with the regressions out measure, this suggests that by the post-critical region, readers are using the CW context to 'neutralise' the RW violation.

Total reading times Reading times in the pre-critical region 2 showed a significant difference between conditions [$F_1(2, 35) = 16.5$, $p < 0.001$; $F_2 (2, 23) = 17.4$, $p < 0.001$]. There were longer total reading times in the RW-inconsistent condition than the RW-consistent condition ($t_1(35) = 4.0$, $p < 0.001$; $t_2(23) = 4.1$, $p < 0.001$) and longer in the RW-inconsistent condition than in the CW-consistent ($t_1(35) = 5.6$, $p < 0.001$; $t_2(23) = 5.7$, $p < 0.001$). The CW-consistent and RW-consistent conditions did not differ (ts < 1.7).

This pattern of total reading times between the conditions persisted in the critical region 3 [$F_1(2, 35) = 23.2$, $p < 0.001$; $F_2 (2, 23) = 20.2$, $p < 0.001$] and into the post-critical region 4 [$F_1(2, 35) = 12.4$, $p < 0.001$; $F_2 (2, 23) = 7.2$, $p < 0.001$]. Planned contrasts revealed longer total reading times in the RW-inconsistent condition than RW-consistent (region 3, $t_1(35) = 6.4$, $p < 0.001$; $t_2(23) = 5.9$, $p < 0.001$; region 4, $t_1(35) = 4.5$, $p < 0.001$; $t_2(23) = 3.4$, $p < 0.001$). The RW-inconsistent condition produced longer times than the CW-consistent (region 3, $t_1(35) = 5.2$, $p < 0.001$; $t_2(23) = 4.9$, $p < 0.001$; region 4, $t_1(35) = 4.1$, $p < 0.001$; $t_2(23) = 3.2$, $p < 0.001$). CW-consistent and RW-consistent did not differ in these regions (region 3, ts < 1.2; region 4, ts < 0.3).[1] Thus, over this later measure of processing in the pre-critical, critical and post-critical regions, there is evidence that the RW anomaly is processed as acceptable following a CW context because readers have integrated the CW context into their current model. In contrast, within a RW context, this violation leads to processing difficulties.

[1] Number of fixations data mirrored total reading times with significant differences between conditions in pre-critical [$F_1(1, 35) = 11.5$, $p < 0.001$; $F_2 (1, 23) = 10.7$, $p < 0.001$], critical [$F_1(1, 35) = 16.6$, $p < 0.001$; $F_2 (1, 23) = 16.2$, $p < 0.001$] and post-critical [$F_1(1, 35) = 6.5$, $p = 0.004$; $F_2 (1, 23) = 3.6$, $p = 0.05$] regions, reflecting more fixations in RW-inconsistent and CW-consistent conditions than RW-consistent.

Regressions in Figure 3 shows the mean number of regressions into each region for each condition.

----------------------FIGURE 3----------------------

Significant main effects of condition were detected in the pre-critical region 2 [$F_1(2, 35) = 12.5$, $p < 0.001$; $F_2 (2, 23) = 12.4$, $p < 0.001$] and critical region 3 [$F_1(2, 35) = 21.2$, $p < 0.001$; $F_2 (2, 23) = 17.2$, $p < 0.001$]. More regressions were made into the pre-critical region in the RW-inconsistent condition than the RW-consistent ($t_1(35) = 4.0$, $p < 0.001$; $t_2(23) = 4.0$, $p < 0.001$) and CW-consistent ($t_1(35) = 4.6$, $p < 0.001$; $t_2(23)= 4.6$, $p < 0.001$) conditions. CW-consistent and RW-consistent conditions did not differ significantly from each other (ts < 0.6). The critical region replicated this effect as the RW-inconsistent condition lead to significantly more regressions in than RW-consistent ($t_1(35) = 5.2$, $p < 0.001$; $t_2(23) = 4.6$, $p < 0.001$) and CW-consistent ($t_1(35) = 6.0$, $p < 0.001$; $t_2(23) = 5.5$, $p < 0.001$) conditions that did not differ from each other (ts < 0.9).

To summarise, a significantly longer first-pass reading time was recorded at the critical region for the CW-consistent and RW-inconsistent conditions compared to RW-consistent. Further, CW-consistent and RW-inconsistent conditions did not differ from each other. This result suggests that processing in terms of real-world knowledge remains active even after the introduction of a counterfactual world. Thus, the present data suggests that statements within the scope of a counterfactual interpretation are initially evaluated against real-world knowledge prior to being accommodated within the counterfactual context. Beyond that, later in processing, and on the basis of all measures considered, the CW-consistent condition does not differ from the RW-consistent condition, and only the RW- inconsistent condition shows a high level of disruption. This shows that after the initial check against the real-world, the CW context becomes the basis of CW processing.

Experiment 2

The question arises as to whether a word that is inconsistent with CW, but is consistent with RW would disrupt processing. It should if CW contexts, when present, are indeed adopted as the basis of processing, after the early use of real-world knowledge. An example of such an inconsistency is shown in (6) where eating fish is inconsistent with cats being vegetarians.

(6) If cats were vegetarians they would be cheaper for owners to look after.
Families could feed their cat a bowl of fish and it would gobble it down happily.

In Experiment 2 we examined such processing of CW-inconsistent information that is congruent in terms of the real world. The aim was to allow a fuller investigation into whether CW information is processed differently from RW information. Additionally, we hoped to explore whether CW inconsistencies are processed in the same way as RW inconsistencies, and specifically whether there is a different pattern or time-course of inconsistency detection for RW and CW information. We expected to replicate the findings of Experiment 1, and additionally to find that the new CW-inconsistent condition led to processing difficulties over the CW-consistent condition.

Furthermore, Experiment 2 ensured that the early effects for RW violations were not the result of semantic priming.[2] Specifically, that information in the CW context sentence was not priming readers' access to the critical word in the critical sentence (i.e. *carrots* being primed by *vegetarians*). Therefore in Experiment 2, RW context sentences contained the same critical words as CW context sentences, but in a realistic framework, as shown in (7).

---

[2] We are grateful to Fernanda Ferreira and Simon Garrod for suggesting this potential problem to us.

(7) Evolution dictates that cats are carnivores and cows are vegetarians.

In fact, it was not anticipated that the early processing effect would be eliminated by these changes for two reasons. First, both consistent and inconsistent critical words have been equally primed in CW context sentences (i.e. *cats* primes *fish* and *vegetarians* primes *carrots*). And secondly, evidence suggests that contextually constrained words are fixated for less time than words not constrained by the semantic context (e.g. Altarriba, Kroll, Sholl & Rayner, 1996; Rayner & Well, 1996; Schustack, Ehrlich & Rayner, 1987). In contrast, the results described above showed increased first-pass reading times in the CW-consistent condition, when the critical word was constrained by prior CW context.

Method

Participants

Thirty-six participants were recruited from the undergraduate population of students from the University of Glasgow and were paid to participate in the study. They were selected through the same criteria as in Experiment 1, and had not taken part in Experiment 1.

Materials and Design

Modified versions of the twenty-four experimental materials used in Experiment 1, plus eight new items, were used. The thirty-two experimental items were as shown in Table 3. A 2x2 within subjects design crossed two context conditions with two consistency conditions. The context condition was split into two levels: a real-world (RW), where the first sentence depicted a realistic circumstance, and counterfactual-world (CW), where the first sentence created a counterfactual alternative. Here, RW context sentences contained the same critical words as the CW context sentences (e.g. *cats* and *vegetarians*) in a realistic scenario, to eliminate possible priming effects. The consistency condition also had two levels and was manipulated by changing the noun in the second sentence: inconsistent, such that events described in the second sentence were inconsistent with the prior context; or consistent, whereby the critical noun was consistent with the pre-specified context. This created a fully crossed experiment with four conditions: RW-inconsistent, RW-consistent, CW-inconsistent, and CW-consistent, as shown in Table 3. Note that the third 'continuation' sentence from Experiment 1 was removed for Experiments 2 and 3 for the sake of brevity and since no significant effects were reported here.

-------------------TABLE 3----------------------

The length and frequency of the critical nouns was matched across conditions. Nouns in the RW-inconsistent and CW-consistent conditions averaged 5.7 (min.= 3, max.= 9) characters, while nouns in the RW-consistent and CW-inconsistent conditions averaged 5.9 (min.= 3, max.= 11) characters. Mean frequency was 87.0 occurrences per million words for the RW-inconsistent and CW-consistent conditions and 64.5 for the RW-consistent and CW-inconsistent conditions.

One version of each item was assigned to one of four lists. The 32 experimental items were assigned to lists so that equal numbers of each condition appeared on each list, with one version of each item appearing on each list, and so that participants did not see more than one version of any given item. The items were displayed with seventy-six filler sentences of various types. At least one filler item intervened between each experimental item. A comprehension question was presented after half of the experimental (i.e., 16) and filler (38) items. Participants did not receive feedback for their responses. All participants used in the data analysis scored at or above 90% accuracy on the comprehension questions.

Eye tracking and Procedure

The eye tracking and experimental procedures were identical to those in Experiment 1.

Results and Discussion

Regions of Analysis

Each target sentence was divided into four regions, as in Experiment 1 and illustrated in (4). Table 4 displays means for each measure for each condition and region.

------------------------TABLE 4 HERE -------------------

First-pass reading times In the critical region 3, no main effects or interactions were detected [all Fs < 0.2].

First-pass regressions out Figure 4 illustrates how each condition affected the mean first-pass regressions out of each region as the sentence progressed.

-------------------FIGURE 4----------------

At the critical region 3, a significant interaction emerged in first-pass regressions-out [$F_1(1, 35) = 4.17$, $p = 0.05$; $F_2(1, 31) = 6.23$, $p = 0.02$]. More regressions were made out of the critical region following a RW violation (RW-inconsistent and CW-consistent), regardless of prior context. The result supports Experiment 1 in that when the critical word is anomalous with respect to the real world, there is an early disturbance in the tracking record, regardless of whether there is a prior counterfactual context (i.e., CW-consistent). This lends further support to the argument that at the critical region, participants automatically process information in terms of their real world knowledge.

The post-critical region 4 revealed a main effect of consistency [$F_1(1, 35) = 4.90$, $p = 0.03$; $F_2(1, 31) = 5.27$, $p = 0.03$], but no interaction [Fs < 0.4], showing that by this region, there was no longer any interference in terms of this measure from real world knowledge in the CW context case. The two inconsistent conditions (RW-inconsistent and CW-inconsistent) led to an increased incidence of first-pass regressions out. These increased regressions out of the post-critical region are likely to be due to readers regressing back in the text to attempt to make sense of the contextual inconsistency. Additionally, it can be noted that, compared to Experiment 1, this study leads to higher incidence of first-pass regressions out of the post-critical region in all conditions. We suggest that this increase is due to several factors, inluding modification and increase in experimental items, particularly the removal of ‘continuation’ sentence 3, which may have accumulated ‘wrap-up’ processes in Experiment 1.

Regression path times The pre-critical region 2 revealed a main effect of context [$F_1(1, 35) = 6.55$, $p = 0.01$; $F_2(1, 31) = 5.13$, $p = 0.03$], with longer reading times following RW than CW contexts. The critical region 3 showed no main effects or an interaction between factors [all Fs < 3.0]. However, by the post-critical region 4, times were longer when the critical word had been inconsistent with the prior context [$F_1(1, 35) = 6.42$, $p = 0.02$; $F_2(1, 31) = 8.13$, $p < 0.01$]. A main effect of context was also found, that was marginal by participants and significant by items [$F_1(1, 35) = 3.63$, $p = 0.06$; $F_2(1, 31) = 4.34$, $p = 0.05$]. This effect was largely due to greatly increased reading times in the RW-inconsistent condition. There was no reliable interaction between consistency and CW/ RW [Fs < 0.1]. Thus, by the post-critical region, CW context is being used without apparent conflict from CW-consistent condition materials being inconsistent with the real world.

Total reading times Similar to Experiment 1, total reading times at the pre-critical region 2 showed a main effect of consistency [$F_1(1, 35) = 4.21, p = 0.05$; $F_2 (1, 31) = 4.03, p = 0.05$] and a main effect of context [$F_1(1, 35) = 14.08, p = 0.001$; $F_2 (1, 31) = 14.47, p = 0.001$]. Inconsistent conditions led to longer total reading times than consistent conditions and longer total reading times when the critical sentence followed a RW context than a CW context. The consistency effect persists into the critical region 3 [$F_1(1, 35) = 4.22, p = 0.05$; $F_2 (1, 31) = 4.23, p = 0.05$], and an interaction between conditions, significant by participants and marginal by items, emerges here [$F_1(1, 35) = 5.18, p = 0.03$; $F_2 (1, 31) = 2.85, p = 0.1$].[3] Effects reported in these pre-critical and critical regions were principally led by increased total reading times for the RW-inconsistent condition and support the suggestion of a larger disruption to the reading process following RW inconsistencies. Thus, over later measures of processing in the pre-critical and critical regions, there is evidence that readers have integrated the CW context into their current knowledge as the RW congruent information is processed as anomalous following a CW context. Similarly, the RW anomaly is processed as acceptable within a CW context.

Regressions in A main effect of consistency was found at the pre-critical region [$F_1(1, 35) = 12.75, p = 0.001$; $F_2 (1, 31) = 16.75, p < 0.001$], revealing increased incidence of regressions into the region following contextually inconsistent than consistent information. Additionally, a significant interaction between context and consistency was found [$F_1(1, 35) = 6.16, p = 0.02$; $F_2 (1, 31) = 6.40, p = 0.02$], showing

[3] Number of fixations data reflected total reading times with main effects of context [$F_1(1, 35) = 9.26, p < 0.005$; $F_2 (1, 31) = 8.07, p < 0.01$] and consistency [$F_1(1, 35) = 5.07, p < 0.05$; $F_2 (1, 32) = 5.14, p < 0.05$] in pre-critical region. Consistency effects persisted into the critical region [$F_1(1, 35) = 6.86, p < 0.01$; $F_2 (1, 31) = 4.03, p < 0.05$].

that there were also more regressions into the pre-critical region when the critical word had been anomalous to our RW knowledge, regardless of prior context.

It should noted that across most measures of reading (namely regressions out, regression path times, total reading times, number of fixations and regressions in) the RW inconsistent condition caused a larger disruption to the reading process than any other condition (including CW-inconsistent), and suggests that RW inconsistencies may have a more powerful effect than CW inconsistencies.

In sum, Experiment 2 showed that world-inconsistent conditions (RW-inconsistent and CW-inconsistent) led to later effects of longer reading times, more fixations and a higher incidence of regressive eye movements around the critical noun than world-consistent conditions (RW-consistent and CW-consistent). This supports findings from Experiment 1 that prior context is rapidly comprehended so that words are processed in terms of that counterfactual world, thus leading to clear inconsistency effects. In addition, an early disruption was observed in response to critical words that violated real world expectations, regardless of prior context (RW-inconsistent and CW-consistent). Such effects were revealed by lengthened first-pass reading times on the critical word in Experiment 1, but in the present study, they were characterised by an increased incidence of regressive eye-movements from and around the critical region. Consequently, we can also claim that RW inconsistencies were detected earlier than CW inconsistencies, as effects of the CW inconsistency were revealed later in the eye-movement record, in regressions from the post-critical region and total reading time measures.

Thus, the results from Experiment 2 allow us to reject the possibility that the effects in initial reading measures are due to priming in the CW conditions, as this effect is still clear when the critical word has been primed by the same critical words in CW and RW conditions. This decision is supported by existing literature (Altarriba, Kroll, Sholl & Rayner, 1996; Rayner & Well, 1996; Schustack, Ehrlich & Rayner, 1987) showing that contextual priming leads to shorter reading times, while our results show increased first-pass reading times in the CW-consistent condition.

Experiment 3

The purpose of Experiment 3 was to confirm that the early effects for RW violations can be replicated, and also to further investigate the role of the modal verb in the critical sentence. Experiments 1 and 2 used an 'If… then…could…' construction to create counterfactual scenarios. However, it was considered that since *could* expresses a conditional possibility or ability, it may imply that a few options are available to the situation, and therefore lends more to a counterfactual interpretation. For example, if a cat were hungry then we *could* feed it bowl of carrots but it would walk away disdainfully. Therefore, Experiment 3 used the modal verb *would* to express that the event is a repeated or habitual action. The use of an 'If…then…would…' composition should draw stronger associations to real-world expectancies. Thus, we examine the strength of a counterfactual context and whether contextual consistency effects can still be identified under these conditions.

Method

Participants

Thirty-six participants from the undergraduate population of Glasgow University took part in the study for a small payment. Participants were selected according to the same criteria as in experiments 1 and 2 and additionally they had not taken part in Experiments 1 or 2.

Materials and Design

Thirty-two experimental items were used in this study. A 2x2 within subjects design crossed two context conditions with two consistency conditions, as described in Experiment 2. Here, the modal verb *could* in the critical sentence was replaced by *would* to investigate the strength of a counterfactual context on comprehension. Additionally, since Experiment 2 found that the early disruptions to the reading process following RW anomalies (regardless of context) were not due to priming in the CW conditions, Experiment 3 used the original RW context design from Experiment 1. An example experimental item is shown in Table 5.

------------ Table 5------------

Length and frequency of the critical nouns were matched across conditions. Mean frequency was 79.0 occurrences per million words for the RW-inconsistent and CW-consistent conditions (mean length = 5.8 characters (min.= 3, max.= 10)) and 65.5 for the RW-consistent and CW-inconsistent conditions (mean length = 6.0 characters (min.= 3, max.= 11)).

One version of each item was assigned to one of four lists for the eye-movement monitoring stage of the experiment. The items were displayed alongside 152 filler sentences of various types.[4] At least one filler item intervened between each experimental item. Comprehension questions occurred after half of experimental (i.e., 16) and half of the filler (76) items. Participants did not receive feedback for their responses and scored at or above 90% accuracy on the comprehension questions.

Eye tracking and Procedure

The eye tracking equipment and experimental procedure were identical to that in the previous experiments.

Results and Discussion

[4] The higher number of filler items in Experiment 3 is attributable to larger experiments serving as filler items for this study.

Regions of analysis

Regions of analysis were as in (4), for consistency with earlier studies. The data were analysed in terms of first-pass reading times, first-pass regressions, regression path times, total reading times and regressions in as shown in Table 6.

------------- TABLE 6 -----------

First-pass reading times The critical region 3 reveals a significant interaction between context and consistency [$F_1(1, 35) = 12.45, p = 0.001$; $F_2\ (1, 31) = 5.22, p = 0.03$]. This interaction is led by increased first-pass reading times following a real-world anomaly, regardless of prior context and was expected here in light of Experiments 1 and 2. These results provide further support that the effects of real-world anomalies may be picked up in measures of very early processing in this experiment. These data are illustrated in Figure 5.

-----------------FIGURE 5----------------------

First-pass regressions out Figure 6 illustrates how each condition affected the mean first-pass regressions out of each region as the sentence progressed.

-------------------FIGURE 6-------------------

No main effects or an interaction between context and consistency variables was found at the critical region 3 (All F's < 0.3). However, a main effect of consistency was revealed at the post-critical region 4 [$F_1(1, 35) = 5.51, p = 0.02$; $F_2\ (1, 31) = 6.21, p = 0.02$], with increased incidence of first-pass regressions out from this region when the critical word had been inconsistent with prior context. These increased regressions out are likely to be due to readers regressing back in the text to attempt to make sense of the inconsistency. Thus, this provides evidence that by the post-critical region, participants are using the CW context to interpret the text, making a real-world anomaly acceptable and real-world congruent information unacceptable.

Regression path times No main effects or interactions were revealed in the regression path measure at any region (All F's < 3.66).

Total reading times Total reading time indicated a significant interaction at the critical region 3 [$F_1(1, 35) = 7.40$, $p = 0.01$; $F_2\ (1, 31) = 4.33$, $p = 0.05$]. Similar to the first-pass reading time data, this interaction showed longer reading times at the critical region when it included a violation of real-world knowledge. This effect, lasting into later measures, suggests that the use of *would* in the critical sentence has led participants to process the passage according real-world expectancies for an extended period. The anticipated main effect of consistency was found at the post-critical region 4, with longer total reading times in the RW-inconsistent and CW-inconsistent conditions than in the RW- and CW- consistent conditions [$F_1(1, 35) = 4.87$, $p = 0.03$; $F_2\ (1, 31) = 4.41$, $p = 0.04$].[5] Thus, by the post-critical region there is evidence that readers have integrated the CW context into their current knowledge.

Regressions in Region 1, that introduced the critical sentence, showed a main effect of context [$F_1(1, 35) = 5.76$, $p = 0.02$; $F_2\ (1, 31) = 4.00$, $p = 0.05$]. This region was more likely to be revisited following a CW context than a RW context. Later at the critical region, a main effect of consistency was revealed [$F_1(1, 35) = 3.97$, $p = 0.05$; $F_2\ (1, 31) = 4.96$, $p = 0.03$], with more regressions into this region when the critical word had been inconsistent with the prior context.

---

[5] Data from number of fixations reflects total reading times, with an interaction at the critical region [$F_1(1, 35) = 9.97$, $p = 0.003$; $F_2\ (1, 31) = 4.06$, $p = 0.05$] and a main effect of consistency at the post-critical region [$F_1(1, 35) = 9.11$, $p = 0.005$; $F_2\ (1, 31) = 4.76$, $p = 0.04$].

The findings in the present study follow those from Experiments 1 and 2. Lengthened reading times at the critical region following a real-world anomaly provides further support for an initial processing mechanism using real-world knowledge. Nevertheless, readers quickly accommodate input to the CW context when one is present. The effects of real-world knowledge are longer lasting in Experiment 3, extending into total reading times and number of fixations at the critical region. We propose that this is due to the use the modal verb *would* to introduce the real-world/ counterfactual event. This prolonged effect was anticipated as *would* suggests that an event is a repeated or habitual, therefore, representing stronger associations to real-world expectancies. However, the fact that counterfactual-world consistency effects were still identified under these conditions provides further evidence that a prior CW context plays an integral, if delayed, role in comprehension.

General Discussion

Processing counterfactual information requires the comprehender to reason within a model that is false with respect to reality. At the same time, it is necessary to know, or to be able to retrieve easily, the fact that the model does not reflect reality, but is an alternative to it, otherwise reality and fantasy would become confounded. At the outset of this paper, we asked whether a CW context would completely remove any trace of a CW-consistent statement being problematic for reading, even though that statement would not make sense with respect to the real world (e.g., *You could feed your cat carrots*). In the reasoning literature, Markovits (1995) pointed out that an initial problem for people presented with counterfactual premises is that these have to be represented without interference from knowledge of the real-world. In normal comprehension, world knowledge is rapidly and automatically recruited to aid interpretation at levels ranging from the lexical (e.g., Rayner, 1998; Rayner & Frazier, 1989; Sereno, Brewer, & O'Donnell, 2003) to the level of situations (e.g. Bower, Black & Turner, 1979; Cook & Myers, 2004; Garnham, 1979; Garrod & Terras, 2000; Rizzella & O'Brien, 2002; see also Sanford & Garrod, 1981, 1998). Classic work has shown that many errors of memory for passages are due to the enrichment of a mental representation of the text by normally appropriate world-knowledge (e.g., Bartlett, 1932; Bower et al., 1979). Sanford and Garrod (1981; 1998; Sanford, 1983; Sanford & Moxey, 1999) have proposed that the mapping of incoming discourse onto existing world knowledge is indeed an automatic, central process, and that without such mapping, rudimentary understanding is impossible. According to their view, it is relating a necessarily fragmentary language input to our understanding of situations that constitutes interpretation, and the richer the mapping, the better the understanding. Frith (1989) termed the tendency to bring real world knowledge to bear on any language input Strong Central Coherence, and has considered the lack of such a capacity to enrich language input a potential problem for persons with autistic spectrum disorder (Frith, 1989; Happé, 1997). It has been argued that an autistic failure to imaginatively elaborate on how counterfactual statements fit with reality may underlie the autistic person's ability to deal with certain types of counterfactual premises (Scott, Baron-Cohen & Leslie, 1999). In sum, there is much evidence to suggest that language input is automatically related to relevant world-knowledge, and that this may well apply to statements that come into the scope of a counterfactual.

In Experiment 1, we used an anomaly paradigm to investigate whether the presence of a counterfactual context would eliminate the effects of a statement that was anomalous with respect to the real world, but which fits the counterfactual context. The results showed that, initially, there was a disruption to eye-tracking even when the anomaly fitted the counterfactual world context, showing as an increase in first-pass reading time at the critical point where the anomaly emerged. Shortly after that, and on measures of later processing, the CW context accommodated the fact completely. Thus within the limits of the paradigm, language input is indeed tested against real-world knowledge, even in the presence of a counterfactual world context. Experiment 2 further showed that when a fact that does not fit the CW is presented in the CW context, that too creates a disruption of processing similar in nature to that obtained in the RW-inconsistent condition, a result replicated in Experiment 3.

Further evidence for the immediate checking of language input against real-world knowledge, even in CW contexts, was found in both Experiment 2 and 3. In Experiment 2, the effect of violating real-world knowledge even in a CW context was revealed in the number of first-pass regressions out of the critical region. An additional purpose of this experiment was to rule out the possibility of priming effects influencing first pass reading times in the four conditions. Therefore, it is possible that because reality in the RW context was not only implied, but explicitly stated (e.g. 'Evolution dictates that cats are carnivores and cows are vegetarians') participants were more alert to the RW anomaly, causing them to immediately regress back in the text to make sense of the anomaly, rather than to seek more information at the critical region. In Experiment 3, a similar effect to Experiment 1 was found, with increased first-pass reading times on the critical word following a real-world anomaly. In this experiment, the effects of the RW anomaly were longer lasting and persisted into later eye-movement measures of total reading time and number of fixations. This is likely to be due to the use of the modal verb *would* in the critical sentence, which requires the reader to accept some event as a usual or repeated behaviour under given conditions. Consequently, it is believed that this should draw stronger associations to real-world expectancies. However, regressions and later effects in the post-critical region revealed that contextual information, though delayed, was used to process the passages. Regardless of the fact that the effects occur across slightly different measures of reading in the three experiments, the statistical analyses led to robust effects on early measures of reading (extending to later measures in Experiment 3), at the same critical region in all three studies. Therefore, we consider the results reported here to support a dual, possibly two-stage, discourse comprehension process (Garrod & Sanford, 1999; Garrod & Terras, 2000; Sanford & Garrod, 2005; Cook & Myers, 2004).

Clearly, when the use of a word violates real-world knowledge, this creates a very early effect upon reading, while contextual information influences later discourse resolution. Thus the basic underlying claim, that language input is tested initially against real-world knowledge even in the presence of a CW context, appears to hold. This effect could, in different ways, fit with both a mental model theory and a suppositional theory of counterfactual conditional processing. Specifically, the mental model theory suggests that people hold in mind the factual possibility that cats are carnivores and therefore do not eat carrots and the counterfactual possibility that cats are vegetarians who do eat carrots. Thus access to either mental space is possible. Additionally, within the mental model theory of conditionals, Santamaria, Espino & Byrne (2005), show that a counterfactual conditional statement primes both factual and counterfactual possibilities whereas a factual conditional primes only the real-world possibility. The effects reported here, where RW-inconsistent information takes longer to process than CW-inconsistent information, support the idea that both possibilities are primed by a counterfactual context while only one is primed by a factual conditional. This suggests that some degree of conflict between reality and the counterfactual world is occurring in a counterfactual scenario. In contrast, the suppositional theory predicts that during the context sentence, readers temporarily add the counterfactual possibility to their store of beliefs by creating a model through a process of *minimal change to reality*. Later processing of events following this context sentence are then dependent on whether that event (e.g. eating carrots) is consistent with the available model. In cases where an inconsistency is initially detected, the suppositional account suggests that participants must refer back to the CW context to evaluate whether it would fit with the alternative world. Accordingly, the early interference caused by RW violations might reflect this extra stage of processing in the case of counterfactuals. It is also interesting to note that given a particular minimal change, some consequents may be easier to infer from the counterfactual state of affairs than others (e.g. *cats eating carrots* given the CW that *cats are vegetarians*). This gives rise to an interesting test of the suppositional theory, as the closeness of the counterfactual to reality will affect the construction of a counterfactual world representation. Consequently, the presence of early effects on processing of real-world violations within a counterfactual context could be predicted by the suppositional model in relation to the distance of change to reality. In comparison, the mental model theory does not make such predictions based on closeness and thus far contains no explanation for the processes by which readers create a counterfactual and factual model. However, the experimental items used in these studies were not constructed to enable a systematic analysis of counterfactual closeness effects on comprehension. Therefore, while this is an interesting issue for future study we feel that involvement in such a debate is not justified here. In sum, we do not commit to either of these theories, only that clearly, for a full understanding of a counterfactual statement, readers must create a representation of both reality and the counterfactual alternative.

A two-stage process of discourse resolution has also been reported by Garrod and Terras (2000). Their study investigated the contribution of purely lexical semantic factors (e.g., that *write* implies using a *pen*) compared to more general contextual factors (e.g., writing on a *blackboard*). They suggest that initial processing of a word is driven by the lexical link between a verb and a dominant role-filler. Thus, dominant role-fillers, such as *the pen,* are integrated automatically with previous material about writing, whereas non-dominant role-fillers, such as *the chalk,* are not and rely on a later process. Further, this early integration process is not influenced by the context in which the role was introduced. Thus *writing on a blackboard* is just as effective for initial integration of *the pen* as is *writing a letter*. Prior context makes an important contribution only at the later second, resolution, stage. The experimental items in the current studies used unbiased verbs to introduce the critical sentence (e.g. *feed* could imply either *carrots* or *fish*). According to Garrod and Terras, the noun in the critical sentence should not influence initial processing. Thus, in terms of lexical priming, *feeding a cat* should be equally effective for the initial integration of *carrots* and *fish*. Additionally, the fact that the strength of associations can be influenced by changing the modal verb in the target sentence (Experiment 3) suggests that more sophisticated language representations are occurring during the comprehension of counterfactuals. On this basis, we believe that the effects reported here, with an early disruption to the reading process following a real-world anomaly, are a genuine consequence of access to real-world knowledge and not simply lexical semantic effects.

Another consideration is the time readers spent processing in a CW context. In order to provide a fair test of the idea that CW contexts would readily accommodate facts congruent with CW, we took care to introduce the counterfactual world in a separate prior sentence, and provided a rationale for the introduction of CW into the discourse. We considered this to be important because firmly establishing the counterfactual would be necessary to fairly answer our initial question. However, the vignettes used in our experiments were very short, and one question is whether with continued exposure to a counterfactual world, continued checking against real-world knowledge would continue. Ultimately, only further experimentation will tell. Nieuwland and Van Berkum (2006) carried out an experiment in which an event impossible in the real world (e.g., *The peanut fell in love*) initially produced the large

N400 effect in the EEG commonly found with such anomalies. However, when a story was produced in which the peanut was firmly established as “a character”, the N400 effect ultimately disappeared. We believe that this observation fits well with our findings that real-world inconsistent information can be accommodated readily within a counterfactual framework when one is clearly provided. Building up the characterisation of a peanut as an animate, feeling entity is one way of doing this (and is commonplace in the world of animated cartoons, for instance). However, it is not at all clear that these results mean that no check against the real world is actually taking place. Furthermore, in many stories, it can readily be argued that keeping contact with real-world knowledge is essential. For instance, stories often contrast what a protagonist believes with what are the “real” facts. A story based in Holland during the tulip fever period concerned an individual who believed a valuable tulip bulb to be an onion, and sliced it up for his lunch (Moggach, 2000, pp 209-211). The whole point of the story is that while in the character’s counterfactual world it was reasonable to eat the ‘onion’, the tension in the story arises from the ongoing recognition that he is devouring a fortune in the form of a valuable bulb. While there is much to be explored on this front, it is generally the case that even in counterfactual worlds, certain constraints of reality have to hold. This has been argued by philosophers of language (e.g., Lewis, 1981, McCall, 1984; von Fintel, 2001). The plausibility of this is easy to see at an intuitive level for the case of the peanut falling in love, discussed above. When we hear that a peanut falls in love, the normal range of expected behaviours, feelings, and actions should follow. There has to be an object of that love (a real-world fact). And even though the peanut may be able to move about, at least some of the real-world rules of physics will constrain possible movements. Indeed, within a literary framework, Ryan (1991) has provided a thorough analysis of what constraints from the real-world have to hold in fictional worlds as a function of genres of writing. In short, mappings from the real world to the counterfactual world are inevitable if reasoning is to be generally possible, and for this reason, we would expect to be able to detect involvement of real world knowledge even within the context of the build-up of a counterfactual scenario.

Evidence has also been provided which implies that processing strategies are different for real-world and counterfactual-world information. The fact that RW inconsistencies are detected immediately in the eye-tracking record, extending into later measures, with lengthened reading times and increased regressions, implies that readers are seeking more information around the critical region following RW-inconsistent items, perhaps re-evaluating the model or reparsing the text. In contrast, Experiments 1 and 3 reveal that while the CW-consistent items also lead to initial problems in processing with lengthened reading times at the critical region, increased regressive eye-movements do not accompany this disruption. Thus, the slowed reading following CW-consistent conditions may reflect extra time required to integrate the counterfactual context. For example, a real-world situation develops strong expectations towards a real-world event, thus any violation of those expectations is likely to cause immediate and comprehensive reanalysis. In contrast, a counterfactual-world situation might produce 'digging-in' effects involving a series of attempted attachments to semantically relevant CW expectancies, which may in turn compete with typical attachments to RW expectancies. Tabor & Hutchins (2004) have suggested a similar model applied to syntax, where reanalyzing text becomes harder the more committed the parser has become to a particular syntactic choice.

To conclude, in three experiments we have shown that when a new counterfactual situation is asserted, people will readily assimilate new input into this situation. If a fact fails to fit the counterfactual world, then it disrupts eye-tracking. If the situation depicted is part of the real world, then a subsequent input that fails to match leads to a similar disruption of eye-tracking. Most interesting, however, is the finding that just prior to the accommodation of a proposition into a counterfactual world representation, the proposition is rapidly evaluated against real-world knowledge. This has the effect of delaying accommodation in the counterfactual world case. We conclude that even in the setting of a counterfactual world context, evaluation against real-world knowledge takes place, and have argued for the rationality of such a process. Finally, the very early influence of situation-specific information provides some very good evidence for the general argument that language comprehension is founded on the fast access to world knowledge.

References

Altarriba, J., Kroll, J., Sholl, A. & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477- 492.

Bartlett, F. (1932). *Remembering.* Cambridge: University Press.

Bower, G.H., Black, J.B., & Turner, T.J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177-220.

Braze, D., Shankweiler, D., Ni, W. & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31 (1),* 25- 44.

Byrne, R.M.J. (1997). Cognitive processes in counterfactual thinking about what might have been. *The Psychology of Learning and Motivation, Advances in Research and Theory,* 37. San Diego, CA: Academic Press. pp. 105-154.

Byrne, R.M.J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences.* 6, 10, 426-431

Byrne, R.M.J. (2007). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, M.A.: MIT Press.

Byrne, R.M.J. & Tasso A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition 27(4)*, 726-740.

Conolly, J.F., & Phillips, N.A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Cook, A.E. & Myers, J.L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language, 50*, 268-288.

Cowper, E. (1999). Grammatical aspect in English. *The Linguistic Review, 16*, 205-226.

Evans, J.St.B.T. & Over, D. E. (2004). *If.* Oxford, England: Oxford University Press.

Evans, J.St.B.T., Over, D.E. & Handley, S.J. (2005). Suppositions, Extensionality, and Conditionals: A Critique of the Mental Model Theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112 (4), 1040–1052

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*, Cambridge, MA: MIT Press.

Fauconnier, G. (1994). *Mental Spaces*. Cambridge, UK: Cambridge University Press.

Fauconnier, G. (1997). *Mappings in thought and language*, Cambridge University Press.

Fauconnier, G. & Turner, M. (2003). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* New York: Basic Books.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford: Blackwell.

Garnham, A. (1979). Instantiation of verbs. *Quarterly Journal of Experimental Psychology*, 31, 207-214.

Garrod, S. & Sanford, A. (1999). Incrementality in discourse understanding. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading*. (pp. 3–27). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Garrod, S.C. & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution*. Journal of Memory and Language, 42*, 526-544.

Happé, F. (1997). Central coherence and theory of mind on autism: Reading homographs in context. *British Journal of Developmental Psychology, 15*, 1-12.

Holcomb, P.J. & Neville, H.J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281–312.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, England: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* London: Erlbaum.

Johnson-Laird, P.N. & Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review.* 109, 646-678.

Kahneman, D. (1995). “Varieties of Counterfactual Thinking”. In *What Might Have Been: The Social Psychology of Counterfactual Thinking,* ed. Neal J. Roese and James M. Olson, Hillsdale, NJ: Erlbaum, 375–96.

Kahneman, D. & Miller, D.T. (1986). Norm Theory - Comparing Reality to Its Alternatives. *Psychological Review 93(2)*: 136-153.

Kahneman, D. & Tversky, A. (1982). The simulation heuristic, In Kahneman, D., Slovic, P. & Tversky, A. (eds.) *Judgement under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Kratzer, A. (1991). Modality. In A von Stechow & D. Wunderlich (Eds.). *Semantik: Ein internationals Handbuch der zeitgnössischen Forschung*, Berlin: Walter de Gruyter, pp 639-650.

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science*, *207*, 203-205.

Kutas, M. & Kluender, R. (1994). What is who violating? A reconsideration of linguistic violations in light of event-related brain potentials. In Heinze, H-J., Münte, T.F. & Mangun, G.R. (Eds.), *Cognitive electrophysiology.* Boston: Birkhäuser.

Kutas, M., Van Petten, C. & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler & M.A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd Edition*. New York: Elsevier.

Lewis, D. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, *1*0, 217-234.

McCall, S. (1984), Conterfactuals based on real possible worlds. *Noûs*, *18*, 462-477.

Markman, K.D. & Tetlock, P.E. (2000). 'I couldn't have known': Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology, 39 (3)*, 313-325.

Markovits, H. (1995). Conditional reasoning with false premises: fantasy and information retrieval. *British Journal of Developmental Psychology, 13*, 1-12.

Moggach, D. (2000). *Tulip Fever*. London: Vintage.

Ni, W., Fodor, J.D., Crain, S. & Shankweiler, D.P. (1998). Anomalous strings: Eye movement patterns. *Journal of Psycholinguistic Research, 27 (5)*, 515-539.

Nieuwland, M.S. & Van Berkum, J.J.A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18(7),* 1098-1111.

Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays.* London: Routledge & Kegan Paul.

Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* 124, 372- 422.

Rayner, K. & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory & Cognition, 15,* 779-790.

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

Rayner, K. & Well, A. D. (1996) Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504- 09.

Rizzella, M.L. & O'Brien, E.J. (2002). Retrieval of concepts in script-based texts and narratives: The influence of general world knowledge. *Journal of Experimental Psychology: Learning , Memory & Cognition, 28*, 780-790.

Ryan, M-L. (1991). Possible worlds and accessibility relations: A semantic typology of Fiction. *Poetics Today, 12*, 553-576.

Sanford, A. J. (1983). *The Mind of Man*. New Haven: Yale University Press

Sanford, A. J. & Garrod, S.C. (1981). *Understanding Written Language*. Chichester: John Wiley & Sons.

Sanford, A.J. & Garrod, S.C. (1998). The role of scenario mapping in text comprehension, *Discourse Processes, 26,* 159-190.

Sanford, A.J. & Garrod, S.C. (2005). Memory-based approaches and beyond. *Discourse Processes, 39,* 205–223.

Sanford, A.J. & Moxey, L.M. (1999). What are mental models made of in: Rickheit, G. & Habel, C. (Eds.) *Mental Models in Discourse Processing and Reasoning* North Holland: Elsevier.

Santamaria, C., Espino, O. & Byrne, R.M.J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology*: *Learning, Memory and Cognition, 31*, 1149- 1154.

Schustack, M. W., Ehrlich, S. F. & Rayner, K. (1987). The complexity of contextual facilitation in reading: Local and global influences. *Journal of Memory and Language,* 26, 322- 40.

Scott, F.J., Baron-Cohen, S. & Leslie, A. (1999). "If pigs could fly': A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology, 17*, 349-362.

Sereno, S.C., Brewer, C.C. & O'Donnell, P.J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science, 14,* 328-333.

Sternberg, R.J. & Gastel, J. (1989). Coping with novelty in human intelligence - an empirical-investigation. *Intelligence, 13 (2)*, 187-197.

Tabor, W. & Hutchins, S. (2004). Evidence for Self-Organized Sentence Processing: Digging-In Effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30 (2), 431–450.

Van Berkum, J.J.A., Hagoort, P. & Brown, C.M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11* (6), 657-671.

Von Fintel, K. (2001). Counterfactuals in a Dynamic Context. In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language*, MIT Press.

Table 1

Examples of experimental sentences (Experiments 1)

RW-inconsistent
If cats are hungry they usually pester their owners until they get fed.
Families could feed their cat a bowl of carrots and it would gobble it down happily.
Cats are loving pets when you look after them well.
RW-consistent
If cats are hungry they usually pester their owners until they get fed.
Families could feed their cat a bowl of fish and it would gobble it down happily.
Cats are loving pets when you look after them well.
CW-consistent
If cats were vegetarians they would be cheaper for owners to look after.
Families could feed their cat a bowl of carrots and it would gobble it down happily.
Cats are loving pets when you look after them well.

Table 2

Mean eye-movement measures for regions 1 to 4, Experiment 1 (standard errors in parentheses)

	Region				
	Context sentence 1	*1 Families could*	*2 feed their cat a bowl of*	*3 carrots and*	*4 it would gobble it down happily.*
First-pass reading time (ms)					
RW-inconsistent	3207 (159.5)	504 (27.2)	833 (43.7)	327 (11.9)	856 (48.2)
RW-consistent	3132 (153.1)	473 (27.7)	804 (39.6)	296 (9.7)	890 (33.0)
CW-consistent	3357 (162.7)	485 (31.1)	740 (28.4)	328 (12.4)	912 (40.7)
Regression path time (ms)					
RW-inconsistent	3207 (159.5)	589 (32.1)	1026 (40.4)	453 (24.5)	1379 (95.0)
RW-consistent	3131 (153.1)	567 (39.2)	1057 (57.9)	389 (17.3)	1074 (67.0)
CW-consistent	3357 (162.7)	577 (35.0)	977 (36.2)	413 (14.9)	1082 (69.2)
First-pass regressions out (%)					
RW-inconsistent	-	7.8 (1.6)	18.4 (2.4)	23.3 (3.0)	31.6 (3.9)
RW-consistent	-	6.9 (1.6)	17.3 (2.0)	21.8 (2.9)	11.8 (2.4)
CW-consistent	-	9.0 (1.4)	19.8 (3.1)	17.5 (2.6)	8.8 (2.2)
Total reading time (ms)					
RW-inconsistent	3398 (191.1)	612 (33.7)	1233 (67.5)	504 (30.4)	1184 (49.0)
RW-consistent	3345 (177.2)	555 (31.9)	1055 (52.7)	357 (14.8)	1029 (36.4)
CW-consistent	3545 (175.8)	604 (39.3)	985 (36.4)	385 (17.1)	1036 (50.2)
Regressions-in (%)					
RW-inconsistent	13.8 (2.1)	26.2 (3.0)	40.2 (3.6)	23.9 (3.2)	5.8 (1.3)
RW-consistent	15.6 (2.6)	22.3 (2.8)	24.4 (2.8)	9.7 (2.2)	6.4 (1.7)
CW-consistent	16.4 (2.6)	22.9 (3.4)	22.1 (3.0)	7.5 (2.0)	7.2 (1.5)

Figure 1

Mean first-pass reading times in critical region for Experiment 1, showing standard error bars.

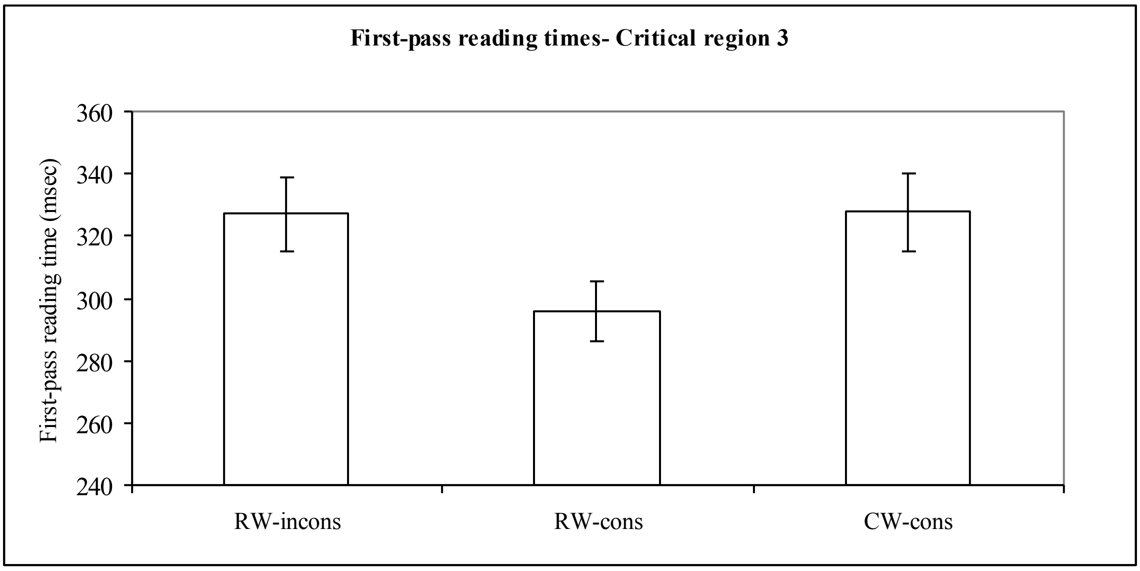

Figure 2

Percentage first-pass regressions out, Experiment 1

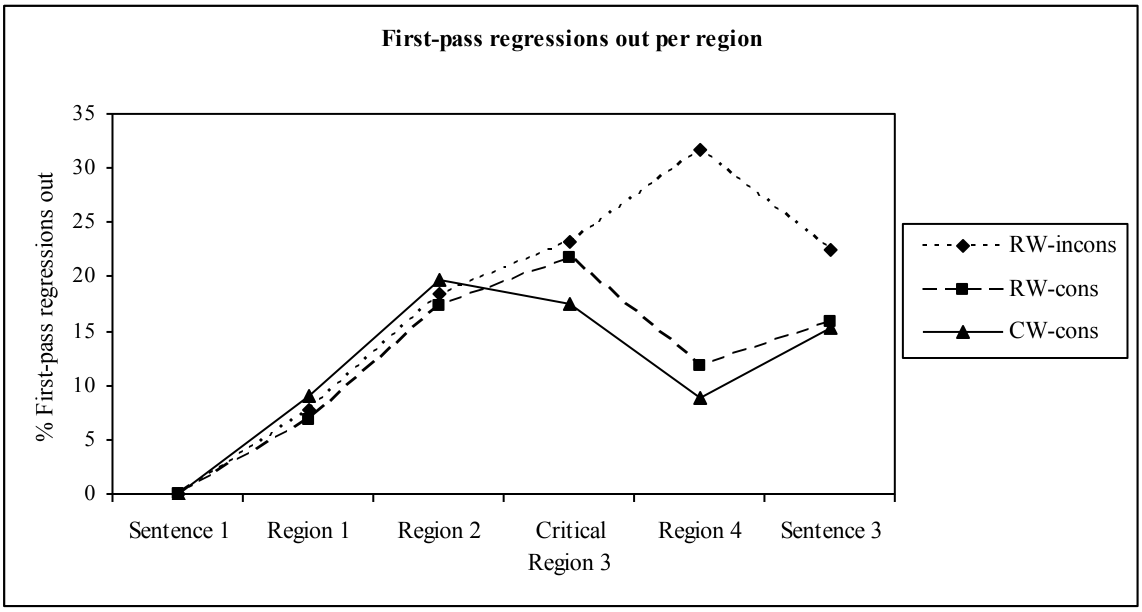

Figure 3

Percentage regressions in, Experiment 1

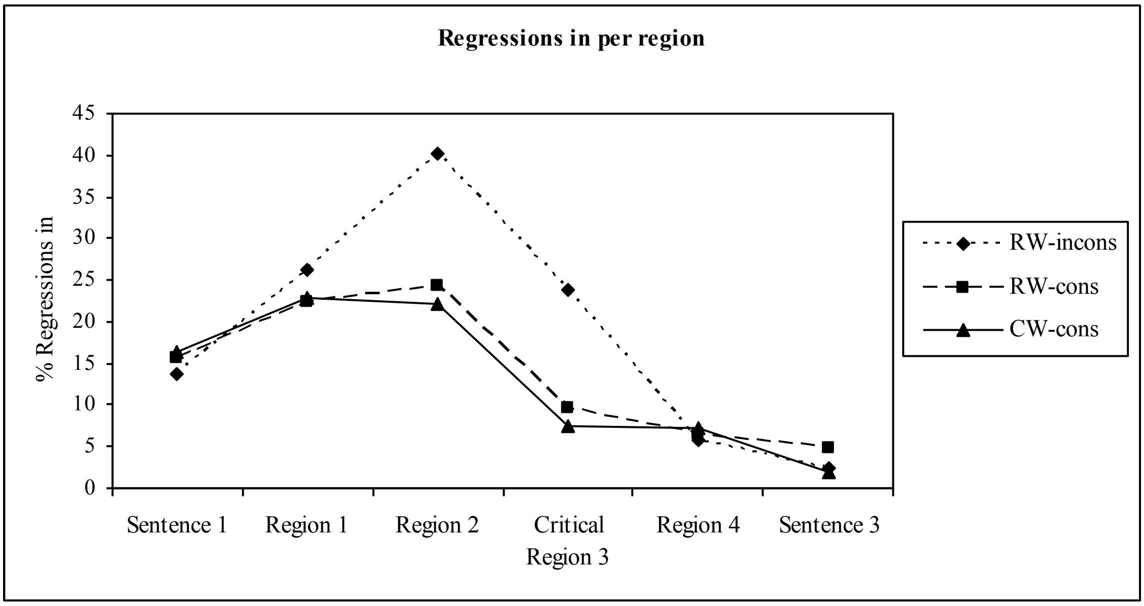

Table 3

Examples of experimental sentences (Experiment 2)

RW-inconsistent Evolution dictates that cats are carnivores and cows are vegetarians. Families could feed their cat a bowl of carrots and it would gobble it down happily.
RW-consistent Evolution dictates that cats are carnivores and cows are vegetarians. Families could feed their cat a bowl of fish and it would gobble it down happily.
CW-inconsistent If cats were vegetarians they would be cheaper for owners to look after. Families could feed their cat a bowl of fish and it would gobble it down happily.
CW-consistent If cats were vegetarians they would be cheaper for owners to look after. Families could feed their cat a bowl of carrots and it would gobble it down happily.

Table 4

Mean eye-movement measures for regions 1 to 4, Experiment 2 (standard errors in parentheses)

	Region				
	Context sentence 1	*1* *Families could*	*2* *feed their cat a bowl of*	*3* *carrots and*	*4* *it would gobble it down happily.*
First-pass reading time (ms)					
RW -inconsistent	3399 (132.2)	493 (24.7)	990 (37.9)	342 (14.3)	921 (41.5)
RW -consistent	3506 (121.2)	480 (22.3)	976 (32.9)	346 (15.5)	1008 (52.6)
CW -inconsistent	3686 (169.4)	482 (23.8)	960 (52.1)	344 (16.2)	988 (58.2)
CW -consistent	3667 (161.3)	485 (22.8)	931 (40.8)	343 (14.7)	955 (57.9)
Regression path time (ms)					
RW -inconsistent	3400 (132.2)	590 (29.1)	1179 (42.1)	421 (20.7)	1426 (95.6)
RW -consistent	3506 (121.2)	603 (47.8)	1218 (53.1)	382 (15.9)	1285 (69.6)
CW -inconsistent	3686 (169.4)	548 (27.3)	1139 (51.7)	391 (21.6)	1338 (77.8)
CW -consistent	3667 (161.3)	557 (35.6)	1105 (44.6)	392 (19.0)	1193 (62.9)
First-pass regressions out (%)					
RW -inconsistent	-	9.8 (1.9)	13.1 (2.0)	14.6 (2.6)	28.2 (3.9)
RW -consistent	-	10.2 (1.8)	15.2 (2.7)	8.1 (1.9)	23.6 (3.8)
CW -inconsistent	-	6.8 (1.7)	13.7 (2.0)	8.1 (1.8)	30.3 (3.6)
CW -consistent	-	8.4 (1.7)	11.5 (2.0)	10.1 (2.6)	22.3 (3.0)
Total reading time (ms)					
RW -inconsistent	3583 (137.1)	606 (27.9)	1298 (52.5)	427 (21.3)	1117 (59.7)
RW -consistent	3700 (121.0)	582 (30.5)	1227 (45.6)	379 (16.2)	1115 (58.2)
CW -inconsistent	3853 (172.3)	555 (27.3)	1175 (55.6)	399 (19.6)	1125 (56.1)
CW -consistent	3775 (166.6)	343 (24.9)	1109 (40.4)	394 (18.4)	1076 (57.2)
Regressions In (%)					
RW -inconsistent	21.3 (3.3)	21.4 (2.9)	23.9 (2.9)	10.9 (1.8)	4.6 (1.3)
RW -consistent	20.4 (3.1)	21.9 (2.9)	10.9 (2.4)	9.6 (1.9)	3.8 (1.4)
CW -inconsistent	18.6 (2.6)	19.7 (2.7)	17.2 (1.8)	12.5 (2.1)	2.5 (1.0)
CW -consistent	16.4 (2.5)	12.7 (1.9)	16.2 (2.8)	12.7 (2.1)	2.7 (1.2)

Figure 4

Percentage first-pass regressions out, Experiment 2

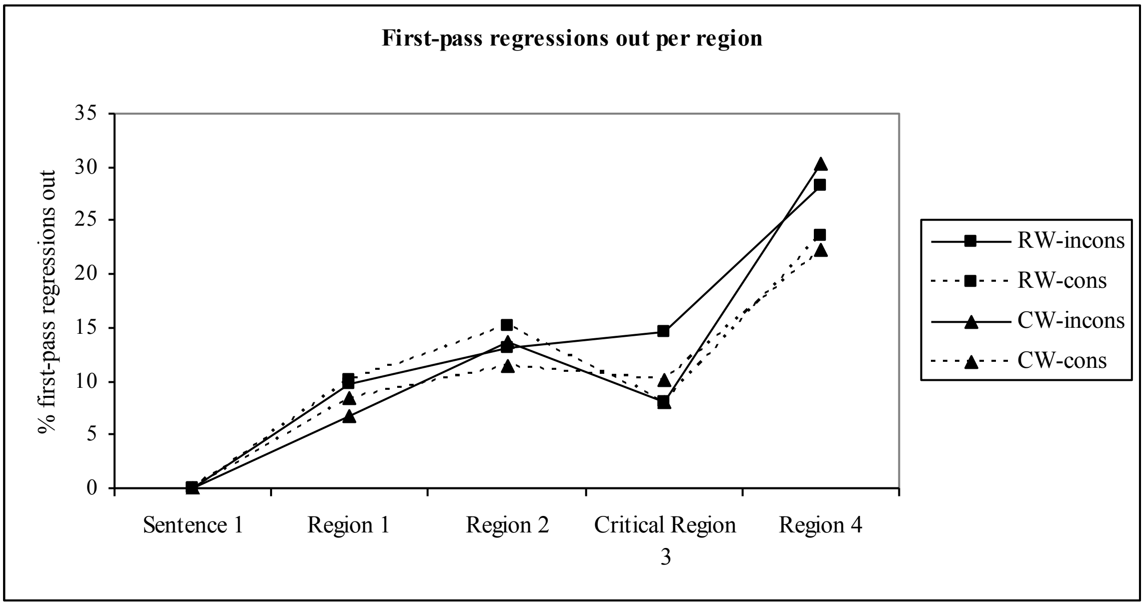

Table 5

Examples of experimental sentences (Experiment 3)

RW-inconsistent
If cats are hungry they usually pester their owners until they get fed.
Families would feed their cat a bowl of carrots and it would gobble it down happily.
RW-consistent
If cats are hungry they usually pester their owners until they get fed.
Families would feed their cat a bowl of fish and it would gobble it down happily.
CW-inconsistent
If cats were vegetarians they would be cheaper for owners to look after.
Families would feed their cat a bowl of fish and it would gobble it down happily.
CW-consistent
If cats were vegetarians they would be cheaper for owners to look after.
Families would feed their cat a bowl of carrots and it would gobble it down happily.

Table 6

Mean eye-movement measures for regions 1 to 4, Experiment 3 (standard errors in parentheses)

	Region				
	Context Sentence 1	*1* *Families could*	*2* *feed their cat a bowl of*	*3* *carrots and*	*4* *it would gobble it down happily.*
First-pass reading time (ms)					
RW-inconsistent	2710 (125.4)	471 (26.1)	779 (35.0)	302 (9.2)	861 (50.0)
RW-consistent	2699 (137.0)	465 (22.1)	785 (37.0)	288 (10.9)	845 (43.4)
CW-inconsistent	2736 (130.8)	462 (21.1)	766 (39.9)	285 (10.4)	893 (41.0)
CW-consistent	2816 (146.6)	483 (23.2)	766 (39.8)	324 (11.4)	870 (44.5)
Regression path time (ms)					
RW-inconsistent	2710 (125.4)	513 (33.4)	892 (43.5)	358 (15.5)	1119 (64.0)
RW-consistent	2699 (137.0)	496 (26.4)	877 (45.1)	338 (16.4)	1089 (74.7)
CW-inconsistent	2736 (130.8)	518 (26.1)	868 (43.8)	344 (21.0)	1158 (62.6)
CW-consistent	2816 (146.7)	517 (26.1)	901 (39.3)	367 (14.5)	1063 (62.2)
First-pass regressions out (%)					
RW-inconsistent	-	3.3 (1.2)	9.1 (2.0)	10.7 (2.3)	22.2 (2.3)
RW-consistent	-	3.3 (1.2)	7.8 (1.8)	12.3 (3.2)	17.1 (3.1)
CW-inconsistent	-	5.1 (1.3)	8.3 (2.0)	11.3 (2.6)	24.1 (3.7)
CW-consistent	-	4.2 (1.1)	11.0 (2.0)	11.9 (2.4)	17.9 (2.7)
Total reading time (ms)					
RW-inconsistent	2782 (130.5)	526 (30.4)	944 (47.7)	360 (17.0)	972 (50.9)
RW-consistent	2781 (143.0)	518 (26.0)	919 (48.3)	327 (14.4)	940 (52.7)
CW-inconsistent	2842 (133.5)	529 (23.9)	916 (45.7)	327 (13.4)	994 (47.6)
CW-consistent	2899 (154.6)	544 (26.0)	916 (41.3)	353 (11.1)	928 (48.7)
Regressions In (%)					
RW-inconsistent	8.8 (2.0)	12.5 (2.1)	19.8 (3.0)	12.7 (2.6)	0.7 (0.5)
RW-consistent	9.6 (2.2)	11.3 (2.2)	13.2 (2.3)	8.0 (1.9)	0.7 (0.5)
CW-inconsistent	13.9 (2.2)	15.6 (3.0)	15.6 (2.6)	8.1 (1.8)	0.6 (0.4)
CW-consistent	10.0 (2.3)	17.4 (2.5)	13.5 (2.5)	6.1 (1.6)	0.7 (0.5)

Figure 5

Mean first-pass reading times in critical region for Experiment 3, showing standard error bars.

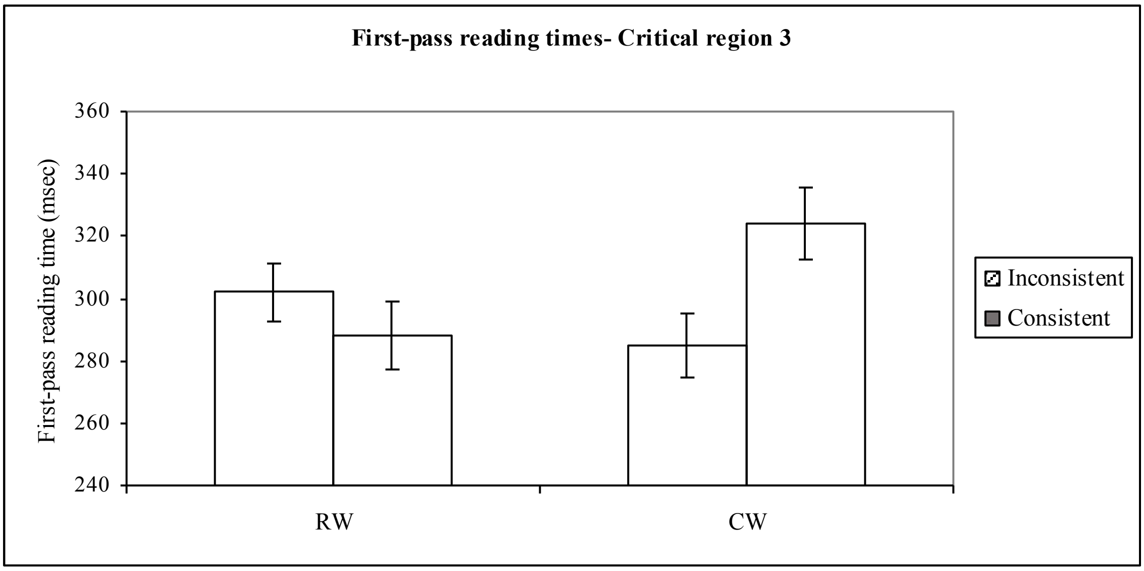

Figure 6

Percentage first-pass regressions out, Experiment 3