Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006)

Keith Rayner and Alexander Pollatsek
University of Massachusetts, Amherst

Denis Drieghe
Ghent University

Timothy J. Slattery
University of Massachusetts, Amherst
and

Erik D. Reichle
University of Pittsburgh

Correspondence to:

Keith Rayner
Department of Psychology
University of Massachusetts
Amherst, MA 01003
413-545-2175 (phone); 413-545-0996 (FAX)
rayner@psych.umass.edu

Running Head: Comment on Kliegl et al.

Abstract

Kliegl, Nuthmann, and Engbert (2006) reported an impressive set of data analyses dealing with the influence of the prior, present, and next word on the duration of the current eye fixation during reading. They argued that outcomes of their regression analyses indicate that lexical processing is distributed across a number of words during reading. In this comment, we question their conclusions and address four different issues: (1) whether there is evidence for distributed lexical processing, (2) whether so-called parafoveal-on-foveal effects are widespread, (3) the role of correlational analyses in reading research, and (4) problems in their analyses with only using cases where words are fixated exactly once.

Reading is a very complex task and to understand the mental processes that occur when we read would be a remarkable achievement (Huey, 1908). Recently, considerable attention has been devoted to the development of models of eye movement control in reading (see Reichle, Rayner, & Pollatsek, 2003). Admittedly, these models do not come close to a full understanding of reading, but they clearly represent a step in that direction. Furthermore, the models do a good job of predicting eye movement behavior during reading. Kliegl, Nuthmann, and Engbert (2006) reported an impressive set of data analyses dealing with the influence of the prior, present, and next word on the duration of the current fixation. They argued, on the basis of estimates from regression analyses (see also Kennedy & Pynte, 2005; Pynte & Kennedy, 2006), that lexical processing is distributed across a number of words falling within the perceptual span region. Their interpretation of their data is important because it is used to argue that parallel models of lexical processing, like their SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005), can better explain the data than a model, like the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006; Rayner, Ashby, Pollatsek, & Reichle, 2004; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006), in which lexical processing is serial (there are parallel components of E-Z Reader dealing with early visual processing and saccadic programming).

While we appreciate the impressive amount of data that Kliegl et al. presented and acknowledge that there is a lot of valuable information inherent in their analyses, we also suspect that some of their conclusions are not fully warranted. In this article, we focus on four issues: (1) distributed lexical processing; (2) parafoveal-on-foveal effects; (3) the use of correlational analyses; and (4) the selection of single fixation cases. We will first briefly discuss current models of eye movement control in reading to clarify the distinction between serial-attention-shift models and parallel-distributed-lexical-processing models. Since E-Z Reader and SWIFT represent prototypes of each type of model, our discussion will be limited to these two models.

Models of eye movement control in reading.

Studies using a gaze-contingent moving window paradigm (McConkie & Rayner, 1975, 1976; Rayner, 1986; Rayner & Bertera, 1979; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981; Rayner, Well, & Pollatsek, 1980; Rayner, Well, Pollatsek, & Bertera, 1982) demonstrated that readers extract word and letter information from a limited region of text – ranging from 3-4 letter spaces to the left of fixation to about 14-15 letter spaces to the right of fixation. That is, these studies showed that reading rates are not normal until the window that was presented to the reader was this large. Considerable work since then with readers of alphabetic writing systems (see Rayner, 1998) has confirmed these results and further demonstrated that although the region that readers extract text from is limited, it is greater than the word being fixated. However, the moving window paradigm indicates the *maximum* region from which readers extract information (Well, 1983) and it does not logically follow that readers extract information from this region on all fixations. Indeed, the finding that reading can proceed at close to normal speed (a 10% reduction) if the window contains the fixated word and the one to the right (Rayner et al., 1982) suggests that the information from these two words is largely what is driving reading.

As noted above, a number of computational models of reading, which differ on many dimensions (see Reichle et al., 2003), have recently appeared[1]. The dimension most relevant to the claims of Kliegl et al. is how words are processed during reading. E-Z Reader posits that words are lexically processed serially (in the forward direction of the text). However, this process occurs rapidly enough that the word to the right of fixation is at least partially processed on most fixations. In addition, in E-Z Reader, the word two to the right of fixation (word n+2) is processed on those fixations in which processing of both the fixated word (word n) and the one to the right (word n+1) of it is very rapid (e.g., when the word to the right of fixation is a function word). Furthermore, the word to the left of the fixated word (word n-1) would be occasionally processed on those occasions when an eye movement was intended for that word, but it overshot that word. (In E-Z Reader, covert attention moves across the text serially and is not directly linked to the triggering of eye movements.) In contrast, SWIFT, which is the framework for the Kliegl et al. article, assumes that, on all fixations, a number of words are processed in parallel.

The second important dimension on which the models differ is the event that controls the forward progress of the eyes. In E-Z Reader, a stage of lexical access (L_1) of the attended word is the trigger for an eye movement to word n+1. Furthermore, skipping is predicted when L_1 processing of word n+1 is rapid enough to initiate a saccadic program to move the eyes to word n+2, which then cancels the saccadic program that would otherwise cause the eyes to move to word n+1. In contrast, in SWIFT, the primary signal to move the eyes is an automatic timer that causes the eyes to move forward at random intervals. Although this timer can be inhibited if the reader experiences difficulty with lexical processing, this inhibition is delayed, so that whatever difficulty is associated with the processing of word n will usually manifest itself as increased fixations on word n+1. Moreover, this inhibition is modulated in a complex way by the progress in lexical processing from the words that are processed in parallel.

It is worth underlining here that the models do not differ all that much in their claims about what regions of text can be processed on a fixation (they are both consistent with the data from the moving window experiments). Furthermore, neither model seriously handles higher-order failures of comprehension or syntactic garden path effects (e.g., Frazier & Rayner, 1982). Our subsequent discussion assumes that we are considering reading situations in which such failures do not exist and that higher order processing lags behind word identification and is virtually invisible in the eye movement record. Where the models differ importantly is with respect to (a) the timing of the processing within a fixation and (b) whether lexical processing directly controls the eyes during reading. In particular, in E-Z Reader, lexical processing of the fixated word is the primary determinant of when the eyes move on in the text, whereas in SWIFT, processing from a larger region of text is controlling where and when the reader moves on virtually all fixations. The data presented in Kliegl et al. are an attempt to document the latter position. However, we feel that this documentation is problematic for two reasons. First, we think several of the claims made from their correlational analyses are contradicted by results from careful experimental studies and, given the likelihood that correlational analyses have serious confounds that make their interpretation problematic, that the results from controlled experiments should receive greater weight. Second, some results reported by Kliegl et al. are claimed to be novel and contrary to a serial-processing model such as E-Z Reader. However, we are not convinced of this and we believe that some of these results are in fact predicted by the model.

Distributed Lexical Processing?

Let's consider the evidence reported by Kliegl et al. (2006) for the distributed nature of lexical processing in eye fixations in reading. There are two findings that they reported which, if true, seem to be the strongest arguments against a serial model of word processing, such as E-Z Reader. The first is that lexical processing of the previous word (word n-1) is reflected almost as strongly as that of the present (i.e., fixated) word (word n) in the current fixation duration. The first point that we wish to make is that it is not clear that this is true even in Kliegl et al.'s data since it is not true for gaze durations (the sum of fixations on a word before moving to another word). That is, for single fixation durations, they report a regression coefficient of word frequency of -4.6 for the effect of word n and -5.1 for word n-1, but for gaze durations, they report a regression coefficient of -33 for the effect of word n and -1.6 for the effect of word n-1 (see their Table 4). The latter effect, in particular, is hardly in line with the distributed nature of lexical processing. That is, virtually all of the effect on gaze duration is accounted for by the frequency of word n. The second point is that the typical finding in the literature is that most of the effect of the frequency of a word is immediate. For example, we examined data from six experiments in which the frequency of a single target word was manipulated (with word n-1 and word n+1 held constant); in these studies, the effect of the frequency manipulation on spillover time (the duration of the fixation on the word which follows the manipulated frequency word) was about 40% of the effect on the current word.

The existence of spillover effects is not controversial and E-Z Reader easily accounts for them. The model posits that there is a time lag between when the eye movement to word n+1 is programmed and when attention shifts to word n+1, and this lag is a function of the processing difficulty of word n. This means that there is less time for processing of word n+1 in the parafovea when word n is more difficult (e.g., lower in frequency) and thus fixation time on word n+1 will be slowed down the more difficult word n is to process. However, for any reasonable assignment of word encoding times, the model predicts that size of spillover effects produced by varying the frequency of a target word are quite a bit smaller than the effects of this frequency manipulation on the gaze duration on the target word. A model such as SWIFT may be able to predict this pattern of results; however, it doesn't fall out as a natural prediction.

We should also point out that there are many phenomena from the eye movement literature that indicate strongly that deeper levels of lexical processing are triggering the decision to move to the next word. For example, Rayner and Duffy (1986) and Duffy, Morris, and Rayner (1988) demonstrated significant lexical ambiguity effects on the gaze duration on the current word (see Rayner, Cook, Juhasz, & Frazier, 2006; Sereno, O'Donnell, & Rayner, 2006 for recent evidence). For example, they found that, in neutral contexts, lexically ambiguous words with two meanings that are about equal in frequency are fixated longer than control unambiguous words. Moreover, these experiments indicated that the prior sentence context modulated these effects, implying that fairly deep semantic processing of a word is occurring before deciding to move off of it onto the next word. Similarly, Rayner, Warren, Juhasz, and Liversedge (2004) found that a prior context that makes a target word semantically anomalous causes gaze durations on the word to be longer than when the word fits in with the prior context but is not predictable (or highly plausible). This also indicates that, instead the identification of the meaning of a word continuing for several fixations after the eyes have moved off of the word, on a significant fraction of the trials that whatever processing is necessary to both identify the word and to determine that its meaning is not congruent with its prior sentence context is done more-or-less immediately. To be sure, there are spillover effects, but they could be due either to higher-order text integration processes or to slower detection of the anomaly in some cases.

Parafoveal-on-foveal effects?

A second main conclusion that comes from Kliegl et al.'s analyses is that parafoveal-on-foveal effects (i.e., that the characteristics of word n+1 significantly influence the time spent on word n) are pervasive. We are not claiming that such effects don't exist. Rayner (1975) found some time ago that when there were orthographically illegal letter strings at the beginning of word n+1, fixation time on word n increased. What is at issue is (a) how widespread these effects are, (b) what aspects of word n+1 produce them, and (c) whether the existence of such effects is a serious concern for a serial-processing model such as E-Z Reader.

The point we wish to make is that such effects are usually limited to when the fixation is on the last few letters of the word preceding a target word. As Kliegl et al. correctly point out, the larger effect of word n+1 on word n from these fixation locations is consistent with the hypothesis that, the closer one's fixation is to word n+1, the more processing of word n+1 will be completed because of the better visual acuity afforded by the proximal viewing location. However, this pattern of results is equally consistent with the hypothesis that, when a fixation is near the end of word n, there is a greater chance that word n+1 was the word that was intended to be fixated. More precisely, this alternate hypothesis is that there can be a discrepancy between the word that is attended to *even at the beginning of a fixation* and the word that is recorded as the fixated word. Such discrepancies can occur for two reasons: (a) inaccuracy in the eye tracker, and (b) inaccuracy in the eye-movement system such that the word that is both the focus of attention and the target of the saccade is either undershot or overshot (in this case, we are primarily concerned with undershoots). If the eye-tracking record indicates that the reader has fixated on word n while attending to word n+1 due to either of these errors, then it is not surprising that characteristics of word n+1 can affect fixation times on word n (Rayner et al., 2004).

The question that this counter-hypothesis raises is, of course, whether or not either of these situations is likely to occur and, if so, whether they occur frequently enough to produce parafoveal-on-foveal effects. Kliegl et al. attempted to rule out the first hypothesis by using only fixations where their eye-tracking device agreed that both eyes were on the same word. We agree that this procedure probably eliminates the worst machine scoring errors, but perhaps not all of them. The question of whether there is inaccuracy in the human spatial targeting system, however, is quite uncontroversial. In fact, Nuthmann, Engbert, and Kliegl (2005) nicely showed that mistargeting is the likely explanation of the paradoxical inverted optimal viewing position (IOVP) effect relating fixation time (single fixations and initial fixations) on a word and the initial landing position on the word (Vitu, McConkie, Kerr, & O'Regan, 2001). We say paradoxical because one would expect that single fixations should be shorter when the eyes land on the center of a word because it has been shown to be the optimal place for processing a word of normal length, but the data are that fixations near the beginnings and ends of words are shorter. Nuthmann et al. presented an elegant analysis indicating that this inverted-U pattern is likely due to the fact that an appreciable percent of the fixations that land near the beginnings and ends of words are due to saccadic errors in which word n–1 or word n+1 (respectively) were the intended targets. They then argue plausibly that these mistargeted fixations are likely to be followed by quick corrective saccades to get the reader nearer to a better place in the attended word. Their analysis indicates that such targeting errors are not at all rare: more than 10% of the saccades miss their intended targets. Such analyses would appear to make it difficult to rule out the null hypothesis that mislocalized fixations account for all (or most) parafoveal-on-foveal effects.

There is one aspect of the data of Kliegl et al., however, that appears to be inconsistent with much of the data from more controlled experiments. That is, Kliegl et al. report that lexical aspects of word n+1 affect processing time on word n, whereas most of the literature indicates that it is mainly fairly low-level aspects of word n+1, such as impossible initial letter sequences (e.g., *qw*), that produce such an effect[2]. One possible resolution of this discrepancy is that, because of the greater power of the Kliegl et al. experiments due to their large corpus of data, the small effects that they reported (about 5 ms) are significant, but are likely to be dismissed in most other experiments. Another possibility is that there is some artifact in the current correlational methods that is causing the difference. For instance, it is known that low frequency words are more likely to contain irregular letter sequences, and as mentioned above, previous research has indicated that such irregular sequences in the parafovea can influence fixation times on the currently fixated word. However, we would not classify this type of processing difficulty as lexical in nature but rather prelexical. Therefore one plausible artifact could be that, consistent with previous research, irregular letter sequences are driving this effect but that their correlation with word frequency is making it appear as though this variable is responsible.

We will return to this latter issue below. First, however, there is a phenomenon that has been observed repeatedly that we think is quite problematic for the Kliegl et al. view of what causes parafoveal-on-foveal effects. It comes from boundary experiments (Rayner, 1975), in which there is a single display change during the reading of a sentence. This is illustrated in Figure 1. In such experiments, readers who are fixating on word n often see stimuli in the position of word n+1 that have the same number of letters as the target word, but are not real words and do not share letters with the word that will appear when it is fixated. However, the fixation time on word n is usually virtually the same (unless the eyes are very close to the target word prior to the display change, as in Rayner, 1975) when the preview is "garbage" as when the preview is the word that belongs in the sentence. Fixation times on word n+1, however, are strongly affected by the nature of the preview information, indicating that the information was processed but had virtually no effect on fixations on word n. This is consistent with the E-Z Reader model, where processing of word n+1 is assumed to be underway before it is fixated on all fixations, but it is the processing of word n that is determining the fixation time on word n. Kliegl et al. might dismiss the boundary technique as unnatural reading, but it seems quite inconsistent with claims about parafoveal-on-foveal processing being pervasive, as the difference between having a zero frequency (and often unpronounceable) stimulus and a normal word in the parafovea is having little or no effect on the fixation time on the prior word (see Rayner & Juhasz, 2004; Rayner, White, Miller, Kambe, & Liversedge, 2002 for further discussion).

Insert Figure 1 about here

We realize that our argument above can be construed as being inconsistent in that we originally argued that parafoveal-on-foveal effects are non-controversial, but later argued that they are generally not observed in controlled studies. In the latter case, we are arguing that they either weren't observed or that hints of an effect were observed but weren't close to significant. Given that Kliegl et al. had a very large set of data, they found a small parafoveal-on-foveal effect to be significant. However, that doesn't make it theoretically interesting, as E-Z Reader can account for such a small effect as being due to mistargeting of saccades.

Before leaving this section, we note two sets of recent results which we think are highly problematic for parallel-lexical-processing models and parafoveal-on-foveal effects. First, McDonald (2005) noted that if processing is distributed over a number of words, then there should be cumulative preview benefit for words that fall within the reader's perceptual span on consecutive fixations. That is, given that the perceptual span extends about 15 letter spaces to the right of fixation, there will be some fixations on a given target word in which there were two prior fixations within the perceptual span region and there should be more preview benefit for such words than for words that had not fallen within the span on the prior fixation. What he found was that there was no evidence for cumulative preview benefit. Second, Rayner, Juhasz, and Brown (2006) used the boundary paradigm to examine if preview benefit is obtained from word n+2. If lexical processing is distributed over a number of words within the perceptual span, then readers should obtain preview benefit from both word n+1 and n+2. However, they found that while there was clear preview benefit for word n+1, there was no preview benefit for word n+2 unless word n+1 was skipped. Consistent with virtually every boundary type of experiment that has been reported, there was no evidence that the lexical properties of word n+1 influenced the fixation time on word n.

The use of correlational analyses

One important methodological issue runs through the current controversy: Whether to put more trust in (a) conclusions from experiments in which aspects of the text or the presentation are carefully controlled and one key variable is manipulated (such as the frequency of a particular word), or (b) conclusions that come from studies in which text is read and inferences are drawn from correlational analyses. We fully realize that no experiment is ever perfect, and that there can be uncontrolled variables in any experiment. Furthermore, we have utilized regression analyses in our work (Niswander-Klement & Pollatsek, 2006; Rayner & Well, 1996), though we typically follow up results obtained via regression analysis with experimental manipulations. For example, Juhasz and Rayner (2003) demonstrated that effects due to age-of-acquisition were apparent via regression analysis techniques and Juhasz and Rayner (2006) followed up on this finding in a controlled experiment where we demonstrated that age-of-acquisition has an effect on fixation time on the target word over and beyond word frequency. Nevertheless, we believe that controlled experiments are the better source of data because there are always serious problems interpreting correlational analyses (e.g., the relevant variables are often quite confounded and may thus be virtually impossible to unconfound using such techniques), and because controlled experiments allow for much stronger inferences about causality. Kliegl et al. seem to imply that the correlational approach is to be preferred because it is "natural reading"[3]. However, in their study participants read individual sentences and responded to comprehension questions after reading some sentences; this is the exact procedure employed in many of the experimental studies that they are claiming to be less natural. Thus, we think that there is every reason to believe that the experimental studies we cited (that did not use display changes) are equally natural examples of reading, and that there is no clear advantage for using correlational techniques. Moreover, while the studies in which there are display changes during saccades are possibly slightly less natural, they are only marginally less so. In most display-change experiments, readers are not aware of the display change (as it occurs during a saccade when vision is suppressed), so again we are left wondering how this situation is any different from the situation that Kliegl et al. use to analyze data and call natural reading.

What are the problems with the correlational techniques used by Kliegl et al. (2006)? They all boil down to the following problem: using (for example) the frequencies of word n–1, word n, and word n+1 as separate predictors is not the same thing as independently varying these factors. That is, nouns and verbs, which probably account for most of the lower frequency words, are probably often preceded by high-frequency words (such as articles) and infrequently preceded by other long words. Thus, for example, if the analysis indicates there is an effect of word n-1 or word n+1, one has little idea of how much of the effect is due to this or other confounds. Similarly, it is quite difficult—if not impossible—to sort out how much of the joint variance is due to one or both of two variables.

To make our argument more concrete, we completed two simulations to show how an uncontrolled covariate can reverse the negative correlation that one might otherwise expect to observe between the fixation duration on one word and the predictability of the next. The logic of doing the simulations is to show how—even when the data come from simulated eye movements using a model in which lexical processing is completed in a strictly serial manner—the introduction of a single uncontrolled covariate can produce patterns of data that seemingly require the assumption of parallel lexical processing to explain them. We decided to focus on Kliegl et al.'s finding that single fixation durations on longer words tended to be longer when word n+1 was predictable—a result that Kliegl et al. claim cannot be explained by E-Z Reader. The simulations employed the standard version of E-Z Reader with its default parameter values (Pollatsek et al., 2006). In both simulations, we systematically varied the length and predictability of pairs of adjacent words (words n and n+1) that were always located near the middle of each of the 48 sentences in our corpus (Schilling, Rayner, & Chumbley, 1998). The simulations ran 100 statistical participants for each combination of length and predictability that words n and n+1 were set equal to, resulting in a minimum of 1600 observations per word for each combination. The frequencies of words n and n+1 were set equal to 126 per million, the mean frequency of the words in the Potsdam corpus (Kliegl et al., 2006).

In the first simulation, the predictability of word n was set equal to .2 (the mean predictability of the words in the Potsdam corpus) while the predictability of word n+1 was incrementally varied from 0-1 in .1 increments. The length of word n was also incrementally varied: from 2-6 letters in the first condition (i.e., short words) and from 7-13 letters in the second condition (i.e., long words). The length of word n+1 was varied from 2-13 letters. (These were the same ranges of lengths that were used in Kliegl et al.'s analyses, ignoring the 7 words in the Potsdam corpus that were longer than 13 letters.) The results of our first simulation indicated that, contrary to what Kliegl et al. observed (e.g., see their Fig. 4d), the predictability of word n+1 was not positively related to single-fixation durations on word n, irrespective of whether word n was short ($r = -.016$) or long ($r = -.027$). Taken at face value, these results support Kliegl et al.'s assertion that E-Z Reader cannot account for their finding that single fixation durations on word n tended to be longer if word n+1 was predictable. This conclusion is premature, however, because of the possibility that some other variable (any of the 50 or so variables that affect word recognition; see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) may have co-varied with the length of word n, and may have therefore contributed to the relationship observed by Kliegl et al.

To demonstrate how this might happen, we completed a second simulation in which we introduced an uncontrolled covariate—the predictability of word n. To do this, we simply incorporated the fact that longer words tend to be less predictable than shorter words, setting the predictability of word n equal to .01 in the second (long word) condition. With this one additional assumption, the predictability of word n+1 was now positively correlated to the single fixation durations on word n in the long-word condition ($r = .062$). Although the absolute strength of this predicted relationship is quite modest, it is comparable to that observed by Kliegl et al. (2006): a modest correlation ($r = .05$) between the logit-transformed predictability of word n+1 and single fixation durations on word n[4]. It is also important to note that our manipulation of the word n predictability covariate was a modest one, with the longer (7-13 letter) words being only slightly less predictable than the shorter (2-6 letter) words (.01 vs. .2, respectively). The fact that even such a modest manipulation was able to change the relationship between single fixation durations on word n and the predictability of word n+1 suggests that other factors (e.g., word class, with the longer words being comprised of fewer function words) may have also co-varied between the two types of words and in a similar manner contributed to the observed positive relationship between the single-fixation durations on word n and the "predictability" of word n+1. The point of our simulations is thus very simple. They caution against putting too much stock in correlational analyses because such analyses are likely to be influenced (in possibly complex ways) by (a potentially large number of) uncontrolled variables.

Another problem with the analyses reported by Kliegl et al. (2006) is that all words were included in their analysis. That is, about 25% of their words were 2-3 letters long, and probably consist largely of function words. Given that function words are often skipped, it may be the case that a lot of the fixations that actually land on these words are due to saccadic error and reflect saccades that were intended for neighboring words. The inclusion of such mislocated fixations in their analyses might explain at least some of their spillover and parafoveal-on-foveal effects and might be tantamount to "the tail wagging the dog." They also did not indicate whether the words prior to or following the target words in the analyses were skipped. It is plausible that this variable (i.e., whether or not a fixation was preceded or followed by a skip) could have a strong effect on fixation time on a word. Finally, both younger and older readers were included in the analyses. This may be problematic in the sense that older readers do show some differing eye movement patterns than younger readers (Laubrock, Kliegl, & Engbert, 2006; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006), and this too may contribute noise to the analysis.

E-Z Reader predicts that the fixation preceding a skip will be longer than a fixation not preceding a skip (all else being equal) because those skips that are not due to simple oculomotor error are caused by later eye movement programs canceling earlier ones. Thus, contrary to what Kliegl et al. report, the finding that having a predictable word as word n+1 lengthens fixations on word n is not a new finding, and is accounted for by E-Z Reader because predictable words are simply more likely to be skipped (this was demonstrated in the simulations that were described above). This is also compatible with the observation Kliegl et al. make that fixation durations on word *n* increase significantly with the length of the outgoing saccade. Further, although Kliegl et al. show that their effect is attenuated (but still exists) when skips of word n+1 are eliminated (i.e., in their triplet analysis, see Table A1), this could be due to trials during which skips were attempted but fell short. Intuitively, one might also expect the same pattern in their data when the frequency of word n+1 is varied, as word frequency influences skipping as well; but, it was not observed. However, this could be due to predictability having a stronger influence on skipping rates than frequency (Brysbaert, Drieghe, & Vitu, 2005) and the fact that an increase in skipping rates as a function of the frequency of the word is usually restricted to predictable target words (Rayner, Ashby et al., 2004).

As mentioned, Kliegl et al. provided an additional analysis that excluded skips to attempt to rule out the possibility that the effects they obtained in the original analysis were due somehow to word skipping (their triplet analysis, Table A1). However, this analysis has problems of its own and we believe it provides an excellent example of the confounding which we have already alluded to that can exist in correlational studies. In this analysis, they examined the main effects of frequency, predictability, the inverse of word length (1/length), and viewing position measures for cases where word n-1, n, and n+1 were not skipped (each of the three words was fixated exactly once in first pass reading). Although this analysis revealed lag and successor effects for frequency and predictability that were similar to those of their original analysis, the effect of 1/length for n-1 and n+1 was in the opposite direction of that in the original analysis. That is, in the original analysis, the coefficient of 1/length for n-1 was 15 ms (meaning that the shorter n-1 was, the longer the fixation on word n was, all other things being equal), whereas in the analysis without skips, the coefficient was –29 ms. Furthermore, these conflicting coefficients were highly significant in both sets of analyses. In the appendix they comment that this reversal is due to the fact that the length of skipped words was unrelated to the fixation durations on word n but that the length of fixated words was related to these durations. However, this is not an adequate explanation. If the length of skipped words was unrelated to the fixation durations on word n, then these word lengths should simply add noise to the data (make a significant coefficient less significant), but not change a coefficient from being significant in one direction to being significant in the other direction. Additionally, the effect of the length of word n+1 also changed drastically between the two analyses, going from a non-significant coefficient of -1.2 ms to a highly significant coefficient of 18 ms. In the next section we discuss a more plausible reason for the reversals in Kliegl et al.'s triplet constrained analysis, one that centers on data selection.

The Selection of Single Fixation Cases

In both corpus studies and experimental studies in reading, there are often analyses reported on a subset of the data; they typically examine variations in eye movements given a certain sequence of events. For example, if one wants to look at the effects of the frequency of word n with respect to the preview benefit on the viewing times of word n+1, it makes sense to restrict the analyses to those cases when the saccade that landed on word n+1 originated from word n. By doing so, one not only eliminates cases when word n+1 was fixated as a result of a regression but also cases where word n was skipped. (Regardless of the frequency manipulation on word n we would always expect a reduced preview effect on word n+1 due to the limited visibility at the far-off launch site of that saccade, making the analysis on word n+1 unnecessarily complicated and potentially lacking power.) Likewise, Kliegl et al. (2006) based their main findings on an analysis restricted to those cases when there was a single fixation on the target word. Whereas restricting the analysis to a certain sequence of events can certainly have its advantages, as illustrated above, we feel that in this particular analysis the choice of single fixation cases comes at a rather high cost in terms of being able to interpret the resulting data – a cost we would like to make more explicit. Before doing so, however, we do acknowledge that Kliegl et al. did report data associated with gaze durations and some results were consistent across single fixation and gaze duration measures. Here, we primarily raise the issue as a cautionary note.

Averaged over all words in text, it is true that the most probable event for a word is to be fixated exactly once. However, it is important to note that the probability of a word having exactly one fixation is highly dependent on a number of factors related to properties of that word, notably word length and frequency. First consider word length; single fixations do not represent the default eye movement pattern when the word is either very short or reasonably long. Very long words tend to be re-fixated at a higher rate than shorter words (Rayner, Sereno & Raney, 1996). Kliegl et al. (2004) reported re-fixation rates in the Postdam corpus (see their Figure 1) that were as high as 30% for 9 letter-words (see also Engbert et al., 2005). Again, this raises questions as to whether single fixation durations are by themselves appropriate to represent the eye movement behavior on long words (13% of the words in the corpus were 9 letter-words or longer). We would argue that for these words only a combination of single fixation times and gaze durations can fully represent the eye movement behavior. A similar argument can be made for very short words which tend to be skipped more often than fixated (for a meta-analysis on word length effects on word skipping, see Brysbaert et al. 2005); in the data reported by Kliegl et al. (2004) both two-letter words and three-letter words were skipped with a frequency of 50% or higher (see also Rayner & McConkie, 1976). Moreover, as argued in a prior section, it is likely that a reasonable fraction of the single fixations on these words are saccades that were intended for the prior or following word. We would argue that for these cases, skipping rates need to be taken into account as well as fixation durations, and that it is unclear whether mean fixation durations on those occasions in which the word is fixated is a very meaningful measure. In sum, the data from the single fixation analyses is likely to have fairly reasonable data from average length words, fairly questionable data from short (2-3 letter) words (24% of the words included in the corpus fall into this category), and a biased sample of data for the long words.

There is a similar relationship between the frequency of a word and how often it receives exactly one fixation during reading. A low frequency word tends to be skipped less often (Henderson & Ferreira, 1993; Rayner et al., 1996) and is re-fixated more often (Rayner et al., 1996). Although the size of these effects are not as large as those for word length, the restriction of the analysis to those cases where a word receives a single fixation will also have an impact on the average frequency of the words analyzed. This is very important given the fact that the amount of data collected in the Potsdam corpus is so huge that quite small effects can produce significant differences in the regression coefficients. In sum, restricting the analysis to single fixation events in the way Kliegl et al. (2006) did, can have profound effects on the characteristics of the words in the analysis (i.e., the words in the analysis will quite often be close to the average word length and their frequency will be somewhat higher than the overall frequency of the words in the corpus, see their Table 2). This undermines the suitability of the analysis to represent the eye movement behavior in normal reading as a whole.

This returns us to the discussion of the Kliegl et al. triplet analysis. In this analysis, the problems just discussed are compounded threefold. That is, each of the three words must receive exactly one fixation (can't be skipped or refixated). These types of words are likely to have very specific characteristics. While Kliegl et al. do not report these characteristics (in fairness, the article was already highly detailed), it is likely a large percentage of these words were 5 or 6 letters long with slightly higher than average frequencies (and a narrow range). The likely lack of variability in these words may have caused a severe restricted range problem which is known to be problematic for correlational analyses. This type of problem could actually explain the reversal in regression coefficients that they report. It is clear that conducting a regression analysis on a subset of a subset of the data in order to remove the influence of word skipping is not the same thing as including word skipping in the regression model. And in fact as we have just shown, this type of analysis cannot truly remove the effects of word skipping as the types of words included in such an analysis will be systematically different (due to the types of words that are skipped) from the types of words in the initial analysis. However, we sympathize with Kliegl et al. in that their regression model was already highly complex without adding the extra skipping variable and the necessary interactions with this variable. This frustration highlights the advantage of conducting experiments where many of these variables can be controlled.

We have already demonstrated by means of simulations how in the analyses of Kliegl et al. an uncontrolled covariate can result in effects seemingly at odds with a serial model of eye movement control in reading (i.e., the longer fixation time prior to a predictable word). Combined with the reservations we have discussed here with regards to the selection of single fixation cases, we believe we have made a convincing argument for being very cautious with effects obtained in the single fixation analyses but not replicated in the gaze duration analyses (where we see few observations that can be taken as a serious threat for a serial model of eye movement control in reading, such as the E-Z Reader model) or indeed in carefully controlled experiments.

Minor Points

Finally, we note a few minor points that we feel are of some relevance. First, Kliegl et al. list Rayner (1977, 1978) as a reference for the cognitive lag hypothesis. Although such a notion was discussed in both papers, it was not advocated in either, and arguments against the view were provided. Second, Kliegl et al. note that there is good evidence that long fixations follow long saccades because fixations preceding long saccades provide less preview of the word to be fixated than when the saccade launch site is closer. We fully agree with this assessment. However, we note that over large segments of text, there is no correlation between successive saccade lengths and fixation durations (Rayner & McConkie, 1976). We raise this point as further evidence that one can reach very different conclusions when experimental data dealing with specific target words are used and when correlational data come from large segments of text (consistent with our arguments above). Third, when discussing studies which use experimentally manipulated target words, we believe that Kliegl et al. tend to blur the distinction between studies dealing with preview benefit and studies dealing with parafoveal-on-foveal effects. While we are certain that they understand the difference, naïve readers may not: the former type of experiments deal with the effect of a preview of word n+1 on the processing time associated with that word following a saccade from word n (where a preview of word n+1 could be obtained); the latter type of experiments do not deal with preview benefit at all, but rather deal with the influence of word n+1 on the fixation time on word n. Fourth, Kliegl et al. argue that the eyes land further into words that are predicted from semantic context and they cite a study by Lavigne, Vitu, and d'Ydewalle (2000). However, while it is clear that strong semantic context influences the probability that a word will be skipped (Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek, 2005; Rayner & Well, 1996), two other studies (Rayner, Binder, Ashby, & Pollatsek, 2001; Vonk, Radach, & van Rijn, 2000) found that semantic context does not influence where in a word the eyes land. This latter finding, which we believe is more reliable than the Lavigne et al. result (for a discussion see Rayner et al., 2001), is consistent with the E-Z Reader model. Finally, Kliegl et al. cite a prior study of ours (Balota et al., 1985) to the effect that visually similar non-words are almost as effective parafoveal previews as real words. However, a more recent study of ours (Drieghe et al., 2005) casts doubt on this conclusion.

Summary

In this article, we have questioned a number of the points made by Kliegl et al. (2006). While we are critical of some of the claims that they make, we want to again note that we think that they have provided some interesting analyses. However, we hope that we have made clear that we think that many of the claims that they make are inconsistent with a great deal of other data, and that much of the data that are inconsistent with their claims have been obtained in well-controlled experiments. Again, we realize that no experiment is ever perfect, but to the extent that extraneous variables can be controlled, we suspect that they are preferable to the type of regression analysis techniques that are relied upon by Kliegl et al. (2006). And, finally, consistent with the prior statement, we strongly urge that claims made from regression analysis techniques should not be accepted until confirmed via controlled experimental techniques.

References

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H., & Yap.,M.P. (2004).Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316.

Balota, D.A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364-390.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance*, Oxford: Oxford University Press, pp. 53-77.

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical and contextual factors. *Spanish Journal of Psychology, 5*, 66-77.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *3*1, 954-969.

Duffy, S.A., Morris, R.K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429-446.

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

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

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, *7*, 70-95.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, *47*, 201-221.

Huey, E.B. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.

Juhasz, B.J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1312-1318.

Juhasz, B.J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*, 846-863.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-354.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45, 153-168.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262-284.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12-35.

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye movements during reading. *Neuroscience and Biobehavioral Reviews*, *30*, 872-884.

Lavigne, F., Vitu, F., & d'Ydewalle, G. (2000). The influence of semantic context on initial landing sites in words. *Acta Psychologica*, *104*, 191-214.

Legge, G.E., Hooven, T.A., Klitz, T.S., Mansfield, J.S., & Tjan, B.S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219-2234

Legge, G.E., Klitz, T.S., & Tjan, B.S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, *104*, 524-553.

McConkie, G.W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578-586.

McConkie, G.W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*, 365-368.

McDonald, S.A. (2005). Parafoveal preview benefit in reading is not cumulative across multiple saccades. *Vision Research*, *45*, 1829-1834.

McDonald, S.A., Carpenter, R.H.S., & Shillcock, R.C. (2005). An anatomically-constrained, stochastic model of eye movement control in reading. *Psychological Review*, 112, 814-840.

Niswander-Klement, E. & Pollatsek, A. (in press). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition*.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*, 2201-2217.

Pollatsek, A., Reichle, E.D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1-56.

Pynte, J., & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research*, in press.

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

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, *4*, 443-448.

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

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*, 211-236.

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

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720-732.

Rayner, K., & Bertera, J.H. (1979). Reading without a fovea. *Science*, *206*, 468-469.

Rayner, K., Binder, K.S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, *41*, 943-954.

Rayner, K., Cook, A.E., Juhasz, B.J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, in press.

Rayner, K., & Duffy, S.A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191-201.

Rayner, K., Inhoff, A.W., Morrison, R.E., Slowiaczek, M.L., & Bertera, J.H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 167-179.

Rayner, K., & Juhasz, B.J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, *16*, 340-352.

Rayner, K., Juhasz, B.J., & Brown, S.J. (2006). 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: Human Perception and Performance,* in press.

Rayner, K., & McConkie, G.W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829-837.

Rayner, K., Reichle, E.D., Stroud, M.J., Williams, C.W., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and elderly readers. *Psychology and Aging*, in press.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188-1200.

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

Rayner, K., & Well, A.D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504-509.

Rayner, K., Well, A.D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*, 537-544.

Rayner, K., Well, A.D., Pollatsek, A. & Bertera, J.H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, *31*, 537-550.

Rayner, K., White, S.J., Kambe, G., Miller, B., & Liversedge, S.P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä , R. Radach, and H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp 213-234). Amsterdam: North Holland.

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*, 4-22.

Reichle, E.D., Rayner, K., and Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparison to other models. *Behavioral and Brain Sciences*,*26*,445-476.

Reilly, R., & Radach, R. (2003). Foundations of an interactive activation model of eye movement control in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds)., *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 429-455). Amsterdam: North Holland.

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

Salvucci, D.D. (2001). An integrated model of eye movements and visual encoding. *Cognitive*

Systems Research, *1*, 201-220.

Sereno, S.C., O;Donnell, P.J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 335-350.

Vitu, F., McConkie, G.W., Kerr, P., & O'Regan, J.K. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Research*, *41*, 3513-3533.

Vonk, V., Radach, R., & van Rijn, H. (2000). Eye guidance and the saliency of word beginnings in reading text. In A. Kennedy, R.Radach, D. Heller, & J. Pynte (Eds), *Reading as a perceptual process* (pp. 269-300). Amsterdam: North Holland.

Well, A.D. (1983). Perceptual factors in reading. In K. Rayner (Ed), *Eye movements in reading: Perceptual and language processes* (pp. 141-150). New York: Academic Press.

Yang, S. (2006). An oculomotor-based model of eye movement control in reading: The competition/interaction model. *Cognitive Systems Research*, *7*, 56-69.

Yang, S., & McConkie, G.W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, *41*, 3567-3585.

Yang, S., & McConkie, G.W. (2004). Saccade generation during reading: Are words necessary? *European Journal of Cognitive Psychology*, *16*, 226-261.

Acknowledgments

Preparation of this article was supported by Grant HD26765 from the National Institute of Health and by Grant R305G020006 from the Department of Education Institute of Education Sciences. Denis Drieghe was supported as a postdoctoral fellow of the Fund for Scientific Research (Flanders, Belgium). We thank Ron Borowsky, Barbara Juhasz, Sachiko Kinoshita, Reinhold Kliegl, and Arnold Well for their helpful comments.

Correspondence should be sent to Keith Rayner, Department of Psychology, University of Massachusetts, Amherst, MA 01003 (rayner@psych.umass.edu).

Footnotes

1. Other models include: EMMA (Salvucci, 2001), Glenmore (Reilly & Radach, 2002, 2006), SERIF (MacDonald, Carpenter, & Shillcock, 2005), Mr. Chips (Legge, Klitz, & Tjan, 1997, Legge, Hooven, Klitz, Mansfield, & Tjan, 2002), SHARE (Feng, 2006), and the Inhibition-Competition model (Yang & McConkie, 2001, 2004; Yang, 2006). All of these models are fully implemented.

2. Actually, while it is very clear that low-level orthographic information about word n+1 can influence the fixation duration on word n, there are major discrepancies with respect to the influence of lexical properties of word n+1 on word n (see Rayner & Juhasz, 2004 for further discussion). Our view is that most studies showing lexical parafoveal-on-foveal effects are either based on tasks that mimic reading (but don't have all the characteristics of reading), are due to mis-located fixations, or result from regression analysis techniques as in the Kliegl et al. article.

3. Actually, we must confess some uneasiness with Kliegl et al.'s distinction between what is called natural reading in their article, and their attribution that many experimental manipulations result in what must be unnatural reading. The two examples that they cite with respect to the latter category are (1) all cases in which experimenters identify a target word and then analyze characteristics of the eye movement record with respect to that word and (2) all cases in which there is a gaze-contingent display change (as with the boundary paradigm). This seems like a very arbitrary and artificial distinction. As we noted with respect to their own corpus of words, the data were originally collected in part to analyze fixation times on specific target words (Kliegl et al. 2004). From the point of view of a participant in such experiments, they aren't aware that a word has been designated as a target word by the experimenter, so why this is unnatural reading isn't obvious.

4. In our second simulation, the correlation between the predictability of word n+1 and the gaze

durations on word n was negligible (r = .001)—a finding that is also consistent with Kliegl et al.'s (2006) results. A third simulation in which an additional covariate was introduced (the overall range of the lengths of word n+1 was reduced from 2-13 letter to 2-7 letter in the condition where word n was long) produced correlations between the predictability of word n+1 and the single-fixation versus gaze durations on word n that were even more in-line with Kliegl et al.'s observations: r = .062 versus r = -.012 for single-fixation and gaze durations, respectively.

Figure Caption

Figure 1. An example of a boundary change experiment. In this example, the target word is *boundary*. However, when the reader begins reading the sentence, the last five letters of the target word are replaced with visually similar letters (so that *bourbcng* is initially present). When the reader's eye movement crosses an invisible boundary at the end of the word preceding the target word, *bourbcng* changes to *boundary*. The asterisks represent the location of each fixation (with the numbers indicating the sequence of fixations).

When the reader’s eyes cross an invisible bourbcng location in the text

* * * * *

1 2 3 4 5

When the reader’s eyes cross an invisible boundary location in the text

* * *

6 7 8