Eye Movements and Display Change Detection during Reading

Timothy J. Slattery, Bernhard Angele, and Keith Rayner

Department of Psychology

University of California, San Diego

Correspondence to:

Timothy J. Slattery
Department of Psychology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093
Email: tslattery@ucsd.edu

Running head: Eye Movements and Display Change Detection

Abstract

In the boundary change paradigm (Rayner, 1975), when a reader's eyes cross an invisible boundary location, a preview word is replaced by a target word. Readers are generally unaware of such changes due to saccadic suppression. However, some readers detect changes on a few trials and a small percentage of them detect many changes. Two experiments are reported in which we combined eye movement data with signal detection analyses to investigate display change detection. On each trial, readers had to indicate if they saw a display change in addition to reading for meaning. On half the trials the display change occurred during the saccade (immediate condition); on the other half, it was slowed by 15-25 ms (delay condition) to increase the likelihood that a change would be detected. Sentences were presented in an alternating case fashion allowing us to investigate the influence of both letter identity and case. In the immediate condition, change detection was higher when letters changed than when case changed corroborating finding that word processing utilizes abstract (case independent) letter identities (McConkie & Zola, 1979; Rayner, McConkie, & Zola, 1980). However, in the delay condition (where d' was much higher than the immediate condition), detection was equal for letter and case changes. The results of both experiments indicate that sensitivity to display changes was related to how close the eyes were to the invalid preview on the fixation prior to the display change, as well as the timing of the completion of this change relative to the start of the post change fixation.

One of the most well-established findings regarding reading is that readers obtain preview benefit from the word to the right of fixation (Rayner, 1998, 2009). When readers have a valid preview of word n+1, they look at it for 20-50 ms less time than when they do not have a valid preview. Virtually all of the research that documents preview benefit in reading derives from the utilization of the gaze contingent boundary paradigm (Rayner, 1975). In this paradigm, readers' eye movements are monitored and a target location is designated in the sentence or text that is read on a video monitor (with a fast decaying phosphor and rapid refresh rate). Prior to crossing an invisible boundary (which is generally located just before the target word), the target word is replaced by a preview stimulus. The nature of this preview stimulus can be manipulated so that it either preserves certain properties of the target word or it does not. When the reader's eyes cross the invisible boundary location, the preview stimulus is replaced by the target word (which remains visible throughout the remainder of the trial). The impact of (1) the nature of the preview and (2) how far the eyes were from the target word location on the prior fixation on fixation times on the target word can then be examined.

Because of saccadic suppression (Matin, 1974) readers typically are not aware of, nor do they see, the display change. The patterns of results from studies using the boundary paradigm have been rather consistent, but there have been two controversial issues. First, while it is very clear that readers obtain preview benefit from word n+1 (the word to the right of fixation), there is some controversy regarding the extent to which they obtain preview benefit from word n+2 (see Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Kliegl, Risse, & Laubrock, 2007; Rayner, Juhasz, & Brown, 2007; Angele & Rayner, 2011; Glover, Vorstius & Radach, 2011; Risse & Kliegl, 2011). If the saccade target is word n+2, there is less controversy than when readers move their eyes first to word n+1 and then to word n+2. Studies dealing with word n+2 preview benefit have been viewed as critical in adjudicating between serial lexical processing models, such as E-Z Reader (Reichle, Fisher, Pollatsek, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006) and parallel lexical processing models, such as SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005) and Glenmore (Reilly & Radach, 2003; 2006). While this debate is interesting, we will only touch upon it briefly in the present article. However, it is at least possible that some of the differences between studies in terms of the patterns of results could be due to some readers being more aware than others of the display change (given that two words, rather than just one word, changes during the critical saccade).

The second controversy has to do with display change artifacts. Specifically, each time a display change occurs (from the preview stimulus to the target word), because of phosphor persistence of letters on the display monitor or the refresh rate of the display monitor (or both), there is a flash on the screen from replacing letters by other letters. O'Regan (1990, 1992) argued that the display change per se might somehow influence the results of boundary paradigm studies or that the amount of change might artifactually influence the data pattern. Thus, according to his argument, the more letters that are replaced, the more disruption there will be, or, given that the number of letters that change is held constant, either the location of the changing letters (closer to or further from fixation) or the characteristics of the changing letters (replacing letters with visually similar letters might cause less of a flash than replacing letters with dissimilar letters) might influence the results.

The present experiments deal more directly with the second controversy than the first (though as we noted above differences in studies related to the first controversy could be related to some subjects being aware of the display change). To return to the second controversy, to what extent is O'Regan's argument an issue in gaze-contingent display change experiments utilizing the boundary paradigm? The currently used display change techniques generally are able to implement the change in something like 8-17 ms, which typically means that the change can be completed during a saccade. Saccades in reading typically take 20-35 ms (depending on the size of the saccade; as saccade size increases, saccade duration increases in a roughly linear fashion), so in principle most display changes *should* occur during saccades when vision is suppressed[1]. If the display change is slowed down so that it takes place during a fixation, readers are aware of the changing letters (Rayner, 1998). Under normal experimental conditions, even though readers are encouraged to report any strange events at the completion of the experiment, they rarely do so, and even skilled observers who know what is happening often do not perceive the changes. On the other hand, it could be argued that it does not matter whether or not readers are consciously aware of the display changes, because there is a change in the stimulus pattern and it is registered by the visual system. For instance, it is known that such changes will induce saccadic inhibition (Reingold & Stampe, 1999, 2000, 2004) if they occur within a fixation. While, the changes in boundary change studies usually occur during a saccade, it remains unknown whether such changes are unconsciously registered by the visual system and interact with the preview manipulations being made. However, Rayner (1998) noted that if so, the interaction is very complex and not easily interpretable and what data there are seem inconsistent with a “flicker” explanation. More critically, in some important experiments designed to examine this issue, Inhoff, Starr, Liu, and Wang (1998 (see also Briihl & Inhoff, 1995) varied the speed of the display change and refresh rate of the display monitor (and found no evidence to suggest that the results of eye-contingent display change experiments were artifacts of the paradigm.

Yet, there are some subjects who do “see” (or are aware of) the display change. White, Rayner, and Liversedge (2005) reported an experiment in which they compared subjects who reported noticing the display change with those who did not. Subjects were categorized into these groups by their response to the question “did you notice anything strange about the appearance of the text?” at the end of the experiment. Among those (n = 16) who were categorized as noticing the change, some reported noticing nonsense letter sequences and some were not aware of exactly what had changed. Some reported noticing something only occasionally, whereas others reported that they often noticed something odd. The other group of subjects (n = 32) were not aware of anything unusual happening during their reading. Importantly, the subjects who were aware of something changing produced a pattern of data that was qualitatively different from those who were not aware of the display changes.

In the present experiments, we introduce a new and more precise way of examining the issue of display change detection. Specifically, we used a signal detection paradigm (Macmillan & Creelman, 2005) wherein on each trial subjects had to indicate if they saw something change in the sentence they were reading. In addition, after making a yes-no judgment about whether or not there was a display change, they had to answer a question about the sentence. Previously, Rayner (1978) used a signal detection and d’ analysis to examine preview benefit in a gaze-contingent type of paradigm. However, in that experiment subjects were not reading meaningful material (as they were in the present study). The present research has both methodological and theoretical implications. At the methodological level, we hoped to gain a better understanding of the extent to which readers are aware of display changes and how this awareness impacts on the ensuing data pattern. At the theoretical level, the present work has implications for attentional processes in reading.

Experiment 1

In Experiment 1, subjects read sentences, presented in alternating case, for comprehension. After every trial they were asked whether or not they noticed a display change. These changes involved a change in letter case (i.e. word shape), letter identity, or a combination of the two.

Method

Subjects

Sixteen undergraduate students at the University of California San Diego participated for course credit. All were native speakers of English, had either normal or corrected to normal vision, and were naïve concerning the purpose of the experiment.

Apparatus

An SR Research Eyelink 1000 eyetracker was used to record subjects' eye movements with a sampling rate of 2000 Hz (i.e. the eye position was sampled twice every millisecond). Subjects read sentences displayed in black letters on a white background in 14 pt Courier New font on an Iiyama Vision Master Pro 454 video monitor with a refresh rate of 150 Hz at a resolution of 1024 by 768 pixels. Viewing was binocular, but only the movements of the right eye were recorded. Viewing distance was approximately 60 cm, with 3.5 letters equaling one degree of visual angle.

Materials.

One hundred and sixty experimental and 60 filler sentences were adapted from the materials used by Angele et al. (2008). For each experimental sentence, an invisible boundary was defined. The boundary was always located between the last letter of the pre-boundary word (*n-1*) and the subsequent space. Table 1 shows length and frequency (determined from the CELEX count using the N-Watch software, Davis 2005) of the pre-boundary word (word *n-1*) as well as the first (word *n*, or target word), and second (word *n+1*) post-boundary words.

Insert Table 1 about here

In order to manipulate word shape and letter identity of the target word preview independently, we presented every other word of both experimental and filler sentences in alternating case (see Table 2). In the experimental sentences, word n was always presented in alternating case and word *n-1* as well as word *n+1* was always presented in lower case. The preview of the target word (word n) that subjects received prior to fixating it either had the same case as the actual word or reverse case (e.g. eXaMpLe vs. ExAmPlE). This case manipulation was counterbalanced such that the first letter of each item was capitalized half of the time and in lower case the other half of the time. Additionally, the preview either consisted of the same letters as the target word (letter same condition) or random letters (letter different condition) generated using the same algorithm employed in Angele et al. (2008) and Kliegl et al. (2007). Note that the previews in the letter different condition were always non-words. The complete experimental design was a 2 (first letter case: upper vs. lower) X 2 (delay: 0 ms vs. 25 ms) X 2 (case change: yes vs. no) X 2 (letter change: yes vs. no) repeated measures design. However, the manipulation of the case of the first letter was only done to prevent subjects from developing a strategy and had little theoretical value for our current purposes. Thus, we averaged over this variable for all subsequent analyses[2].

Once a readers' eyes crossed the boundary, the preview was replaced by the target word after the delay period of zero or 25 milliseconds. Note that for three quarters of the experimental conditions the letters present at location *n* during the preview were different (in either case, letter identity, or both) from those present after the boundary was crossed. These trials will be referred to as "display change" trials. The other quarter of the experimental conditions also technically included a boundary change, however in these conditions the letters present at location *n* during the preview were the same as those present after the boundary was crossed. In these trials there was no detectable change on the display monitor; these trials will be referred to as "identical" trials. These identical trials (along with the filler trials) were used to estimate false alarm rates for the d` calculations. There were 120 trials in which there was a detectable change during reading and 100 trials in which there was no change (40 experimental item trials and 60 filler item trials). Table 2 shows examples of the preview conditions.

In most gaze-contingent display change experiments, target words are placed in the middle of sentences or close to the middle. However, in the current experimental paradigm doing this could cause subjects to develop attentional strategies. In order to prevent them from adopting such strategies, the location of the target word in the experimental sentences was more variable than in most studies. This was accomplished by having the target word occur equally often in the first, second, and last third of the word positions within the sentences (excluding the first two word positions and the final word position). This additional manipulation prevented the adoption of any attentional strategy that favored the middle words of each sentence.

Insert Table 2 about here

Procedure

Subjects were presented with the 160 experimental sentences along with the 60 filler sentences in random order. Every sentence presentation was preceded by a gaze target. Once subjects fixated the gaze target, the target disappeared and the sentence was displayed on the screen. The location of the gaze target always matched the subsequent location of the first word in the sentence. After every trial, subjects were prompted to state whether a display change had occurred while they were reading. They answered the two-alternative forced choice prompts by pressing one of two buttons on a button box. Approximately 50% of the prompts were followed by a comprehension question that subjects answered by pressing the button corresponding to their answer choice. This is a higher percentage of comprehension questions than is standard but we wanted to stress the importance of reading for understanding in the current study. Before the experiment, subjects practiced responding to the detection prompt and the comprehension questions during 10 practice trials, 50% of which contained display changes. Outside of the practice trials, subjects did not receive feedback on their responses.

Custom-made software ensured that display changes were executed either as fast as possible (approximately 8 ms on average) after crossing the boundary (no delay condition) or with a 25 ms delay (25 ms delay condition). Since such a delay almost invariably resulted in the display change occurring during a fixation, we expected subjects to detect more of the changes in the 25 ms delay condition than in the no delay condition.

Results

Trials in which there was a blink or track loss on the target word or during an immediately adjacent fixation were removed prior to analysis (2.3% of trials), as were trials in which the display change was triggered early by a saccade that ended left of the boundary[3] (7.0% of trials). Also, trials in which the display change failed to occur or occurred more than 25 ms after the delay period were removed (1.8% of trials). Fixations shorter than 80 ms, which were within 1 letter of a previous or subsequent fixation, were combined with that fixation; all other fixations less than 80 ms were eliminated (1.9% of all fixations). Data loss affected all conditions similarly ($F < 1$). Additionally, the average accuracy to the comprehension questions was 93% and did not differ significantly by experimental condition, $F < 1$. This high accuracy to the comprehension questions is important and indicates that the subjects were primarily engaged in reading for understanding.

The remainder of the results will be broken into three sections. The first section will consist of analysis of the d` data. The second section will consist of analyses of three standard measures of eye movements (Rayner, 1998): first fixation duration (the duration of the first fixation on a word), gaze duration (the sum of all fixations before moving to another word), and skipping rate. We will also present an analysis of the fixation duration immediately prior to crossing the display change boundary in order to assess potential parafoveal on foveal effects. The final section will consist of a series of post hoc analyses designed to explore the relationship between eye movements and display change detection.

Analysis of d' Data

For each subject, hit rates were calculated for each display change condition, and a single false alarm rate was calculated from the combination of the identical conditions and filler items. Note that the use of identical trials was necessary to estimate false alarm rates and to provide a baseline for which to compare eye movement measures of reading such as gaze duration. However, this also means that we only have d` values for 6 of our 8 conditions (d` requires both a hit rate and a false alarm rate). These hit and false alarm rates were used to calculate d` with the formula Z(hit rate) – Z(false alarm rate). In the current experiment the false alarm rate was generally quite low and occasionally hit rates were quite high. There are problems calculating d` when hit rate is equal to one and/or false alarm rate is equal to zero as Z(1) = -infinity, and Z(0) = infinity. However, there are standard corrections that work well in these situations. We used a correction in which a half of a false alarm was added to a subject's mean when they made no false alarms for a condition and we subtracted half a hit when their hit rate was equal to one. The mean d` values after these corrections are shown in Table 3 along with the means for the various eye movement measures.

Insert Table 3 here

The d` values were submitted to a 2 (delay: 0 or 25 ms) X 3 preview (letter same case changed, letter changed case same, letter and case changed) ANOVA with subjects as a random factor[4]. As predicted, there was a large effect of delay as d` was more than 3 times as large when the display change was delayed by 25 ms than when it occurred immediately, $F(1,15) = 174.95$, $\eta_p^2 = .92$, $p < .001$. Additionally, there was a main effect of preview condition, $F(2,30) = 12.38$, $\eta_p^2 = .45$, $p < .001$. Bonferroni adjusted pairwise comparisons confirmed that the condition in which the letters remained the same and only the case was changed elicited lower d`s than the letter change or letter and case change conditions ($ps < .05$). However, these latter two conditions were not significantly different from each other $t < 1$. That is, changing the case of the letters provided no additional signal beyond changing the identity of the letters. There was also a significant interaction between the delay and preview conditions, $F(2,30) = 10.16$, $\eta_p^2 = .40$, $p < .001$. Bonferroni adjusted contrast comparisons indicated that this interaction was due to a larger effect of the delay for the condition in which the letters remained the same and only the case changed ($ps < .05$) than for the conditions in which the letters changed as well. In fact, there was no indication that the various invalid previews differed in their d` values at the 25 ms delay ($F < 1$). The results for the immediate conditions indicate that detection of these display changes are sensitive to the letter identities of the preview and target, which agrees with the results of McConkie and Zola (1979) and Rayner et al. (1980) who used an alternating case manipulation and changed the case of all the letters from fixation to fixation. They found that this had virtually no effect on reading times and argued that lexical processing occurs through the use of abstract letter identities where case is ignored. This argument is strengthened by the current results. Additionally, the interaction of preview condition with delay indicates that the preview differences in d` were not due to a larger visual overlap between upper and lower case letters when they were identical than when they were different (as this should have resulted in an effect of preview even in the delay condition).

Analysis of Eye Fixation Data

The eye movement data for word *n* were submitted to a linear mixed model (LMM) analysis using the lme4 package (Bates & Maechler, 2010) for the R statistical software (R Development Core Team, 2010). This model contained delay (0 vs. 25 ms), case change (yes vs. no), letter change (yes vs. no), and the interactions of these predictors as fixed effects and subjects and items as crossed random effects. These means appear in Table 3. For the fixed effects, we report coefficient and standard error estimates as well as p-values estimated from Markov chain Monte Carlo (MCMC) simulations (see Baayen, 2008 for a discussion as to why MCMC methods are preferred to estimate p-values for this type of analysis)[5].

We first assessed potential parafoveal on foveal effects with an analysis of the duration of the fixation immediately prior to crossing the boundary. These data were submitted to an LMM that contained the manipulated elements of the preview, letter case and letter identity, as fixed effects and subjects and items as random effects. There was a main effect of letter identity, $b = -3.74$, $SE = 1.84$, $p = .038$, as these fixations were shorter when the preview letters did not form a word. No other effects approached significance, $ps > .38$.

On average, the target word (n) was initially skipped on 2.9% of the trials. In a logistic LMM analysis with preview letter case and preview letter identity as fixed and subjects and items as random effects, the only significant effect on skipping rate was whether the preview consisted of a word or a non-word with higher skipping rates for word previews (identical letter preview conditions) than non-word ones (letter change preview conditions) which would be expected if subjects were reading for meaning (b = -

0.46, SE = .16, p < .01). None of the other experimental manipulations influenced skipping rates (ps > .15). We averaged over the delay variable in this analysis since the display change had not yet occurred at the time that saccades were being planned to skip the target. The overall low skipping rate is likely due to the use of 50% non-word previews, the novel appearance of the alternating case, and the inclusion of some long target words (10 letters).

There was a high degree of similarity between the first fixation and gaze duration data and the LMMs indicated the same pattern of significance values for the two measures. For brevity, we present only the LMM results for gaze duration here. Gaze durations were longer in the delay condition than in the no delay condition (b = 37.4, SE = 2.77, p <. 01). They were also longer when the case changed (b = 16.08, SE = 2.76, p < .01), or when the letters changed (b = 40.19, SE = 2.77, p < .01). Additionally, the three way interaction and all three of the two way interactions were significant (all ps < .008).

It is perhaps easiest to explore this seemingly complex data pattern by fitting separate models for the 0 ms and the 25 ms delay conditions. First consider the no delay conditions: there was no significant difference between the conditions in which the letters remained the same (identical vs. case change only conditions: b = 5.94, SE = 3.43, p > .05). There was, however, a difference between the conditions in which the letters changed and those conditions in which they did not (b=28.89, SE = 3.44, p < .01). The interaction between letter change and case change was not significant in the no delay condition (b = -0.41, SE = 3.44, p > .05). Thus for the no delay conditions, it would appear that as long as the identity of the letters remained the same (independent of case) before and after the boundary was crossed there was little if any effect on the fixation times.

These results contrast with those of the 25 ms delay condition. Here there was a difference between the two conditions in which the letters remained the same, ($b = 25.87$, $SE = 4.22$, $p < .01$), likely due to the fact that in the case change only condition there was a highly visible change occurring within a fixation that was absent in the identical condition. As with the no delay condition, both of the letter same conditions differed from each of the letter change conditions ($b = 51.24$, $SE = 4.22$, $p < .01$). The interaction between the case and letter change conditions was also significant ($b = -29.69$, $SE = 4.22$, $p < .01$), suggesting that, if the letter identity changed, an additional change in case did not result in longer fixation times. Therefore, the results of the 25 ms delay conditions are similar to those of the no delay condition with one main difference: there was an apparent cost in the form of much longer fixation durations for all conditions in which there was a visible change during fixation. Crucially, there was an additional cost for having different letters before and after the boundary was crossed indicating a preview benefit for the case change only condition over the letter change conditions even at this disruptive delay.

Post Hoc Analyses

Next, we explored the relationship between detecting a display change and various aspects of the eye movement record. For instance, there are a number of factors that may play a role in a subject's ability to detect a display change on a given trial. We explored five such variables: (1) the duration of the fixation immediately prior to crossing the boundary, (2) the proximity of this pre-change fixation to the boundary (in pixels), (3) the timing of the completion of the display change relative to the start of the post boundary change fixation, (4) the duration of the post boundary change fixation, and (5) the proximity of this post-change fixation to the boundary (in pixels). Due to the extremely high hit rate in the 25 ms delay condition (91.6%) we restricted these analyses to the no delay letter change conditions, which had an overall hit rate of 24.8%. The means for these variables in the two conditions of interest appear in Table 4. Since detecting or not detecting a change on a given trial is a binary outcome, we used a logistic LMM to model the influences of these variables on hit probabilities. Our regression equation predicted hit probability from the independent factors mentioned above which appear in Table 4 using subjects and items as random effects. According to this analysis the closer the subjects' fixation was to the boundary before crossing it, the more likely they were to correctly respond "changed" ($b = -.017$, $SE = .006$, $p = .006$). There was also a positive relationship between the duration of the post boundary change fixation and subjects likelihood of correctly responding "changed" ($b = .008$, $SE = .001$, $p < .001$). However, the other three variables did not significantly affect the likelihood of correctly responding "changed".

Insert Table 4 here

Given the influence that delaying the display change, and the proximity of the pre-boundary change fixation, had on hit rates we wanted to determine if sensitivity to detect display changes would be all but eliminated in a 'clean' data set. Therefore, we recalculated d` for the 0 ms delay conditions after excluding cases in which the pre-change fixation was within 11 pixels of the boundary (11 pixels equals 1 letter) and cases in which the change completed after the start of the post boundary change fixation. This exclusion procedure removed an additional 31% of the trials across the 0 ms delay conditions. However, although the overall d` values decreased somewhat (reduced by .03 on average), the overall pattern was unchanged. Additionally, the d` for the case change only condition was still less than for the letter change only condition, $t(15) = 4.38$, $p < .001$, or the case and letter change condition, $t(15) = 4.75$, $p < .001$, and the case and letter change condition did not differ from the letter change only condition, $t < 1$. Therefore, the detection of changes was not simply due to 'artifacts' in the data, nor was the pattern of readers' sensitivities to such changes.

We also addressed another issue that is directly related to O'Regan's criticism of display change methodology. That is, to what extent does the existence and pattern of preview effects in fixation times depend on a reader's awareness of or sensitivity to preview changes? We conducted two further LMM analyses using the data from the no delay conditions. The first predicted target gaze duration from the letter identity and letter case factors as well as their interaction, including subjects and items as random effects. This analysis yielded a main effect of changing the letters, $b = 51.153$, $SE = 9.855$, $p = .0001$, but there was no effect of changing the case nor was there an interaction, $ps > .60$. Next we repeated this analysis after excluding all the 'hit' and 'false alarm' trials. As with the first analysis there was again a main effect of changing the letters, $b = 34.102$, $SE = 10.124$, $p = .001$, but again there was no hint of a main effect of changing the case nor was there an interaction, $ps > .90$. Therefore, while display change detection did increase the size of the preview effects, it was not the sole cause of these effects nor did it change the pattern of effects over the different preview change conditions.

Discussion

There are a number of important d` results from Experiment 1. The least surprising of these is that sensitivity to detect display changes was much better when the display change was delayed such that the change occurred after the start of the post boundary fixation. However, the finding that delaying the change interacts with the relationship between the preview and target characteristics is very interesting and novel. It is perhaps easiest to examine this interaction by considering the effect of the relationship between preview and target at the 0 ms and 25 ms delay conditions separately. When the display change was initiated immediately upon the eyes crossing the boundary, subjects' sensitivity to detect changes was strongly impacted by the relationship between the preview and target characteristics. That is, simply changing the case of all the letters between the preview and the target word resulted in very low d` values compared to changing the identity of the letters. Additionally, if letter identities were changed there was no additional increase in d` values for changing the case of the letters as well. This finding strongly suggests that at the 0 ms delay, when saccadic suppression can be assumed, detecting such changes is the result of comparisons between preview and target characteristics, in which abstract letter identities play an important role with minimal influence of purely visual input (Rayner et al., 1980). This can be starkly contrasted with the results of the 25 ms delay condition in which saccadic suppression is far more limited if operating at all. Here, sensitivity is near ceiling for all dissimilar preview conditions suggesting that the comparison of preview with target is based on purely visual input with little to no impact of abstract letter identities.

The gaze duration results were similar to the d` results as conditions with larger d` also tended to have longer gaze durations. However, there was a crucial difference in gaze durations between the preview conditions that did not follow this simple pattern. When the letters in the preview and target were the same there was clear evidence of preview benefit regardless of the delay condition. This is perhaps best exemplified in the 25 ms delay condition where d` was equally large in all of the dissimilar preview conditions. Here the gaze durations were significantly shorter in the condition in which only the case of the letters differed between preview and target than in either of the conditions in which the identity of the letters differed between preview and target. This suggests that while the visual change, which occurred within fixation, disrupted eye movements, readers were still able to make use of the accurate preview letter identities to aid in lexical access of the target word. We take this as further evidence that subjects in the current experiment were primarily engaged in the task of normal reading rather than simply attending to the text for the purpose of detecting display changes. Additionally, this visual disruption in the delay conditions appears similar to saccadic inhibition (Reingold & Stampe, 1999, 2000, 2004). Saccadic inhibition is the finding that sudden visual onsets disrupt the programming of saccades beginning roughly 100 ms post onset. Evidence for saccadic inhibition in studies with onsets that are time locked to the start of fixations can be seen as dips in fixation time distributions. In Figure 1, the fixation time distributions (with 50 ms bins) are presented for various conditions from Experiment 1. As can be seen, the delay conditions involve such distributional dips beginning around 130 ms (~100 ms post onset) relative to their respective no-delay distributions. However, relative to the no-change fixation distribution, the no-delay distributions aren't nearly as influenced by apparent saccadic inhibition. What's more, the case change only and the no-change distributions are nearly identical despite a very large visual change occurring as the eyes cross the boundary. This further supports the notion that as long as the change occurs within fixation subjects can be largely unaware of it depending on the relationship between the preview and target letters. The issue with display change timing and saccadic inhibition is both important and complex. Saccadic inhibition has been shown to be tightly linked to the timing of the visual onset (display change relative to fixation onset for our discussion). If the underlying fixation time distributions are different at the time point of saccadic inhibition over experimental conditions, then saccadic inhibition would be expected to have very a different influence on mean fixation durations for these conditions. Thus late display changes have the potential to distort data in ways that may not be easy to assess from mean fixation times, which further highlights the importance of well controlled timings in such display change studies.

Insert Figure 1 here

Finally, the post hoc analyses, which examined the influence of eye movements on hit rates, confirmed what many researchers in the field likely believed all along. That is, subjects were more likely to notice such changes the closer their fixation was to the invalid preview prior to triggering the change. This result may have some implications for n+2 studies of preview benefit. In both the Angele et al. (2008) and the Angele and Rayner (2011) studies, the assumption was that the comparison between the condition in which both word n+1 and n+2 were masked and the condition in which word n+1 was masked, but n+2 was available is a more valid test for effects of n+2 preview than the comparison between the condition in which only n+2 was masked and the condition in which both previews were available. The reason for this was that failed skippings of word n+1 can easily lead to mislocated fixations, resulting in apparent effects of n+2 preview on word n+1. However, our findings in the present study show that subjects are much more likely to detect a display change when they were fixating near a preview that subsequently changed. In the n+1 masked conditions of the aforementioned experiments, this was clearly the case. Also, in the present study, detecting the display change resulted in inflated fixation durations. As a consequence, possible effects of n+2 preview may have been masked by subjects detecting the changes. It is important to mention, however, that in both n+2 studies by Angele and colleagues, data from subjects who saw a significant number of display changes were removed from the analysis, limiting the impact of this issue. However, in the current study, d` values were only slightly reduced after excluding trials with 'artifacts', and the pattern of means across the experimental conditions was not affected. Additionally, significant preview effects did not depend on the detection of display changes.

While it would appear from Experiment 1 that the likelihood of detecting a change would be greater anytime the letters in the preview are different from those of the target, in Experiment 1 the previews in these conditions always consisted of nonword strings. Therefore, the increase in d` for the letter change conditions could have been the result of nonword previews. We address this possibility in Experiment 2 by including both word and nonword letter change preview conditions.

Experiment 2

In Experiment 2 we were primarily interested in testing the difference between two different letter change preview conditions (words vs. nonwords). This comparison allowed us to better interpret the results of Experiment 1. However, we were also interested in further exploring the influence that delaying the display change has on sensitivity and eye movements. Therefore, in Experiment 2 we used a 15 ms delay condition instead of the 25 ms delay condition that had been used in Experiment 1 in hopes that this would result in hit rates and d` values below the near ceiling values obtained in Experiment 1. Finally, in Experiment 2 we collected confidence-rating data for detecting a display change in order to construct Receiver Operating Characteristic (ROC) curves. These data have a number of advantages over simple binomial hit rate data and allow greater ability to explore the differences between the immediate and delayed conditions as well as differences between the word and nonword conditions. Additionally, by collecting confidence ratings we can confirm whether or not d' is the appropriate measure of sensitivity to be using with this task as well as gain an understanding of the underlying signal and noise distributions assumed by signal detection theory for this task.

Method

Subjects

Twelve undergraduate students at the University of California San Diego participated for course credit. All were native speakers of English, had either normal or corrected to normal vision, and were naïve concerning the purpose of the experiment.

Apparatus

Experiment 2 was conducted with the same equipment/software and under the same viewing conditions as Experiment 1.

Materials.

One hundred and twenty of the experimental sentences from Experiment 1 were used along with 40 filler sentences. As with Experiment 1, an invisible boundary was placed between the last letter of the pre-boundary word (word *n-1*) and the subsequent inter-word space. Table 5 shows length, word frequency, and mean log bigram frequency (determined from the CELEX count using the N-Watch software, Davis 2005) of the target words and as well as the previews.

Insert Table 5 about here

In Experiment 2 there were six different preview conditions: (1) identical, (2) case change only, (3) word change, (4) word and case change, (5) nonword change, and (6) nonword and case change. The word previews were selected to have the same word shape as the targets. Additionally, an attempt was made to choose word previews with lexical and bigram frequencies that were close to the targets. The nonword previews were created by replacing the first letter of the word preview with a letter of a similar shape. Table 6 shows examples of the preview conditions.

As with Experiment 1, we presented every second word of all sentences in alternating case with word n always presented in alternating case (see Table 6). In Experiment 1 whether the first letter of the target was presented in upper or lower case was a within item variable intended to prevent subjects from developing a strategy. In Experiment 2 this was done as a between item variable; for half of the experimental sentences, the first letter of the target was in upper case and for the other half it was in lower case. However, as this was done only to prevent subjects from developing a specific change detection strategy that might deviate from normal reading we will not discuss it further.

Once readers crossed the boundary, the preview was replaced by the target word after the delay period of zero or 15 milliseconds. The identical trials (along with the filler trials) were used to estimate false alarm rates for the d` calculations as in Experiment 1. Across the experiment there were 100 trials in which there was a detectable change during reading and 100 trials in which there was no change (20 experimental item trials and 40 filler item trials). As with Experiment 1, the target word was evenly distributed over the first, second, and last third of the word positions within the sentences (excluding the first two word positions and the final word position).

Insert Table 6 here

Procedure

Subjects were presented with the experimental and filler sentences in random order. After every sentence presentation, subjects were prompted to respond, with a button press, as to how confident they were that a display change had occurred on a six-point scale (1 = very confident that there was no display change, 6 = very confident that there was a display change). As with Experiment 1, approximately 50% of the prompts were followed by a comprehension question that subjects answered by pressing the button corresponding to their answer choice. Subjects practiced responding to the prompt and the comprehension questions during 10 practice trials, 50% of which contained display changes. Outside of the practice trials, subjects did not receive feedback on their responses.

Results

Data exclusion criteria were the same as for Experiment 1. Fixations shorter than 80 ms, which were within 1 letter of a previous or subsequent fixation, were combined with that fixation, all other fixations less than 80 ms were eliminated (1.9% of all fixations). In total, 8.4% of the experimental trials were removed prior to analysis (0.2%

due to blinks or track loss; 5.3% due to early triggering of display changes; 2.9% due to late or failed display changes). Data loss affected all conditions similarly, $F(11,121) = 1.05$, $p > .4$). As with Experiment 1, the average accuracy to the comprehension questions was high (92%) and did not differ significantly by experimental condition, $F < 1$, again confirming that subjects were reading for understanding. As with Experiment 1 we will first present the analysis of d` data followed by fixation duration data. The means for the various dependent measures are given in table 7.

Analysis of d' Data

We calculated d` based on hits to 'change' trials and false alarms to 'identical' trials. However, given that we had confidence rating data, instead of simple binomial data, hits and false alarms were determined based on confidence ratings of 3 or greater in the respective conditions. As with Experiment 1, zero false alarm rates were corrected by adding half a false alarm to the subject mean, and hit rates of one were corrected by subtracting half a hit from the subject mean.

There were three main purposes of Experiment 2: (1) to confirm that, in the no delay condition, changing the letters between preview and target resulted in larger d` values than simply changing the case of these letters, (2) to explore possible differences between word and non-word letter replacement conditions, and (3) to examine the ROCs for the different preview and delay conditions. Therefore, to provide a more simplified and principled analysis, we will first focus on specific contrasts in which we compared d` in the case change condition with d` in each of the two different letter change conditions for each level of delay. We will then examine possible differences between the word and non-word conditions by comparing the conditions in which letters changed between the preview and the target with a 2 (delay: 0 ms vs. 25 ms) by 2 (type: word vs. nonword) by 2 (case: same vs. changed) ANOVA with subjects as a random factor. The ROCs for the letter change conditions averaged over the case change variable (which did not significantly impact d`) are presented in Figure 2.

Insert Table 7 about here

In the immediate condition (0 ms delay), the case change only condition resulted in significantly smaller d` values than either the word, $t(11) = 2.35$, $p < .05$, or non-word, $t(11) = 5.10$, $p < .001$, letter change conditions. However, in the 15 ms delay condition there was no significant difference in d` between the case change only and the word change, $t(11) = 1.85$, $p = .09$, or the non-word change, $t < 1$, conditions. This finding replicates the d` finding from Experiment 1 and further extends this finding to include changes where the letter replacement previews comprise a word.

In the ANOVA that included only those conditions in which letters were changed, there was a large effect of delaying the change as d` was more than twice as large if the change was delayed by 15 ms, $F(1,11) = 71.42$, $\eta_p^2 = .867$, $p < .001$. There was also an interesting interaction between delay and the type of preview, $F(1,11) = 10.07$, $\eta_p^2 = .478$, $p < .01$. Paired comparisons indicated that this effect was driven by a significantly larger d` for non-words in the immediate change condition, $t(11) = 3.38$, $p < .01$, with no significant difference between conditions when the change was delayed, $t(11) = 1.25$, $p = .235$. This interaction was somewhat unexpected as the non-word previews only differed from the word previews in their first letter. There were no other effects that approached significance ($F < 1.3$).

The ROCs paint an interesting picture. It is clear from simple visual inspection of Figure 2 that the delay caused quite a difference in these curves. This difference seems to be driven by an overall higher hit rate across all confidence levels in the delay condition. Additionally, these ROC curves are clearly curvilinear suggesting that a measure such as d` or d_a is appropriate for this task[6].

Insert Figure 2 here

Analysis of Eye Fixation Data

As with Experiment 1, we assessed potential parafoveal on foveal effects with an analysis of the duration of the fixation immediately prior to crossing the boundary. These data were submitted to an LMM that contained the manipulated elements of the preview, letter case, and letter identity, as fixed and subjects and items as random effects. Since there were three preview letter identity conditions, we used two orthogonal contrasts, the first (identity contrast) compared the identical with the average of the two dissimilar letter conditions and the second (lexicality contrast) compared the dissimilar non-word with the dissimilar word conditions. These fixation times were not significantly impacted by the manipulations, all ps > .12. Therefore, with dissimilar letter previews that were words or highly word-like, we find no evidence of parafoveal on foveal effects.

On average, the target word (n) was initially skipped on 7.9% of the trials. In a logistic LMM analysis with preview letter case and preview letter identity as fixed and subjects and items as random effects, there were no significant main effects or interactions on skipping rate (ps > .20). As with Experiment 1, we averaged over the delay condition in this analysis since the display change had not yet occurred at the time that saccades were being planned to skip the target.

In order to more closely examine the word and non-word letter replacement conditions that were the main impetus for Experiment 2, we submitted the gaze durations for the eight conditions in which letters changed to an LMM analysis, with delay, preview case, preview letter identity, and their interactions as fixed and subject and item as random effects. Since there were three preview letter identity conditions, we used two orthogonal contrasts, the first (identity contrast) compared the identical with the average of the two dissimilar letter conditions and the second (lexicality contrast) compared the dissimilar non-word with the dissimilar word conditions. As expected, gaze durations were significantly longer when the display change was delayed by 15 ms ($b = 29.2$, $SE = 3.66$, $p < .01$) and when the preview consisted of dissimilar letters (identity contrast: $b = 37.23$, $SE = 5.18$, $p < .01$). Additionally, there was a significant three-way interaction between delay, case change, and each of the preview letter contrasts (identity contrast: $b = -16.176$, $SE = 5.178$, $p < .01$; lexicality contrast: $b = -9.44$, $SE = 4.47$, $p = .03$).

We explored these interactions further by fitting separate models for the 0 ms and the 15 ms delay conditions. In both models, the identity contrast was highly significant, (0 ms: $b = 29.13$, $SE = 7.01$, $p < .01$; 15 ms: $b = 46.26$, $SE = 7.611$, $p < .01$) indicating a sizable preview benefit effect. However, in the 0 ms delay condition, neither the interaction between the identity contrast and case, nor the interaction between the lexicality contrast and case reached significance ($p = .098$ and $p = .235$, respectively). This can be compared with the 15 ms delay condition, where we found a significant interaction between the identity contrast and case change ($b = -44.585$, $SE = 7.6$, $p < .01$) and a marginally significant interaction for the lexicality contrast and case change ($b = -12.3$, $SE = 6.45$, $p = .065$). The first interaction indicates that, when the case changed between preview and target word, there was no further effect of the preview letters being identical to those of the target. However, when the letter case was the same for preview and target, changing the letter identities caused a large increase in fixation duration. The second (marginal) interaction indicates that, changing the case of the letters had a larger impact on fixation durations for word previews than for non-word previews.

In summary, while letter identity had an effect even in the no delay condition, letter case only affected fixation durations on the target in the 15 ms delay condition by modulating the effect of letter identity. These results therefore replicate the basic pattern of gaze duration results from Experiment 1.

Post Hoc Analyses

Similar to Experiment 1, we explored how aspects of the eye movement record may have influenced subjects' detecting display changes. We used the same five variables that we tested in the initial post-hoc analysis of Experiment 1: (1) the duration of the fixation immediately prior to crossing the boundary, (2) the proximity of this pre-change fixation to the boundary (in pixels), (3) the timing of the completion of the display change relative to the start of the post boundary change fixation, (4) the duration of the post boundary change fixation, and (5) the proximity of this post-change fixation to the boundary (in pixels). However, in Experiment 2 we have confidence rating data instead of binomial hit rate data. Therefore, our regression equation predicted confidence ratings from the independent factors mentioned above which appear in Table 8 using subjects and items as random effects. Again, we conducted these analyses using the LME4 package of the R statistical software (Bates & Maechler, 2010; R Development Core Team, 2010), and report coefficient, and standard error estimates as well as p-values (estimated from Markov chain Monte Carlo simulations). We first present these analyses for the no delay data. As with Experiment 1, the closer the subjects' fixation was to the invalid preview before triggering the display change, the higher their confidence rating was that a change had occurred, b = -.0106, SE = .0029, p = .0001. There was also a positive relationship between the duration of the post boundary change fixation and subjects confidence rating, b = .0023, SE = .0006, p = .0003. As with Experiment 1, none of the other three variables had a significant effect. Therefore, the current analysis using confidence ratings was in complete agreement with the analysis from Experiment 1 which used binomial 'hit' data.

Insert Table 8 here

Unlike Experiment 1, the delay condition sensitivity data were below ceiling. This along with the use of confidence ratings allowed us to more directly investigate how the timing of the display change influenced sensitivity. This timing variable did not influence confidence ratings in the no delay data. However, it should be noted that in the no delay conditions the change completed prior to the start of the fixation on over 83% of the trials. Additionally, the average timing for the trials in which the display change completed after the start of the fixation was only 3.8 ms so it is not surprising that in the immediate change condition this timing variable did not impact the confidence ratings. The initiation of the display changes in these conditions was always delayed by 15 milliseconds. However, as this delay was initiated upon crossing the boundary, there was quite a bit of variability in the timing of the completion of the change relative to the beginning of the post boundary fixation[7]. We therefore conducted the same LMM as with the immediate change (no delay) data with the delay data. There were again only two significant predictors of confidence ratings in this analysis. However, the proximity of the pre-change fixations to the invalid preview no longer influenced confidence ratings as it had in the immediate change conditions. First, as with the immediate change data, there was a positive relationship between the post-change fixation durations and confidence ratings, b = .0023, SE = .0005, p = .0001. Second, there was a strong influence of the timing of the change relative to the start of the post-change fixation, b = .0864, SE = .0137, p = .0001.

Insert Table 9 here

In order to further examine the influence of the timing of the display change relative to the start of the post boundary change fixation, we plotted (see Figure 3) the mean confidence rating over this timing variable in 9 ms time bins centered around -10, -5, 0, 5, 10, and 15 ms timings for both the display change and no display change trials. This plot indicates that the increase in sensitivity to display changes that occur with delayed timing of the change is most dramatic between the 5 and 10 ms delay bins. This suggests that as long as display changes are completed within the first 5 ms of the display change, sensitivity to the changes would be unlikely to create significant ‘artifacts’ in the data.

Insert Figure 3 here.

Discussion

Experiment 2 replicated many of the important novel effects from Experiment 1 while incorporating a number of crucial changes. First, in Experiment 2 confidence rating data were collected rather than simple binomial ‘hit’ data. This allowed a more detailed examination of readers’ sensitivity to detect these boundary changes. For instance, these confidence rating data were necessary to determine that the underlying signal detection model of this task was one in which the variability in the signal distribution was larger than the variability in the noise distribution.

Second, the delay was reduced from 25 to 15 ms in an attempt to reduce the hit rate in the delay condition from the ceiling values that were obtained in Experiment 1. Such ceiling performance may have hidden or obscured important effects in the delay condition of Experiment 1. As hoped, the hit rate in the delay condition of Experiment 2 was only 78% overall (compared to 91% in Experiment 1). Additionally, the results of the delay condition of Experiment 2 were quite consistent with those from Experiment 1 suggesting that the ceiling performance in the delay condition of the first experiment was not obscuring the results. This difference in hit rate also highlights the importance of experimental timing when conducting such display change experiments with reading. Typically, saccade durations during reading are in the range of 20-35 ms. However, the change is only initiated after the boundary is crossed, which can occur at any time during the saccade. The post hoc results from the delay conditions indicate that the longer this change is delayed relative to the start of the post-boundary change fixation (the end of the triggering saccade) the greater subjects' confidence will be that a change in fact occurred. While the post hoc results from the immediate condition did not find a significant impact of this display change timing on confidence ratings, in the immediate condition the change completed prior to the start of the fixation on over 83% of trials, Additionally, when the change completed after the start of the fixation, the average timing relative to the start of the post boundary change fixation was only 3.8 ms so it is not surprising that in the immediate condition this timing variable did not impact the confidence ratings.

This finding provides further support for the use of such boundary changes to study normal reading in situations where the timing of such changes is well controlled.

The third major difference between Experiment 1 and 2 was the use of both word and non-word letter change conditions. This was done for two main reasons. First, we wanted to rule out the possibility that the larger d` values for the letter change vs. case change previews in the no delay condition of Experiment 1 were due to these letter change previews being non-words. In Experiment 2 both the word and non-word letter change conditions resulted in larger d` values than the case change only condition when the change was not delayed. The second reason for using the word and non-word letter change previews in Experiment 2 was to investigate differences in detection rates and gaze durations between these conditions. Quite unexpectedly, we found a significant interaction in d`s between the lexical status of the preview (word vs. non-word) and the delay (0 ms vs. 15 ms) with larger d`s for non-words than for words in the 0 ms delay condition and similar d`s for non-words and words at a 15 ms delay. The increased display change sensitivity for non-word previews in the immediate change condition is consistent with the idea that detection of these changes is due in part to having noticed something ‘odd’ in the parafovea. However, these data could also be explained by assuming that word previews are represented in a more abstract form than non-word previews and then assuming that this abstract representational format impedes display change detection. In the gaze duration data, we also found an interesting marginal three-way interaction between delay, case change, and the lexicality contrast. This interaction was such that changing the case and letters of word previews reduced fixation times compared to only changing the letters but only in the delay condition. Gaze durations are known to be very sensitive to lexical variables (Rayner, 1998, 2009). Therefore, it may be that with word previews and delayed changes there was residual lexical activation of these previews that interferes with the lexical processing of the target. If this is the case, changing the case of all the letters may better mask the preview letters thereby disrupting this residual lexical activation. It is also possible that this effect is due in part to saccadic inhibition. Figure 4 presents the fixation time distribution plots (with 50 ms bins) for the word preview conditions (no case change vs. case change). As seen in this figure, the apparent saccadic inhibition dip is greater for the condition in which the case of the letters remains the same between the preview and target letters. This may be related to differences in the underlying fixation time distributions in the absence of sudden visual onsets and the saccadic inhibition that ensues. Comparing the no-delay distributions for these conditions, the peak of the no case change condition is shifted to the right compared to the case change condition. Therefore the effects we see in mean fixation duration for the delay conditions could be caused by differential influences of saccadic inhibition related to the underlying fixation time distributions (when no delay is added).

Insert Figure 4 here.

Whatever the underlying cause of this interaction in gaze duration, it represents a somewhat divergent finding from the results of the d` data. With the d` data, we find a significant effect of lexicality but only in the no delay condition. In the gaze duration data, we find a significant interaction between lexicalilty and case change but only in the delay condition. We also found a similar dissociation between the gaze duration and d` data in Experiment 1.  Here, gaze durations to the case change only condition were shorter than the letter change conditions in both the delay and no delay conditions, while the d` data indicated that the case change only condition yielded lower d`s but only in the immediate condition. Therefore, it appears that while fixation times on the post boundary target are influenced by detections of display changes (with longer times associated with greater detection likelihoods), they are also independently influenced by ongoing lexical processing.

General Discussion

Boundary change studies have been used and continue to be used extensively in eye tracking studies of reading to examine issues related to preview benefit. Usually the conclusions drawn from these studies rely on comparisons of invalid preview conditions, in which there is the potential for detecting a change, with a valid preview condition in which no detectable change occurs. However, often the crucial comparisons are between two or more invalid preview conditions. As these conclusions normally relate fixation duration differences to lexical processing, it is crucial that comparisons between invalid conditions not be contaminated with differences in sensitivity to detect such changes. The current studies confirm that sensitivity to detecting display changes can influence target fixation durations as suggested by O'Regan (1990, 1992), providing the first demonstration that differences in the properties of the previews can result in significant differences in detection rates and subsequently in fixation times. However, preview effects were still robust even in the absence of display change detection, indicating that while detection artifacts may increase the size of such preview effects they are not solely responsible for them.

Perhaps one of the most interesting finding from the current studies is the complex pattern in sensitivity to detecting display changes. This pattern indicates a clear interaction between the timing of the display change and the relationship between the abstract letter identities of the preview and the target. When changes were initiated immediately upon crossing the boundary abstract letter identities were all that mattered. However, in the delay conditions sensitivity was visually based with no influence of abstract letter identities.

Target gaze durations generally increased with increasing sensitivity to detect the display changes. However, target gaze durations were also consistently influenced by abstract letter identities in that accurate letter identity information in the preview always provided a gaze duration benefit over inaccurate letter identity information. These are important findings with regards to the use of display change methodology to study normal reading as they indicate that display change detection ‘artifacts’ can substantially increase target gaze durations and the pattern of this increase will be a delicate interplay between target and preview relationships (i.e. visual, lexical) and the timing of the display change. This is highlighted by the literature on saccadic inhibition, which indicates that sudden visual onsets can disrupt saccade plans approximately 100 ms post onset. The influence of this saccadic disruption on mean fixation time will depend on the underlying distribution of saccades at this time point. In the current study evidence for such saccadic inhibition was present for our delayed change conditions. The post-hoc analyses of Experiment 2 paint a cautionary tale as they indicate a linear increase in display change sensitivity with increasing display change timing relative to the start of the post boundary change fixation. Figure 3 indicates that this increase in sensitivity to changes becomes most dramatic between 6 and 14 ms after the start of the post boundary change fixation. However, even the sensitivity in the 1 to 9 ms interval was significantly greater than that of the -4 to 4 ms interval. Therefore, it is crucial that researchers using these display change techniques maintain a high degree of control over the timing of such changes. Alternatively, researchers could monitor change detections on a trial-by-trial basis for all or a subgroup of subjects to ensure that detection rates are comparable across the invalid preview conditions. However, this introduces a new task that is unrelated to reading to the experiment and may not be desirable for that reason.

The post hoc analyses of both experiments suggest that the proximity of the fixation to the preview prior to crossing the boundary significantly influences display change detection. As we alluded to at the outset, there is currently an ongoing debate as to whether lexical processing is distributed in parallel across multiple words or is relatively constrained to one word at a time. There have now been quite a few studies that have investigated this issue by manipulating the preview of words n+1 and n+2 during reading (Angele et al., 2008; Kliegl et al. , 2007; Rayner et al., 2007; Angele & Rayner, 2011; Glover et al., 2010; Risse & Kliegl, 2011). These studies have found conflicting results with some concluding that readers do not obtain significant preview benefit from word n+2 and others finding small but significant preview benefits from these words. One hypothesized reason for such differences is the length of word n+1, as evidence for preview benefit for word n+2 has come mainly from studies using short n+1 words. This would be consistent with the finding from the current studies showing that display change detection is more likely when the eyes are closer to the to-be-changed word.

In the experiments reported here, the subject's task was somewhat different from most boundary experiments. Specifically, on each trial readers had to first make a judgment as to whether or not they had noticed a display change (and in Experiment 2 they also had to provide a confidence rating). Then (on 50% of the trials) they had to answer a comprehension question regarding the sentence. Subjects at the beginning of the practice trials found this a bit difficult, but by the end of the practice trials they could do the task quite well. Indeed, the fact that the comprehension questions were answered correctly 93% and 92% of the time in Experiments 1 and 2 respectively suggests that readers were reading quite normally. However, it may be the case that they were reading more cautiously in both experiments than is typical. Specifically, the probability of skipping a word was lower in the present experiments than is typical. On the other hand, the average fixation times were not longer than normal. If readers were reading more cautiously in the present studies than in most prior studies they may have been more likely to notice display changes in the present study than is typical. But, many of them were largely insensitive to these changes and we therefore believe that the results of the present studies are rather diagnostic regarding display changes per se. Additionally, simply adopting a more careful reading strategy would not explain the complex pattern of results obtained in the current studies. Our results are quite consistent with prior research reported by Inhoff et al. (1998) and White et al. (2005) while still providing more precise information than these latter studies were able to do.

In summary, the results of the present study demonstrate that the use of signal detection analyses when combined with standard eye movement measures can provide valuable information regarding processing during reading. However, the confidence ratings of display change detection were actually more sensitive to the lexical status of previews than gaze durations were. Therefore, this new method of exploring display change detection with the use of confidence rating data may prove especially valuable in resolving controversies related to the distribution of attention and lexical processing during reading.

References

Angele, B., & Rayner, K. (2011). Parafoveal processing of word n+2 during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance*, in press.

Angele, B., Slattery, T., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n 1 and n 2 previews simultaneously. *Visual Cognition, 16*(6), 697-707.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press, Cambridge, UK.

Bates, D., Maechler, M., & Dai, B. (2009). Lme4: Linear mixed-effects models using S4 classes. *R Package Version 0.999375–32.*

Briihl, D., & Inhoff, A. W. (1995). Integrating information across fixations during reading: The use of orthographic bodies and of exterior letters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 55-67.

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*(5), 621-636.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777-813.

Glover, L., Vorstius, C., & Radach, R. (2011). Exploring the limits of distant parafoveal processing during reading: A new look at N+2 preview effects. *Journal of Eye Movement Research, in press.*

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n 2. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250-1255.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Lawrence Erlbaum Associates, Mahwah, New Jersey, USA.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899-917.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading. *Perception & Psychophysics, 25*(3), 221-224.

O'Regan, J. K. (1992). Solving the" real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*(3), 461-488.

O'Regan, J. K. (1990). Eye movements and reading. *Reviews of Oculomotor Research, 4,* 395-453.

R Development Core Team. (2010). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65-81.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618-660.

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

Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word n 2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology, 33*(1), 230-245.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology, 12*(2), 206-226.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*(1), 4-22.

Reilly, R. G., & Radach, R. (2003). Foundations of an interactive activation model of eye movement control in reading. *In: Hyönä, J., Radach, R. & Deubel, H. (Eds.). The Mind's Eye: Cognitive and Applied Aspects of Eye Movements. Elsevier Science, Oxford, UK.*

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*(1), 34-55.

Reingold, E. M. & Stampe, D. M. (1999). Saccadic inhibition in complex visual tasks. In W. Becker, H. Deubel, & T. Mergner (Eds.), Current oculomotor research: Physiological and psychological aspects. (pp. 249-255). Plenum: London.

Reingold, E. M. & Stampe, D. M. (2000). Saccadic inhibition and gaze contingent research paradigms. In A. Kennedy, R. Radach, D. Heller & J. Pynte (Eds), Reading as a perceptual process. (pp. 119-145). Elsevier: Amsterdam.

Reingold, E. M. & Stampe D. M. (2004). Saccadic inhibition in reading. Journal of Experimental Psychology: Human Perception and Performance, 30, 194-211.

Risse, S., & Kliegl, R. (2010). Adult age differences in the perceptual span during reading. *Psychology and Aging*, in press.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review, 12*(5), 891.

Acknowledgments

This research was supported by grant HD26765 from the National Institute of Child Health and Human Development. We thank Albrecht Inhoff and an anonymous reviewer for their helpful comments. Correspondence should be addressed to Timothy J. Slattery (tslattery@ucsd.edu).

Footnotes

1. However, a good practice is to have the software keep track of exactly when the display change occurred relative to eye position.
2. As an additional check, we conducted two ANOVAs that included the case of the first letter as an independent measure, one on d` values and one on gaze durations. There was a marginal effect of the case of the first letter on d` values ($p = 0.081$) with higher values when the first letter had been capitalized. However, there was no impact of this variable on gaze durations $Fs < 1$.
3. Saccades that trigger the display change early are sometimes referred to as hooks and are believed to be the result of saccadic overshoots and corrections.
4. It is standard in psycholinguistic studies to do two separate ANOVAs, one with subjects as a random effects factor and the other with items as a random effects factor. There are two reasons why we are not doing this with the current d` data. First, d` is a measure that is influenced by bias inherent to a given subject and most studies using this measure do not report items analyses for this reason. The second reason is that in the current studies we use the filler sentences to help better estimate subject false alarm rates. This cannot be done for the items as it would be similar to estimating a given subjects false alarm rate from the false alarm rate other subjects.
5. We also conducted the fixation time analyses using standard F1 and F2 ANOVAs and follow up t-tests where appropriate. These analyses were highly consistent

with the LMMs that we report in the main text and yielded the same pattern of statistical findings.

6. When plotting the data in zROC space it is clear that the variances of the two distributions (signal and noise) are not equal. This suggests that d_a is a more appropriate measure of sensitivity here. For this reason we calculated d_a for the current study and the results did not differ from those of the d` analyses.
7. This variability is do both to variability in the CRT monitor’s raster sweep, and variability due to the remaining duration of the saccade once the boundary is crossed.

Table 1. Target word characteristics of Experiment 1 stimuli.

	N-1	N	N+1
Word frequency	598.21 (4938.28)	169.68 (158.17)	666.77 (2979.85)
Word length	6.96 (2.03)	6.23 (1.44)	6.76 (2.01)

Note: frequencies are occurrences per million words; word length is in number of characters. The length of the target word varied between 4 and 10 letters.

Table 2. Example items Experiment 1.

Letter preview	*Case preview*	*Example sentence*
		`BoYs' voices WiLl noticeably ChAnGe during PuBeRtY.`
identical	identical	`BoYs' voices WiLl noticeably ChAnGe during PuBeRtY.`
identical	dissimilar	`BoYs' voices WiLl noticeably cHaNgE during PuBeRtY.`
dissimilar	identical	`BoYs' voices WiLl noticeably RbEcPa during PuBeRtY.`
dissimilar	dissimilar	`BoYs' voices WiLl noticeably rBeCpA during PuBeRtY`

Table 3. Experiment 1 means

Change Type	Delay	None	Case	Letter	Both
d`	0 ms	NA	0.77 (.09)	1.40 (.19)	1.55 (.21)
	25 ms	NA	3.78 (.12)	3.84 (.13)	3.83 (.13)
Skipping Rate	NA	4.7 (1.5)	3.0 (1.0)	2.1 (0.7)	1.8 (0.6)
First Fixation	0 ms	249 (9.9)	257 (12.9)	295 (15.3)	296 (12.7)
	25 ms	242 (10.2)	342 (16.9)	364 (15.5)	372 (14.5)
Gaze Duration	0 ms	289 (12.2)	303 (15.2)	348 (14.4)	360 (12.6)
	25 ms	292 (11.0)	403 (16.0)	449 (19.4)	445 (13.5)

Note: Skipping rate is the percentage of trials in which the target word was skipped during first pass reading. First fixation and gaze duration are given in milliseconds. Standard errors are reported in parenthesis below each cell mean.

Table 4. Experiment 1 Post-hoc variable means for the zero ms delay letter change conditions.

Change Type	Letter	Letter and Case
Pre-change Fix Distance	48.39 (1.77)	49.12 (1.60)
Pre-change Fix Duration	232.99 (5.19)	226.19 (4.89)
Timing of Change	-3.48 (.39)	-3.56 (.42)
Post-change Fix Distance	-36.81 (1.03)	-35.17 (1.00)
Post-change Fix Duration	300.29 (6.51)	298.89 (6.97)

Note: Distances given in pixels (1 character = 11 pixels), times given in milliseconds, standard errors are in parentheses.

Table 5. Word characteristics of Experiment 2 target stimuli.

	Target	Unrelated preview	Nonword preview
Word frequency	175.09 (168.61)	49.89 (154.17)	NA
Word length	5.95 (1.32)	5.95 (1.32)	5.95 (1.32)
Mean Log Bigram frequency	2.91 (0.34)	2.82 (0.32)	2.74 (0.33)

Table 6. Example item Experiment 2.

Letter preview	*Case preview*	*Example sentence*
		BoYs' voices WiLl noticeably ChAnGe during PuBeRtY.
identical	identical	BoYs' voices WiLl noticeably ChAnGe during PuBeRtY.
identical	dissimilar	BoYs' voices WiLl noticeably cHaNgE during PuBeRtY.
word	identical	BoYs' voices WiLl noticeably AlWaYs during PuBeRtY.
word	dissimilar	BoYs' voices WiLl noticeably aLwAyS during PuBeRtY.
nonword	identical	BoYs' voices WiLl noticeably ElWaYs during PuBeRtY.
nonword	dissimilar	BoYs' voices WiLl noticeably eLwAyS during PuBeRtY.

Table 7. Experiment 2 means

Change Type	Delay	None	Case	Word	Word + Case	NW	NW + Case
d`	0 ms	NA	0.39 (.12)	0.77 (.23)	0.76 (.24)	0.96 (.22)	1.14 (.21)
	15 ms	NA	2.23 (.25)	2.53 (.24)	2.41 (.30)	2.28 (.23)	2.37 (.23)
Skipping Rate	NA	8.4 (3.0)	7.6 (3.1)	8.8 (3.1)	5.9 (3.1)	7.9 (2.6)	8.9 (3.4)
First Fixation	0 ms	255 (15.7)	260 (14.2)	294 (22.3)	278 (11.3)	295 (19.6)	281 (20.5)
	15 ms	259 (11.9)	319 (21.7)	341 (19.1)	310 (21.7)	336 (23.2)	346 (25.6)
Gaze Duration	0 ms	283 (17.1)	291 (13.0)	337 (25.1)	325 (18.3)	348 (19.0)	315 (22.1)
	15 ms	284 (14.3)	372 (18.0)	423 (26.4)	366 (26.4)	411 (27.9)	389 (23.5)

Note: Skipping rate is the percentage of trials in which the target word was skipped during first pass reading. First fixation and gaze duration are given in milliseconds. Standard errors are reported in parenthesis below each cell mean.

Table 8. Experiment 2: Post-hoc variable means (0 ms delay condition)

Preview Type	Identical	Identical	Unrelated	Unrelated	Nonword	Nonword
Case	Changed	Unchanged	Changed	Unchanged	Changed	Unchanged
Pre-change Fix Distance	45.60 (3.03)	52.70 (3.62)	49.34 (3.03)	50.96 (3.37)	49.34 (3.10)	50.72 (2.77)
Pre-change Fix Duration	240.37 (23.80)	234.81 (8.85)	240.24 (10.44)	233.13 (8.04)	227.61 (8.03)	243.15 (11.43)
Timing of Change	-5.81 (0.67)	-6.28 (0.64)	-5.93 (0.74)	-5.91 (0.65)	-4.98 (0.66)	-5.29 (0.64)
Post-change Fix distance	-44.57 (2.07)	-45.04 (1.91)	-46.84 (2.88)	-43.15 (2.00)	-43.17 (2.08)	-42.04 (2.08)
Post-change Fix duration	258.46 (10.65)	255.22 (9.40)	273.44 (9.65)	291.96 (11.62)	271.11 (12.01)	295.24 (10.92)

Note: Distances given in pixels (1 character = 11 pixels), times given in milliseconds, standard errors are in parentheses.

Table 9. Experiment 2: Post-hoc variable means (15 ms delay condition)

Preview Type	Identical	Identical	Unrelated	Unrelated	Nonword	Nonword
Case	Changed	Unchanged	Changed	Unchanged	Changed	Unchanged
Pre-change Fix Distance	50.37 (3.14)	43.86 (2.52)	52.48 (3.58)	44.03 (2.85)	45.44 (3.02)	45.31 (2.55)
Pre-change Fix Duration	233.75 (10.17)	242.45 (8.99)	215.97 (7.66)	229.15 (8.56)	237.06 (10.41)	263.81 (12.75)
Timing of Change	18.51 (0.60)	19.16 (0.55)	18.07 (0.71)	19.49 (0.57)	18.94 (0.58)	19.43 (0.56)
Post-change Fix distance	-43.63 (2.16)	-47.03 (2.15)	-42.98 (2.13)	-42.97 (1.81)	-44.45 (1.76)	-40.94 (1.91)
Post-change Fix duration	313.01 (13.73)	251.74 (8.45)	316.94 (13.51)	339.17 (12.94)	336.33 (12.23)	341.13 (14.33)

Note: Distances given in pixels (1 character = 11 pixels), times given in milliseconds, standard errors are in parentheses.

Figure captions

1. Fixation time distribution plots with 50 ms bins for Experiment 1.
2. ROC curves for Experiment 2.
3. Mean confidence rating over display change timing bins.
4. Fixation time distribution plots with 50 ms bins for Experiment 2.

Figure 1.

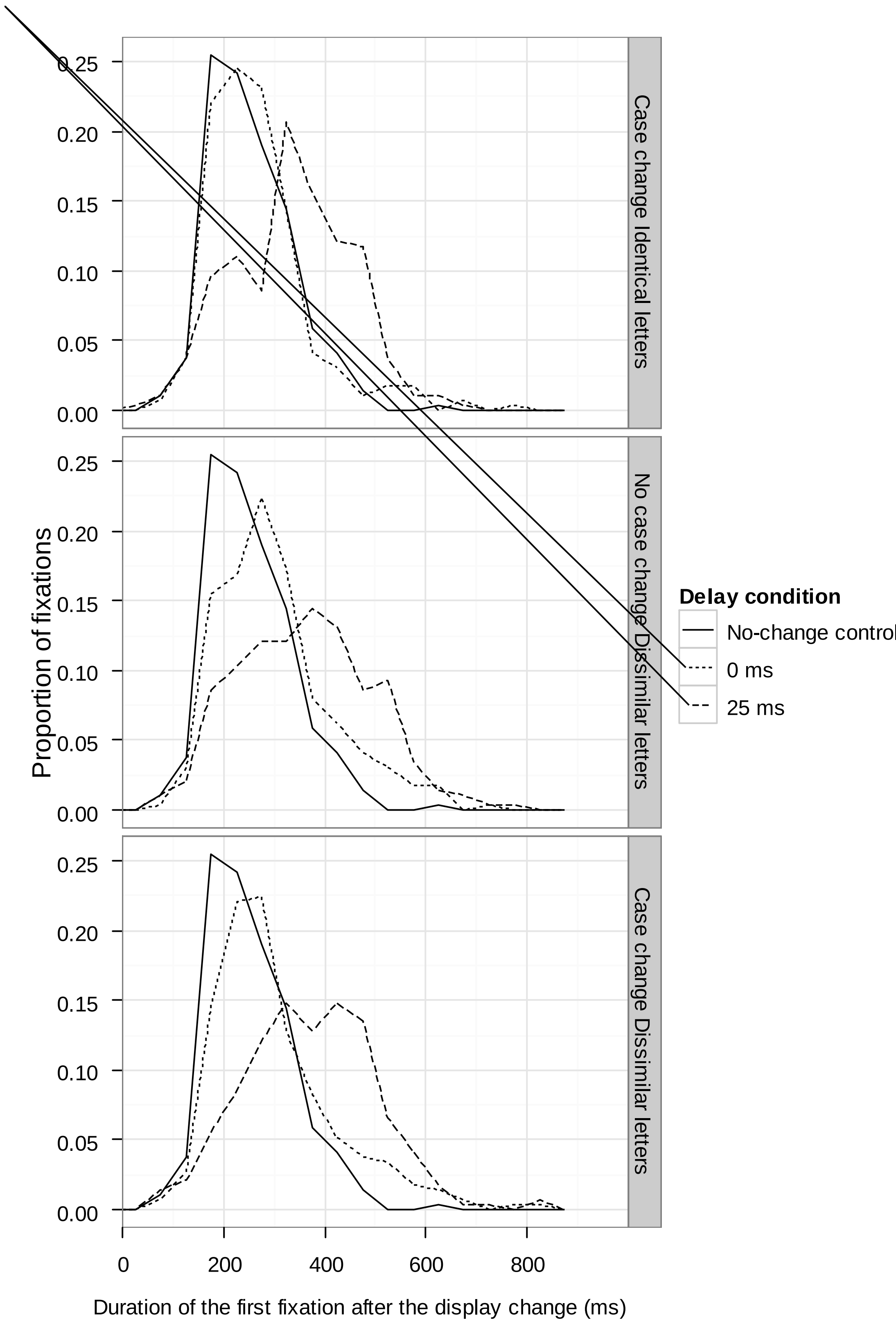

Proportion of fixations
0.25
0.20
0.15
0.10
0.05
0.00
Case change Identical letters
No case change Dissimilar letters
Case change Dissimilar letters
Delay condition
No-change control
0 ms
25 ms
0
200
400
600
800
Duration of the first fixation after the display change (ms)

Figure 2.

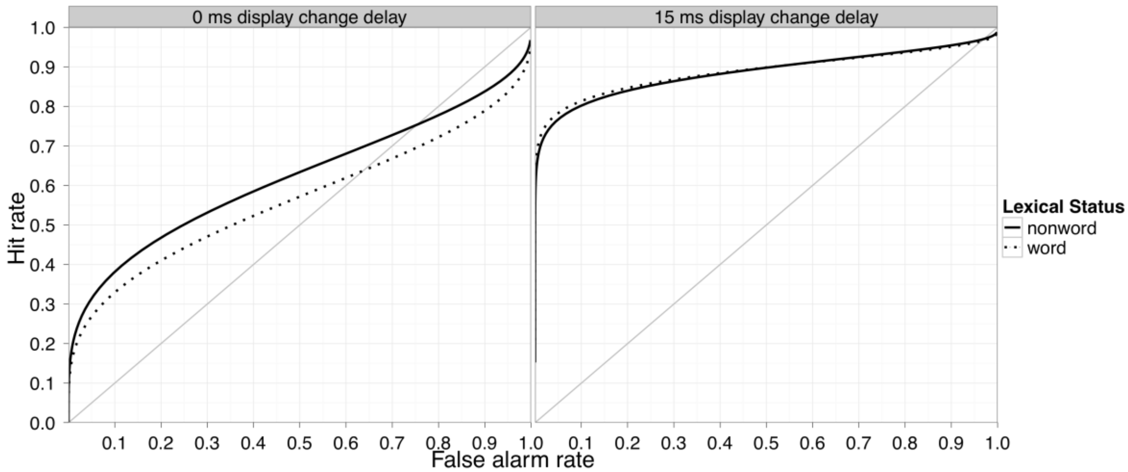

0 ms display change delay
15 ms display change delay
Hit rate
False alarm rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Lexical Status
nonword
word

Figure 3.

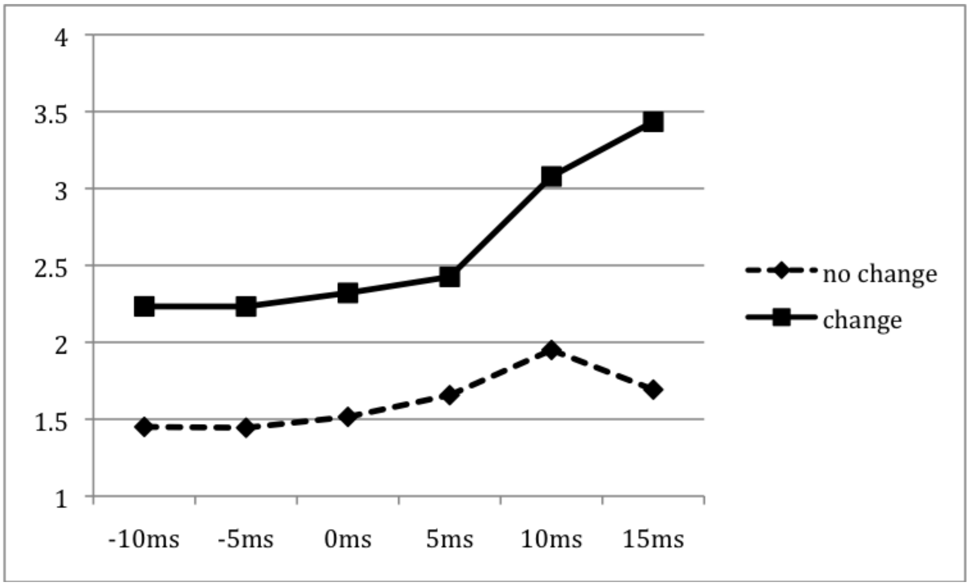
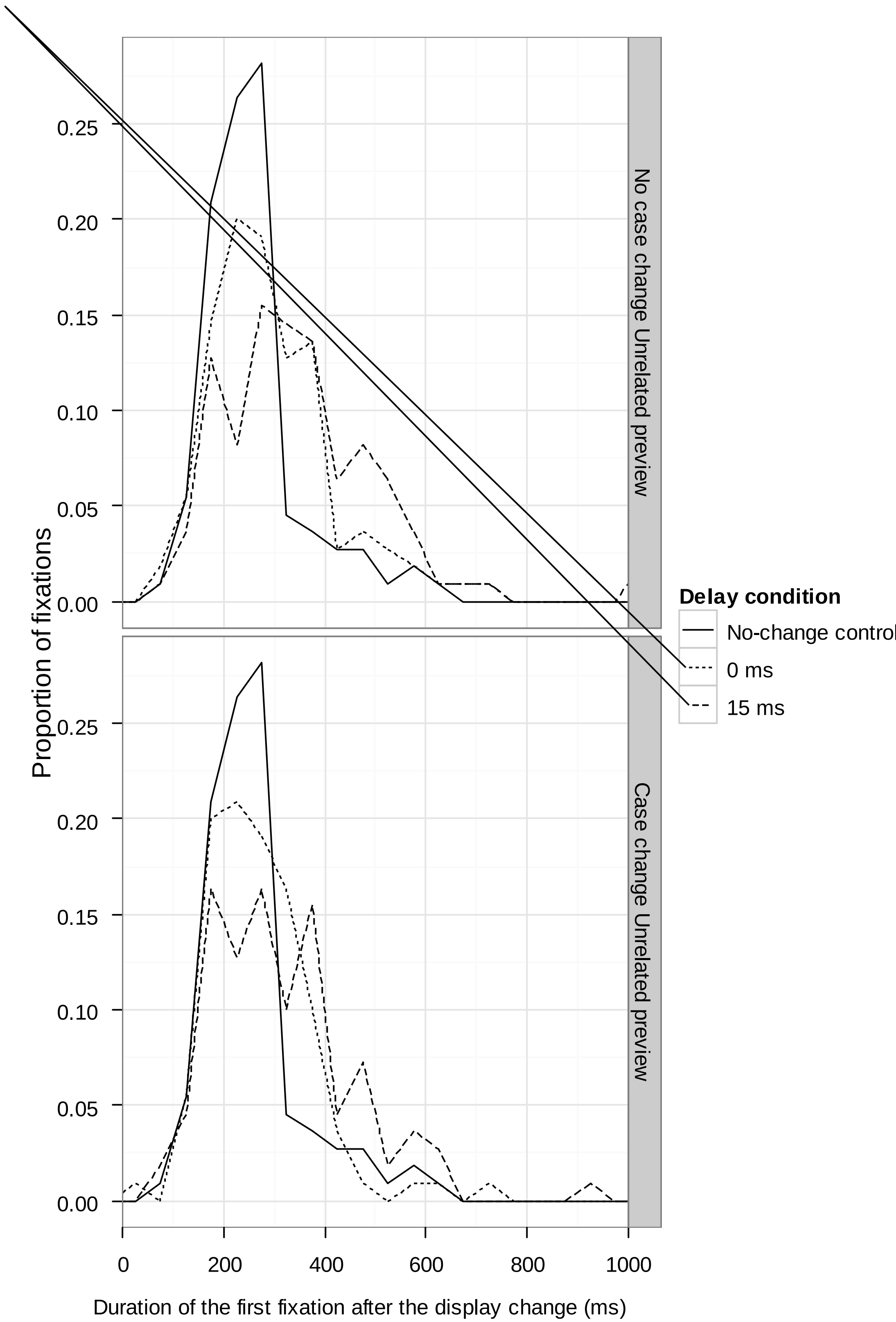


Figure 4.



0.25
0.20
0.15
0.10
0.05
0.00
No case change Unrelated preview
Case change Unrelated preview
Proportion of fixations
Delay condition
No-change control
0 ms
15 ms
0
200
400
600
800
1000
Duration of the first fixation after the display change (ms)