Published in final edited form as:

J Exp Psychol Hum Percept Perform. 2008 June ; 34(3): 726–750. doi:10.1037/0096-1523.34.3.726.

Immediate and Delayed Effects of Word Frequency and Word Length on Eye Movements in Reading: A Reversed Delayed Effect of Word Length

Alexander Pollatsek,

Department of Psychology, University of Massachusetts, Amherst

Barbara J. Juhasz,

Department of Psychology, Wesleyan University

Erik D. Reichle,

Department of Psychology, University of Pittsburgh

Debra Machacek, and

Department of Psychology, University of Massachusetts, Amherst

Keith Rayner

Department of Psychology, University of Massachusetts, Amherst

Abstract

Three experiments examined the effects in sentence reading of varying the frequency and length of an adjective on (a) fixations on the adjective and (b) fixations on the following noun. The gaze duration on the adjective was longer for low frequency than for high frequency adjectives and longer for long adjectives than for short adjectives. This contrasted with the spillover effects: Gaze durations on the noun were longer when adjectives were low frequency but were actually shorter when the adjectives were long. The latter effect, which seems anomalous, can be explained by three mechanisms: (a) Fixations on the noun are less optimal after short adjectives because of less optimal targeting; (b) shorter adjectives are more difficult to process because they have more neighbors; and (c) prior fixations before skips are less advantageous places to extract parafoveal information. The viability of these hypotheses as explanations of this reverse length effect on the noun was examined in simulations using an updated version of the E-Z Reader model

Keywords

reading; eye movements; attention; models; E-Z Reader

The frequency and length of a word both influence how long people fixate it when silently reading (Rayner, 1998). Numerous studies have demonstrated that the frequency of a word influences fixation time on that word using several different measures: (a) first-fixation duration, the duration of the first fixation on a word independent of the number of fixations made on the word, (b) single-fixation duration, the duration of a fixation when only a single fixation is made on the word, and (c) gaze duration, the sum of all fixations on a word prior to moving off of the word. The size of the frequency effect is typically on the order of 20–30 ms for first-fixation duration and single-fixation duration and on the order of 50–60 ms for gaze

duration. Although the length of a word typically has a much smaller (if any effect) on the first-fixation and single-fixation duration on that word, it usually has a robust effect on gaze duration. Given that gaze duration sums multiple fixations on a word, it is not surprising that as a target word gets longer, the probability of making more than one fixation on it increases (Rayner & McConkie, 1976), producing increased gaze durations (Rayner, Sereno, & Raney, 1996).

Most of the eye movement research on the effects of variables such as the frequency and length of a word has typically focused on how those variables affect fixation times on that word. Although such research has yielded a considerable amount of data on various reading measures, it is also important to understand how the properties of a word influence eye movement measures on adjacent words as well. Here, we review the results of some studies which are relevant to the research reported here. In each of these studies, properties of adjacent words were varied (though the manipulations typically involved only a frequency manipulation and not a length manipulation).

Rayner, Sereno, Morris, Schmauder, and Clifton (1989) asked readers to read sentences such as the following:

- 1a. The tiger started to growl.
- 1b. The fierce tiger started to growl.
- 2a. A theory will explain the facts.
- 2b. A simple theory will explain the facts.
- 3a. Jerry's acoustic guitar needs new strings.
- 3b. Jerry's electric guitar needs new strings.

In sentences like 1, the target noun (*tiger*) was low frequency (fewer than 10 occurrences per million according to the frequency norms of Francis & Kučera, 1982), and in sentences like 2, the target noun (*theory*) was high frequency (more than 50 per million). The adjectives in sentences like 1b and 2b had frequencies that matched the noun that they modified. However, in sentences like 3a and 3b, the adjective preceding the target noun was either low frequency (*acoustic*) or high frequency (*electric*). (The frequency of the noun was not systematically controlled.) The low and high frequency words were matched on word length. Rayner et al. (1989) found that there were clear frequency effects: Readers looked at low-frequency words longer than high-frequency words. Interestingly, fixation times on target nouns were shorter when there was no adjective than when there was an adjective. Most relevant to the current study, however, there was a spillover effect (Rayner & Duffy, 1986) from the adjective to the noun (i.e., the gaze duration on the noun was longer when it was preceded by a low-frequency adjective than when it was preceded by a high-frequency adjective.)

Experiments by Henderson and Ferreira (1990) and Kennison and Clifton (1995) reported similar spillover effects. Henderson and Ferreira manipulated the frequency of a noun preceding a verb whereas Kennison and Clifton factorially manipulated the frequency of adjacent adjectives and nouns, thereby creating four different conditions. Both experiments obtained frequency spillover effects; Henderson and Ferreira found that verbs preceded by high-frequency nouns were processed faster than verbs preceded by low-frequency nouns, and Kennison and Clifton found that nouns preceded by high-frequency adjectives were processed faster than nouns preceded by low-frequency adjectives. More recently, Slattery, Pollatsek, and Rayner (2007) varied the frequency of three consecutive content words; the three words were either all high frequency or all low frequency (as in the high-frequency sequence *enemy soldiers attacked* versus the low-frequency sequence *rival warriors ambushed*). The size of the frequency effect did not differ significantly across the three words on either first-fixation duration or gaze duration (thus indicating no spillover on immediately succeeding words);

however, there were effects of the frequency manipulation at the end of the sentence, indicating a much later spillover effect. These three studies all suggest that word frequency may have effects beyond initial lexical access in reading.

The studies we have discussed are consistent in indicating frequency effects associated with specific target words. However, in all of the studies, the length of the high- and low-frequency target words was matched. Although word length effects have not been studied as often as frequency effects, it is clear (as we noted at the outset) that the length of a word has a large effect on the gaze duration on that word. Word length also has a large effect on skipping behavior (Brysbaert, Drieghe, & Vitu, 2005; Rayner, 1998; Rayner & McConkie, 1976; Rayner et al., 1996; Vitu, O'Regan, Inhoff, & Topolski, 1995); not surprisingly, short words are much more likely to be skipped than longer words. One study that did factorially vary the length and frequency of target words was reported by Schroyens, Vitu, Brysbaert, and d'Ydewalle (1999); however, the task they used mimicked reading but was not reading per se. They had native Dutch speakers read word triads, and the length and the frequency of the first word were manipulated. The first word was either three letters (short) or five letters (long) and either high or low frequency. The second word (the target word) was seven letters long and was either high or low frequency. Five-letter high-frequency pretarget words were processed faster than five-letter low-frequency pretarget words, but there was no difference in the processing times of the high- and low-frequency three-letter pretarget words. They reported a spillover effect (of 13 ms for frequency), but did not differentiate between three- and five-letter words.

The final study that is relevant to the present experiments is by Kliegl, Nuthmann, and Engbert (2006). They used regression analysis techniques (see also Kennedy & Pynte, 2005; Pynte & Kennedy, 2006) to examine the processing associated with any set of three words in a passage of text. Specifically, they examined fixation time measures, as a function of length, frequency, and predictability of a target word for the target word (word n), as well as for the prior word ($n - 1$) and the following word ($n + 1$). Interestingly, they found effects of these variables on word n , but they also found effects on word $n - 1$ (which they called successor effects) and word $n + 1$ (which they called lag effects, and which are basically spillover effects) due to the properties of word n . Although their finding using these techniques mirrored the studies we described above in finding a substantially larger frequency effect on word n than on word $n + 1$ for gaze duration, they actually obtained a larger word-frequency effect on word $n + 1$ than on word n for first-fixation duration. Although the results reported by Kliegl et al. are quite interesting, we have some reservations about the conclusions that one can draw from such regression analyses (see Rayner, Pollatsek, Drieghe, Slattery, & Reichle, 2007).

The present experiment used adjective–noun pairs to further investigate the effect of the difficulty of word n on the processing time on an adjacent word during on-line sentence processing. Following the work of Henderson and Ferreira (1990) and Kennison and Clifton (1995), the frequency of an adjective was manipulated (high vs. low). Furthermore, an additional, factorial, length manipulation of the adjectives was employed. The short adjectives were 3–4 letters and the long adjectives were 7–9 letters (differing from Schroyens et al.'s comparison of three- and five-letter words); and, in all conditions, the adjectives immediately preceded a noun. This created four distinct conditions: (1) high-frequency long adjectives (*private restaurant*), (2) low-frequency long adjectives (*festive restaurant*), (3) high-frequency short adjectives (*dark restaurant*), and (4) low-frequency short adjectives (*tidy restaurant*). We expected that nouns preceded by high-frequency adjectives would be processed faster than nouns preceded by low-frequency adjectives. Because length also affects foveal difficulty and parafoveal preview, we also expected that nouns preceded by short adjectives would also be processed faster than nouns preceded by long adjectives.

As will be apparent following the presentation of Experiment 1, the results did not match our initial expectations, as we obtained the counterintuitive finding that short adjectives actually led to longer processing times on the target noun. Experiments 2 and 3 were attempts to determine why we obtained this result and our motivation for the experiments (especially Experiment 3) was guided by the E-Z Reader model (Pollatsek et al., 2006c; Rayner, Ashby, Pollatsek, & Reichle, 2004; Reichle et al., 1998; Reichle, Rayner, & Pollatsek, 2003). We present the E-Z Reader model in some detail when we describe our attempts to model these data; however, in our exposition of the experiments, we describe it only in enough detail to motivate the experiments.

Experiment 1

The design of Experiment 1 was straightforward. Two salient attributes of the adjective that should affect fixation times on the adjective were varied factorially: the length of the adjective and its frequency in the language. The usual finding is that, all else being equal, gaze duration on a word is longer if it is (a) lower in frequency or (b) has more letters, even when the other factor is controlled. Moreover, spillover effects, although sometimes small, are usually of the form that any variable that causes fixation time on a word (word n) to be longer will tend to produce a smaller, but similar, effect on the time to fixate the subsequent word (word $n + 1$).

One way this relationship has been explained is in terms of the E-Z Reader model. In the model, the duration of an earlier stage of lexical processing (called the *familiarity check*, or L_1) determines the fixation time on word n and the later stage of lexical processing (called the *completion of lexical access*, or L_2) influences the fixation time on word $n + 1$. If one assumes that the variables, word frequency and word length, affect the two stages of lexical processing similarly (i.e., that L_1 and L_2 are each longer for lower frequency and longer words), then one would predict that the frequency and length effects observed on the adjective should be mirrored (at least qualitatively) by the spillover effects on the noun.

Method

Participants—The 32 participants were students at the University of Massachusetts. They either were paid or received course credit for their participation. They all had normal vision or corrected vision (via soft contact lenses).

Materials and design—The target words were adjective–noun pairs that were embedded in sentences that appeared on a single line on a video monitor. There were 48 adjectives that were divided into 12 sets, with each set containing a high-frequency (HF) long adjective, a low-frequency (LF) long adjective, a high-frequency (HF) short adjective, and a low-frequency (LF) short adjective. See Appendix for the experimental sentences. The following example shows the variations for Sentence 42:

- 4a. Long-HF: Martha witnessed the serious accident at the race track.
- 4b. Long-LF: Martha witnessed the colossal accident at the race track.
- 4c. Short-HF: Martha witnessed the big accident at the race track.
- 4d. Short-LF: Martha witnessed the vile accident at the race track.

The 24 high-frequency adjectives were in the range of 100–1,000 occurrences per million in the Francis and Kučera (1982) word corpus database, with a mean of 197; the 24 low-frequency adjectives were in the range of 0–10 occurrences per million, with a mean of 5, and the 48 nouns were in the range of 40–130 occurrences per million with a mean of 64. The long adjectives were 7–9 letters ($M = 7.8$); the short adjectives were 3–4 letters ($M = 3.8$); and the nouns were 7–10 letters ($M = 7.9$).

As can be seen from the example, the target noun was held constant for each set, so that there were sets of four sentences in which only the adjective differed. However, in order to maximize the power of the experiment, we wanted to get an observation from each participant for each adjective. Accordingly, there were four sentence frames for each set of four adjectives. Each frame had a different noun, and it was paired with all four adjectives (as above), so that another sentence frame for this set was *The war prisoner's [serious/colossal/big/vile] struggle ended victoriously*. The design was a counterbalanced one, so that each participant read each of the adjective–noun combinations once and read each of the sentence frames once, with the particular sentence frame that each adjective–noun pair was read in being counterbalanced across participants. The order of the experimental sentences was randomized for each participant.

All the sentences were tested for naturalness and predictability. The naturalness norms required 10 participants in each of the four counterbalancing conditions to rate sentences on a 7-point scale (1 = *unnatural*, 4 = *somewhat natural*, 7 = *natural*) according to how natural they sounded. There were no significant differences between sentences containing adjectives differing in length or frequency, and there was no significant interaction between length and frequency. In addition, all the sentences satisfied a minimum naturalness criterion of 4.40. The norming items that assessed the predictability of the adjective displayed the stimuli sentences up until the adjective, but not including it, and the participants' task was to provide a possible adjective given the sentence beginning. There were 10 participants in each of the four counterbalancing conditions. None of the adjectives was predictable; the percent completions for the four adjective conditions were 0.7%, 0%, 1.2%, and 0.2% in the long high-frequency, long low-frequency, short high-frequency, and short low-frequency conditions, respectively. High-frequency words were slightly more predictable than low-frequency words, $F(1, 44) = 3.19$, $MSE = 109$, $p = .08$, but there were no significant predictability differences due to length or due to the interaction between the length and frequency of the adjective, $F_s < 1$. The norming items assessing the predictability of the four nouns displayed stimuli sentences past the adjective up to, but not including, the noun. The participants' task was to provide a possible noun given the sentence beginning. The nouns were also not predictable; the average percent of completions for the noun following the four types of adjectives were 0.46%, 0.46%, 0.35%, and 0.40% in the long high-frequency, long low-frequency, short high-frequency, and short low-frequency conditions, respectively. No predictability differences of the noun due to the length, the frequency, or the interaction of the preceding adjective's length or frequency were significant, $F_s < 1$. None of the participants in the norming studies participated in the eye-tracking experiments.

Apparatus—A Fourward Technologies Dual Purkinje Eye-tracker (Buena Vista, VA), having a resolution of 10 min of arc, recorded all eye movements. The eye tracker was interfaced with a 486 microcomputer that ran the experiment. Participants' viewing was binocular, but only eye movements from the right eye were recorded. The computer sampled eye movements every millisecond. All sentences presented on the monitor were on a single line not exceeding 80 letters. During the experiment, participants were seated 61 cm from the monitor, where 3.8 letters equal 1° of visual angle. All letters were presented in lower case except when uppercase was called for (sentence beginnings or proper names). The brightness on the computer screen was adjusted to a comfortable level for each participant and was held constant throughout the experiment.

Procedure—When participants first arrived for the experiment, a bite bar was prepared to eliminate all head movements. Participants were given a set of written instructions in addition to being verbally informed of all procedures before the start of the experiment. After the initial calibration of the eye tracker (which took no longer than 5 minutes) subjects were instructed to read each sentence for meaning.

At the beginning of each trial, a row of five calibration boxes appeared across the middle of the screen. Participants were instructed to look at the center box, blink (in order to minimize track losses due to blinks on the sentence), and then to look left one box at a time, until the leftmost box where the beginning of each sentence would appear. If the initial calibration still remained accurate, meaning that a small red light following the participant's eyes was centered within each box as the participant looked at each box, the experimenter would present a sentence to the participant. If the initial calibration no longer remained accurate, the experimenter would recalibrate before presenting a sentence. Subjects read 6 practice sentences before reading the experimental sentences: 48 target sentences mixed with 96 filler sentences. For the experimental sentences, participants were asked to respond to simple yes–no questions after 20% of the sentences. Responses were recorded by *yes* or *no* lever bar presses on a response box. Participants responded to the questions accurately 96% of the time.

Results

Data were excluded from the analyses if there was (a) a track loss on the adjective or noun target regions (usually caused by a blink), (b) if the first fixation on the sentence was on one of the target regions (all adjective–noun pairs were embedded in the middle of the sentence), (c) if the first fixation on the sentence started more than three words from the beginning, and (d) if a subject made 20 or more fixations on a single sentence. Altogether, 6.64% of the data were lost for the reasons listed above.

Many different indices of eye movements in reading were computed. However, to clarify presentation, we focus on a few key indices that capture most of how the adjective and noun were processed. (Our later discussion includes a few more.) To capture the initial encoding of the target words, our primary focus was on the gaze-duration measure (the total fixation time on a word before it is exited for the first time). Two alternative measures of somewhat earlier processing on a word are first-fixation duration (the mean of the duration of the first fixation on a word) and single-fixation duration (the mean fixation duration on a word, conditional on the word being fixated exactly once). All three of these measures are conditional on the first fixation on a word occurring before any fixation to the right of that word. For the sake of completeness, the total fixation duration (sum of all fixations on a word, including regressions) is presented in the tables, but we do not discuss it further.

Fixation time on the adjective—As seen in Table 1, both the length and the frequency of the adjective had substantial effects on the time spent fixating the word. On average, gaze durations were 42 ms longer on long adjectives than on short adjectives, $F_1(1, 31) = 61.62$, $MSE = 964$, $p < .001$; $F_2(1, 47) = 42.47$, $MSE = 2,270$, $p < .001$, and gaze durations on low-frequency adjectives were 20 ms longer than those on high-frequency adjectives, $F_1(1, 31) = 13.39$, $MSE = 990$, $p < .001$; $F_2(1, 47) = 16.36$, $MSE = 1,675$, $p < .001$. However, as indicated by a significant Frequency $\times$ Length interaction, $F_1(1, 31) = 17.33$, $MSE = 1,016$, $p < .001$; $F_2(1, 47) = 13.04$, $MSE = 1,795$, $p = .001$, the frequency effect was confined to the long adjectives, as there was a 44-ms frequency effect for the long adjectives but there was actually a –3-ms frequency effect for the short adjectives.

The pattern was similar for the other measures; however, the effects were smaller for the two early measures, especially the length effect. For first-fixation duration, the 7-ms length effect was marginally significant, $F_1(1, 31) = 3.15$, $MSE = 486$, $p < .10$; $F_2(1, 47) = 3.84$, $MSE = 1,107$, $p < .10$, but neither the 6-ms frequency effect nor the interaction were close to significant, $F_1(1, 31) = 1.77$, $MSE = 434$, $p < .20$; $F_2(1, 44) = 3.18$, $MSE = 790$, $p < .10$; and $F_1(1, 31) = 2.40$, $MSE = 328$, $p < .20$; $F_2(1, 44) = 1.27$, $MSE = 810$, $p > .20$, respectively. Even though the means for single-fixation duration were similar to those for first-fixation duration, the effects were more reliable, with a significant 11-ms length effect, $F_1(1, 31) = 5.35$, $MSE = 618$, $p < .$

05; $F_2(1, 47) = 8.96$, $MSE = 1,293$, $p < .01$, a marginally significant 6-ms frequency effect, $F_1(1, 31) = 2.64$, $MSE = 413$, $p < .20$; $F_2(1, 47) = 4.24$, $MSE = 927$, $p < .05$, but no significant interaction, $F_1(1, 31) = 1.42$, $MSE = 376$, $p > .20$; $F_2(1, 47) = 1.08$, $MSE = 895$, $p > .20$.

Probability of skipping the adjective—Most strikingly, the short adjectives were skipped with a much higher probability than were the long adjectives, 25% vs. 2%, $F_1(1, 31) = 121.1$, $MSE = 134$, $p < .001$; $F_2(1, 44) = 68.6$, $MSE = 212$, $p < .001$. There was also a marginal 2.8% frequency difference in skipping rates, $F_1(1, 31) = 2.59$, $MSE = 92$, $p < .20$; $F_2(1, 47) = 4.07$, $MSE = 120$, $p < .05$, and, although there was a suggestion that the frequency effect on the skipping rate was larger for the short adjectives, the interaction was not close to significant, $F_1 < 1$; $F_2(1, 47) = 1.42$, $MS = 171$, $p > .20$. Moreover, the data from the next two experiments indicate it is unlikely that this interaction is a real effect.

Fixation time on the noun—The most striking effect was that of adjective length on the gaze duration on the noun (see Table 2), which was in the “wrong direction”: Processing times on the noun were longer for short adjectives; the effect was only 14 ms, but it was significant, $F_1(1, 31) = 5.80$, $MSE = 1,034$, $p < .025$; $F_2(1, 47) = 4.23$, $MSE = 2,428$, $p < .05$. In contrast, there was a significant (nonreverse) length effect that was about the same size for first-fixation duration (16 ms), $F_1(1, 31) = 13.32$, $MSE = 617$, $p = .001$; $F_2(1, 47) = 16.67$, $MSE = 790$, $p < .001$, and single-fixation duration (15 ms), $F_1(1, 31) = 12.63$, $MSE = 565$, $p = .001$; $F_2(1, 47) = 9.34$, $MSE = 1,176$, $p < .005$. This pattern suggests that the reverse length effect is occurring relatively late during the period when the noun is initially fixated. There also appeared to be a small spillover effect of the frequency of the adjective in the appropriate direction (i.e., longer fixation times on the noun if there was a low-frequency adjective). However, it was not significant on any measure, although the 12-ms effect on gaze duration was marginally significant, $F_1(1, 31) = 2.88$, $MSE = 1,412$, $p = .10$; $F_2(1, 47) = 3.02$, $MSE = 1,875$, $p < .10$. The interaction between length and frequency was not close to significant on any measure (all $F_s < 1$) and, as seen in Table 2, the pattern was not consistent across the various measures.

Probability of skipping the noun—The noun was virtually always fixated (see Table 2), and no effect on the probability of skipping it was close to significant ($F_s < 1$).

Later measures that reflect regressing back to the adjective—For reasons that are expanded on below, we were also interested in even later effects of the adjective beyond the first-pass time on the noun. Although total time on the noun is one such measure, it does not directly reflect regressing back to the adjective, as there could be a regression back to the adjective followed by a skip of the noun. We report two such measures in Table 2. The first, the probability of regressing back from the noun, was 3.2% greater for low-frequency adjectives than for high-frequency adjectives, $F_1(1, 31) = 4.25$, $p < .05$; $F_2(1, 47) = 5.09$, $p < .05$, and 1.9% greater for short adjectives than for long adjectives, $F_1(1, 31) = 2.37$, $p < .20$; $F_2(1, 47) = 1.85$, $p < .20$ ($F_s < 1$ for the interactions). The second measure, go-past time, is generally defined as the sum of all fixation durations between the time that the noun is first fixated and when the eyes first move to the next word; thus, it counts the fixation durations of all regressions from the target word. However, in this case, we adopted a more restrictive definition and only counted regressions back to the adjective in computing the measure. Thus, it was the gaze duration plus any fixation time in going back to the adjective plus any refixation time on the noun before going forward from it. As a result, our go-past measure can be viewed as a fixation measure that combines gaze duration and this regression measure. The pattern of results for this measure is similar to the regression measure. The go-past time for the noun was 33 ms longer for low-frequency adjectives, $F_1(1, 31) = 9.32$, $p < .005$; $F_2(1, 47) = 12.40$, $p < .001$, and 12 ms longer for short adjectives, $F_1(1, 31) = 2.17$, $p < .20$; $F_2(1, 47) = 1.19$, $p > .20$ ($F_s < 1$ for the interactions).

Additional analysis—We did one additional analysis on the gaze duration on the noun to try to understand the reverse length effect more fully: Gaze durations on the noun were analyzed conditional on there being exactly one fixation on the adjective (see the values in parentheses in Table 2). In this analysis (as can be seen from the table), the length effect was 8 ms (i.e., not reversed), but not close to significant, $F_1(1, 31) = 1.28$, $MSE = 1,470$, $p > .20$; $F_2(1, 47) = 1.55$, $MSE = 2,778$, $p > .20$. It thus appears that the reversed length effect is due, at least in part, to the trials on which the adjective is skipped.

Discussion

The effects in Experiment 1 on the adjective were about what would be expected. First, the fixation data on the adjective showed the expected effects of frequency and length that one generally observes. Although these two variables are obviously confounded in the language, they have each been shown to have independent effects on fixation times (Rayner et al., 1996). The one slightly odd feature of these data was that there was no frequency effect for the short adjectives. The failure to find such an effect might have been due to the high skipping rate for the adjectives; however, there was no evidence for a differential skipping rate for the high-frequency and low-frequency short adjectives that would offer an obvious mechanism for why this high skipping rate would eliminate the frequency effect for the short adjectives.

The effect of the frequency of the adjective on the gaze duration (and other measures) on the noun were in line with typical spill-over effects, although they were admittedly weak. However, spillover effects are frequently not all that large (e.g., Henderson & Ferreira, 1990; Drieghe, Rayner, & Pollatsek, 2007). In contrast, the length manipulation on the adjective had a significant and quite unexpected result. That is, increasing the length of the adjective, which had a quite substantial effect on the gaze duration on the adjective, had a reverse effect on the processing time on the noun: longer adjectives, which were fixated longer, actually led to shorter fixation times on the subsequent noun. We were originally quite mystified as to what could have produced such an effect, but ultimately thought of three plausible explanations for such an effect. We quickly sketch them now in order to motivate Experiments 2 and 3 but reserve a full discussion of them for the General Discussion section.

The first explanation is that early and later stages of word identification may be differentially affected by different processing variables. As indicated above, the prediction in the E-Z Reader model that spillover effects should generally mirror fixation duration effects on a target word rests on the assumption that a manipulated variable should affect early and later stages of word identification in a similar manner, at least qualitatively. However, this may not be true for word length, at least for these materials. A notable difference between our long and short adjectives is that their orthographic neighborhood characteristics differed. For example, using Coltheart's (1981) N measure (the number of words that are the same length as the target word but differ by exactly one letter), the short words have many more neighbors than the long words. The mean N values were 0.1, 0.5, 10.2, and 8.7, for the long high-frequency, long low-frequency, short high-frequency, and short low-frequency adjectives, respectively. A second neighborhood measure that is implicated in inhibitory effects on later processing of words is the number of higher frequency neighbors. Here again, there are similar differences between the long and short adjectives on this measure: 0, 0, 1.0, and 4.3 for the long high-frequency, long low-frequency, short high-frequency, and short low-frequency adjectives, respectively. Thus, if having more neighbors and/or having more high-frequency neighbors (i.e., competitors) slows down later stages of lexical access, one might expect the observed inverse length effect if this inhibitory effect comes very late in processing. In fact, Perea and Pollatsek (1998) found that there were inhibitory effects of having higher frequency neighbors (the number of higher frequency neighbors was varied while holding length constant), and Pollatsek, Perea, and Binder (1999) found that there were similar inhibitory effects due to

having more neighbors. In both cases, these inhibitory effects were largely quite late (e.g., on regressions back to the target word and spillover).¹ Thus, our observed inverse length effect could plausibly be due either to a neighborhood-size or to a number-of-higher-frequency-neighbors effect. We elaborate on this hypothesis later, as it was a motivation for Experiment 3.

A second possible reason for the inverse spillover effect is that the difference in length between the adjectives had an effect on the pattern of eye movements, and this differential pattern of eye movements accounted for the inverse spillover effect. There are two (not mutually exclusive) intuitively plausible mechanisms for how eye fixation locations could be less optimal after short adjectives. The first is related to the differential skipping rates of the long and short adjectives. That is, although some of this differential skipping rate may be due to it being easier to identify the short adjectives in the parafovea, some of it may be due to errors in eye movement programming. That is, because readers tend to overshoot shorter intended saccades (McConkie, Kerr, Reddix, & Zola, 1988), readers may often plan to fixate the short adjective but overshoot it. Thus, they may often be fixating the noun when they are still attending to and processing the adjective. The second mechanism is that, for related reasons due to differences between targeted fixation location and actual fixation location, readers may fixate in different places on the noun after long and short adjectives, and thus may be in a somewhat less advantageous place on the noun following a short adjective. We also discuss this hypothesis in more detail later, after we have presented the results of all three experiments.

The third possible mechanism for the reverse length effect is that the reader obtains, on average, less parafoveal preview information about the noun in the short-adjective condition because of the high skipping rates in that condition. That is, in the short-adjective condition, on a significant fraction of the trials, the fixation prior to the noun was from the word before the adjective, whereas in the long-adjective condition, the fixation prior to the noun is from the adjective and hence closer to the noun. The fact that the reversed length effect on the noun went away when trials were removed on which the adjective was skipped suggests that this mechanism may account for much of the reversed length effect.

As we will see, removal of these trials did not eliminate the reversed length effect in either Experiments 2 or 3. However, the major motivation for Experiment 2 was to try to minimize the skipping rates for the adjectives to determine whether doing so would eliminate the reverse length effect observed in Experiment 1. The specific manipulation we employed was a version of the boundary technique (Rayner, 1975) in which the adjective was replaced by random letters until the readers fixated on it. As a result, readers were denied an opportunity to encode the adjective until they fixated it, which should reduce skipping rates considerably. If this also eliminates the reverse length effect on the noun, it would imply that the latter was mainly due to the higher skipping rates for the short adjectives.

Experiment 2

Method

Participants—The 28 participants were drawn from the same population as in Experiment 1.

¹There is a recent report of a failure to replicate Perea and Pollatsek (1998) by Sears, Campbell, and Lupker (2006). However, when they used the same materials as Perea and Pollatsek, they obtained late inhibitory effects similar to those in Perea and Pollatsek except that the effects were not significant. They also reported null effects when using different materials; however, there are problems with these materials, such as the target words often being the second word in the sentence.

Stimuli and design—The stimuli were identical to those in Experiment 1, and the counterbalanced design was the same as well. The only difference was that participants were prevented from encoding the adjective until they fixated it using the boundary-change technique mentioned above. In the experiment, the (invisible) boundary was the beginning of the space before the adjective. Prior to the eyes crossing the boundary, the letters of the target word were replaced by a preview of random letters. During the saccade when the eyes crossed the boundary, the preview was replaced by the adjective. An example of a boundary change is given in Figure 1. As vision is suppressed during a saccade and little low-level visual information is retained from fixation to fixation, participants were rarely, if ever, aware of the display change.

Apparatus and procedure—These were the same as in Experiment 1 except for the boundary-change manipulation described above.

Results

Trials were excluded in the same manner as in Experiment 1. In addition, trials were excluded if the display changes did not occur with the correct timing (i.e., it occurred too early or too late). Altogether, 18.75% of the data were lost.

Fixation time on the adjective—As can be seen in Table 3, there were clear effects of both length and frequency on gaze durations on the adjective. Gaze durations on the adjective were 46 ms longer for the long adjectives than for the short adjectives and 34 ms longer for the low-frequency adjectives than for the high-frequency adjectives, $F_1(1, 27) = 27.30$, $MSE = 2,115$, $p < .001$; $F_2(1, 47) = 29.53$, $MSE = 3,041$, $p < .001$; $F_1(1, 27) = 23.10$, $MSE = 1,418$, $p < .001$; $F_2(1, 47) = 25.61$, $MSE = 2,432$, $p < .001$, respectively. (The Length $\times$ Frequency interaction was not close to significant, $F_s \leq 1$.)

There were also significant frequency effects on first-fixation duration (17 ms) and single-fixation duration (24 ms), $F_1(1, 27) = 17.00$, $MSE = 474$, $p < .001$; $F_2(1, 47) = 13.30$, $MSE = 1,115$, $p = .001$; $F_1(1, 26) = 17.27$, $MSE = 926$, $p < .001$; $F_2(1, 47) = 19.01$, $MSE = 1,876$, $p < .001$, respectively. The 8-ms length effect was not significant for first-fixation duration, $F_1(1, 27) = 2.01$, $MSE = 777$, $p = .168$; $F_2(1, 47) = 1.14$, $MSE = 1,292$, $p = .290$, but the 18-ms effect on single-fixation duration was, $F_1(1, 26) = 10.57$, $MSE = 859$, $p = .003$; $F_2(1, 47) = 8.33$, $MSE = 1,810$, $p = .006$. The frequency effect appeared to be larger for the short adjectives on the two early measures and larger for the long adjectives on the two later measures, but the Length $\times$ Frequency interaction was not significant on any of the measures (most $F_s < 1$, all $F_s < 3.15$).

Probability of skipping the adjective—As expected, denying a preview of the adjective produced lower skipping rates on the short adjectives than in Experiment 1; however, there still was a 12.6% skipping rate for the short adjectives, and the 11.8% difference in skipping rate between the short and long adjectives was significant, $F_1(1, 27) = 26.17$, $MSE = 146$, $p < .001$; $F_2(1, 47) = 63.04$, $MSE = 116$, $p < .001$. As can be seen in Table 3, as in Experiment 1, there is a suggestion of a frequency effect for the short adjectives, but this cannot be due to the actual frequency of the word, as the adjective was not visible before it was fixated or the preview was skipped. The fact that there is still a differential skipping rate between the long- and short-adjective conditions (even when what is skipped is a meaningless and unpredictable string) indicates that at least part of the differential skipping rate is merely due to the length of the string (i.e., due to oculomotor error) and not either to the speed of encoding the adjective or to guessing it. We return to this issue in depth later.

Fixation time on the noun—The manipulation of denying a preview for the adjective appeared to enhance the frequency effects on the adjective, especially for the short adjectives, by forcing processing of the adjective to wait until it was fixated. A reflection of this is that gaze durations on the adjectives, on average, were 41 ms longer in Experiment 2 than in Experiment 1. In spite of these overall processing differences in the time spent fixating the adjective and the probability of skipping it, the gaze durations on the noun following the adjective were virtually unchanged from those in Experiment 1 (see Table 4). There was a 19-ms effect of the frequency of the adjective on the gaze duration on the noun and also a 19-ms reverse length effect; however, both effects were only marginally significant, $F_1(1, 27) = 3.84$, $MSE = 1,066$, $p < .10$; $F_2(1, 47) = 4.06$, $MSE = 2,934$, $p < .05$; and $F_1(1, 27) = 4.55$, $MSE = 2,086$, $p < .05$; $F_2(1, 47) = 3.33$, $MSE = 4,246$, $p < .10$, respectively. The interactions were not close to significant ($F_s < 1$). The first- and single-fixation measures on the noun showed a somewhat different pattern, as the fixation durations were shorter in the short-high-frequency-adjective condition than the other three conditions; however, the only effect that was close to significant was a (nonreversed) length effect on first-fixation duration, $F_1(1, 27) = 3.52$, $MSE = 673$, $p < .10$; $F_2(1, 47) = 4.29$, $MSE = 1,485$, $p < .05$.

Probability of skipping the noun—As with Experiment 1, the noun was almost always fixated (see Table 4) and there were no differences among conditions ($F_s < 1$).

Later measures reflecting regressing back to the adjective—The pattern was different here than for Experiment 1 in that the reverse adjective-length effects were large and significant and the adjective frequency effects were quite small. The percent of time that there was a regression back to the adjective was 6.1% greater for short adjectives, $F_1(1, 27) = 10.69$, $p < .005$; $F_2(1, 47) = 10.35$, $p < .005$, and 2.1% greater for high-frequency adjectives (however, $F_s < 1$). Similarly, the go-past time on the noun was 40 ms longer for short adjectives, $F_1(1, 27) = 9.84$, $p = .005$; $F_2(1, 47) = 8.31$, $p = .01$, but it was only 14 ms longer for the low frequency adjectives (however, $p_s > .20$). There appeared to be interactions in both measures, such that the length effect was bigger in the high-frequency-adjective condition; however, it was only marginally significant for the regressions, $F_1(1, 27) = 3.96$, $p < .10$; $F_2(1, 47) = 4.02$, $p < .10$, and not close to significant for the go-past measure, $F_1(1, 27) = 1.95$, $p < .20$; $F_2(1, 47) = 1.69$, $p > .20$.

Additional analysis—Although we replicated the reverse length effect of Experiment 1 in gaze duration on the noun in this experiment, the effect just failed to reach significance. Again, we thought it would be instructive to determine how much of the effect was due to skipping the adjective and did an analysis of the gaze duration on the noun conditional on there being exactly one fixation on the adjective (see the values in parentheses in Table 4). Unlike Experiment 1, however, the reverse effect of the length of the adjective on the noun did not go away, but instead increased. There was, on average, a 32-ms reverse adjective-length effect, $F_1(1, 26) = 8.70$, $MSE = 3,146$, $p < .01$; $F_2(1, 47) = 15.41$, $MSE = 4,128$, $p < .001$, and a 23 ms adjective frequency (spillover) effect, $F_1(1, 26) = 5.38$, $MSE = 2,640$, $p < .05$; $F_2(1, 47) = 5.95$, $MSE = 4,304$, $p < .025$. However, there was clearly only a frequency effect in the short-adjective conditions, and the Length $\times$ Frequency interaction was significant, $F_1(1, 26) = 6.74$, $MSE = 2,312$, $p < .025$; $F_2(1, 47) = 6.39$, $MSE = 4,096$, $p < .025$. This analysis indicates that the reversed length effect was not merely due to actually skipping the adjective, at least under these display conditions.

Discussion

The major motivation for Experiment 2 was to determine whether the reverse length effect on fixation times on the noun was likely to be due to the very high skipping rates for the adjectives in Experiment 1. This turned out not to be the case, as the overall reverse length effect in

Experiment 2 was about the same as that in Experiment 1 even though skipping rates were quite a bit lower in Experiment 2. Moreover, when we analyzed gaze durations on the nouns conditional on the adjective not being skipped, there was a significant reverse length effect, although it largely came from the low-frequency-adjective conditions.

We defer further discussion of these results until after presenting Experiment 3. As indicated above, one hypothesis for the reverse length effect is that it was due either to the short adjectives having many more neighbors than the long adjectives, the short adjectives having more higher frequency neighbors than the low-frequency long adjectives, or both. We also indicated that these neighborhood effects are likely to appear late in processing and that, in the Perea and Pollatsek (1998) and Pollatsek et al. (1999) data, they mainly appeared in spillover effects.

It might be helpful to distinguish two possible mechanisms for such neighborhood effects. The first is the one we have mentioned above: that these effects are in the normal, but later, stages of encoding a word and that neighborhood effects primarily affect later stages. For example, in the activation-verification model of Paap, Newsome, McDonald, and Schvaneveldt (1982), the first stage is one in which the letters seek out lexical entries and the second is one in which the lexical system computes which excited lexical entry is the “winner.” In such a model, having more neighbors or more high-frequency neighbors inhibits later stages. The second mechanism for inhibitory neighborhood effects on spillover is that some of the time a word is misperceived as a neighbor and, although sentence context will usually quickly make clear that the encoding was wrong, this correction process lengthens processing on the subsequent word.

We thought that an interesting further exploration of the reverse length effect would be to do another boundary experiment, this time denying a preview of the noun (instead of the adjective). If the first neighborhood explanation of the reverse length effect is correct (i.e., that the short adjectives having more neighbors is affecting later normal stages of word processing), then the reverse length effect should go away when there is no preview of the noun. That is, in this view, in normal reading conditions, the longer, later stages of processing would continue on the adjective, so that there is less processing of the noun before it is fixated. In other words, there is less “preview benefit” from the noun in the short-adjective condition than in the long-adjective condition. If this view is correct, then, by preventing normal parafoveal processing of the noun, any of the reverse word-length effect that would result from there being less parafoveal processing of the noun in the short-adjective condition would be attenuated. It is less clear what the other neighborhood explanation (i.e., that more neighbors or more high-frequency neighbors causes momentary misidentification of a word) would predict. One possibility is that denying a preview might instead increase the effect because processing of the following noun is delayed by the lack of preview. This may allow misidentification of the adjective to be more definite, and hence make it harder for the reader to correct the misidentification when the noun is fixated.

Experiment 3

Method

Participants—The 40 participants were drawn from the same population as in Experiments 1 and 2. No one participated in more than one of the three experiments.

Stimuli and design—The stimuli were identical to those in Experiments 1 and 2, and the counterbalanced design was the same as well. The only difference was that participants were prevented from viewing the noun until they fixated it using the boundary-change technique mentioned above. (There was no display change involving the adjective.) The procedure was

exactly as in Experiment 2 except that the (invisible) boundary was defined to be the beginning of the space before the noun. As in Experiment 2, the preview was a string of random letters.

Apparatus and procedure—These were the same as in Experiment 2 except for the different location of the boundary.

Results

Trials were excluded in the same manner as in Experiments 1 and 2. Eight additional participants were excluded from analysis due to an excessive amount of data loss. Altogether, 7.60% of the data were lost.

Fixation time on the adjective

As seen in Table 5, there was a 52-ms length effect and a 49-ms frequency effect on the gaze duration on the adjective, $F_1(1, 39) = 40.41$, $MSE = 2,660$, $p < .001$; $F_2(1, 47) = 32.58$, $MSE = 4,145$, $p < .001$; and $F_1(1, 39) = 30.72$, $MSE = 2,545$, $p < .001$; $F_2(1, 47) = 33.10$, $MSE = 2,668$, $p < .001$, respectively. There was a somewhat larger frequency effect for the long adjectives, although the interaction was only marginally significant, $F_1(1, 39) = 3.52$, $MSE = 2,405$, $p < .10$; $F_2(1, 47) = 3.16$, $MSE = 3,387$, $p < .10$. The pattern for single-fixation duration was similar to that observed in Experiments 1 and 2, with a marginally significant 12-ms length effect, a 26-ms frequency effect, and a nonsignificant interaction, $F_1(1, 39) = 4.03$, $MSE = 1,530$, $p = .052$; $F_2(1, 47) = 3.16$, $MSE = 2,632$, $p < .10$; $F_1(1, 39) = 12.48$, $MSE = 2,187$, $p = .001$; $F_2(1, 47) = 14.19$, $MSE = 1,911$, $p = .001$; and $F_s < 1$, respectively. The pattern for first-fixation duration was somewhat different with a 19-ms frequency effect, $F_1(1, 39) = 9.55$, $MSE = 1,558$, $p = .004$; $F_2(1, 47) = 9.63$, $MSE = 1,652$, $p = .003$, and virtually no length effect or interaction ($F_s < 1$). One other aspect of the adjective data should be noted: The fixation times on the adjective, overall, were about the same as in Experiment 2 (e.g., the mean gaze duration in Experiment 3 was actually 6 ms longer than that in Experiment 2) even though readers had a preview of the adjective. It is also worth noting that slower reading in Experiment 3 was not confined to the adjective. Readers in Experiment 3 had longer overall sentence reading times (3,002 ms) compared to both Experiment 1, (2,193 ms), $t(70) = -5.89$, $p < .001$, and Experiment 2, (2,722 ms), $t(66) = -1.93$, $p = .058$.

Probability of skipping the adjective

Again, there was a large difference in skipping rates between the short and long adjectives (20.6% vs. 2.4%), $F_1(1, 39) = 69.23$, $MSE = 191$, $p < .001$; $F_2(1, 47) = 138.99$, $MSE = 118$, $p < .001$, but no hint of either a frequency effect or an interaction ($F_s < 1$).

Fixation time on the noun

As indicated above, the main focus in Experiment 3 was to determine what the elimination of the preview of the noun would do to the reverse length effect. The answer is straightforward; the effect was larger in Experiment 3 (45 ms) than in either Experiment 1 (12 ms) or Experiment 2 (19 ms). For gaze duration, both this reverse effect of adjective length and the 25-ms effect of adjective frequency were significant, and there was virtually no interaction, $F_1(1, 39) = 34.58$, $MSE = 2,335$, $p < .001$; $F_2(1, 47) = 29.28$, $MSE = 2,935$, $p < .001$; $F_1(1, 39) = 15.00$, $MSE = 1,454$, $p < .001$; $F_2(1, 47) = 11.76$, $MSE = 2,247$, $p = .001$; and $F_s < 1$, respectively (see Table 6).

There was a 14 ms adjective frequency effect on both first-fixation duration and single-fixation duration, $F_1(1, 39) = 4.86$, $MSE = 1,441$, $p < .05$; $F_2(1, 47) = 6.86$, $MSE = 1,383$, $p = .025$; and $F_1(1, 38) = 4.63$, $MSE = 1,476$, $p = .05$; $F_2(1, 47) = 2.34$, $MSE = 1,648$, $p < .20$, respectively. For first-fixation duration, there was virtually no effect of adjective length or interaction

between adjective frequency and length ($F_s < 1$). For single-fixation duration on the noun, there was a hint of a reverse length effect, although this effect did not reach significance, $F_1(1, 38) = 3.23$, $MSE = 2,046$, $p < .10$; $F_2(1, 47) = 3.19$, $MSE = 1,648$, $p < .10$. There was also a nonsignificant interaction between adjective length and frequency, $F_1(1, 38) = 1.79$, $MSE = 1,675$, $p < .20$; $F_2(1, 47) = 3.86$, $MSE = 1,988$, $p < .10$. The nature of this interaction is that single-fixation durations on the noun were shorter when the adjective was long and high frequency compared to the other three conditions.

Probability of skipping the noun

As with Experiments 1 and 2, the noun was almost always fixated (see Table 6), and there were no differences among conditions ($F_s < 1$).

Later measures reflecting regressing back to the adjective

The pattern was similar to that of Experiment 2 in that there was a 4.4% reverse adjective-length effect on the percent of time there was a regression back to the adjective, $F_1(1, 39) = 7.26$, $p < .01$; $F_2(1, 47) = 7.29$, $p < .01$, and a 51-ms reverse length effect on the go-past time on the noun, $F_1(1, 39) = 33.04$, $p < .001$; $F_2(1, 47) = 16.17$, $p < .001$. Similar to Experiment 2, there were actually 2.6% more regressions back to high-frequency adjectives, $F_1(1, 39) = 2.38$, $p < .20$; $F_2(1, 47) = 1.91$, $p < .20$, but the go-past time on the noun was in the “right” direction: 20 ms longer for low-frequency adjectives, $F_1(1, 39) = 3.05$, $p < .10$; $F_2(1, 47) = 4.36$, $p < .05$. For the regression measure, there was a suggestion of an interaction, $F_1(1, 39) = 3.08$, $p < .10$; $F_2(1, 47) = 1.05$, $p > .20$, but in the go-past measure, there was little interaction ($F_s < 1$).

Additional analysis

As in the previous experiments, we did an analysis of the gaze durations on the noun conditional on there being exactly one fixation on the adjective (see the values in parentheses in Table 6). The pattern of data was similar to the unconditional analysis above except that both the reverse length effect (33 ms) and the frequency effect (15 ms) were somewhat smaller, and again there was no interaction, $F_1(1, 39) = 18.15$, $MSE = 2,452$, $p < .001$; $F_2(1, 47) = 22.02$, $MSE = 4,288$, $p < .001$; $F_1(1, 39) = 3.49$, $MSE = 2,479$, $p < .10$; $F_2(1, 47) = 5.55$, $MSE = 3,227$, $p < .05$; and $F_s < 1$, respectively.

Discussion

The major finding of Experiment 3 was that denying a parafoveal preview of the noun did not reduce the reverse length effect but, instead, increased it. This is inconsistent with the hypothesis that part of the reduced length effect was caused by neighbors of the adjective delaying later stages of lexical processing and thus reducing the amount of processing of the noun that could be accomplished parafoveally. However, it is consistent with the other neighborhood hypothesis: that an adjective having a higher frequency neighbor and/or many neighbors may sometimes be (at least tentatively) misidentified as a neighbor, producing a time-consuming correction process on the noun. It seems plausible that delaying encoding of the noun could increase the difficulty of this correction process, as the memory trace of the correct encoding may weaken with increasing delay. However, as we posited earlier that the reverse length effect may be due, at least in part, to the length of the adjective modulating where people land on the noun, we need to present these landing-site data. We do so here because it is best to see the data from all three experiments together.

Landing position on the noun

First consider the mean landing position on the noun. (Remember that the nouns were 7–10 letters long with a mean of 7.9.) The participants in fact landed, on average, significantly further into the noun when preceded by the longer adjectives in all three experiments. The means for the long-versus short-adjective conditions were, respectively, 4.2 vs. 3.1 letter positions into the noun (Experiment 1), $F_1(1, 31) = 45.3$, $MSE = .30$, $p < .001$; $F_2(1, 47) = 35.6$, $MSE = .62$, $p < .001$; 4.0 vs. 3.5 (Experiment 2), $F_1(1, 27) = 13.4$, $MSE = .65$, $p < .001$; $F_2(1, 47) = 24.0$, $MSE = .55$, $p < .001$; and 3.5 vs. 2.9 (Experiment 3), $F_1(1, 39) = 31.6$, $MSE = .44$, $p < .001$; $F_2(1, 47) = 45.9$, $MSE = .37$, $p < .001$, respectively. Although these mean differences may not seem large enough to cause substantial differences in the speed of processing the noun, the distributions of landing positions suggest otherwise. That is, as shown in Figure 2, these mean differences are largely due to there being substantially more fixations in the short-adjective condition near the beginning of the noun—likely a very nonoptimal place to fixate a 7–10 letter word.

This pattern makes sense given a general fact about eye movements in reading mentioned earlier: There is a bias for programmed short saccades to be longer than intended and for programmed long saccades to be shorter than intended (McConkie et al., 1988). A consequence of this is that fixations that were intended for the short adjectives will tend to overshoot the short adjectives and land on the first few letters of the noun. One way to test this hypothesis is to consider only those trials on which the adjective was not skipped (see Figure 3). In fact, on these trials, there was virtually no difference in the mean landing position between the short and long adjectives: Experiment 1, 0 letters, $F_s < 1$; Experiment 2, 0.1 letters, $F_s < 1$; Experiment 3, 0.1 letters, $F_1 < 1$, $F_2(1, 47) = 1.71$, $p < .20$. Inspection of Figure 3 indicates that there remains only the slightest suggestion of the increased probability of fixation near the beginning of the word observed in the overall data. These analyses indicate that part of the reversed length effect is likely due to misplaced fixations in the short-adjective condition, but the fact that there was still a significant reversed length effect in Experiments 2 and 3 when trials were removed when the adjective was skipped indicates that these misplaced fixations cannot be the entire explanation. However, to be really sure of that conclusion (as there was still a slight tendency for fixations in the short-adjective conditions to be higher in the initial two positions) we did analyses of go-past time on the noun and the probability of regressing back from the noun in Experiment 3 conditional on the noun not being initially fixated on either the first two or last two letters. These analyses indicate that both effects were at least as large as those in Table 6 and still significant. The length effect on go-past time was 51 ms (it was 51 ms in the main analysis), $F_1(1, 38) = 6.93$, $p < .025$; $F_2(1, 45) = 7.81$, $p < .01$, and the regression effect, $F_1(1, 39) = 6.84$, $p = .025$; $F_2(1, 45) = 4.68$, $p < .05$, was actually larger than in the main analysis (7.2% vs. 4.4%) in the participant analysis, although it was about the same size in the item analysis.² Thus, we feel quite confident that little of the reverse length effect in the analyses conditional on the adjective being fixated is due to the length of the adjective influencing the initial landing position on the noun.

General Discussion

The three experiments all provided evidence for the standard word-frequency and word-length effects on fixation durations on the target adjectives. All three also showed relatively small frequency spillover effects on the subsequent noun (ranging from 12 ms to 25 ms on gaze duration on the noun), which are typical of the size of frequency spillover effects reported in the literature (Drieghe, Rayner, & Pollatsek, 2007; Henderson & Ferreira, 1990). The novel

²Gaze duration also continued to show a significant reverse length effect for Experiment 3 when trials were removed contingent on the initial landing position not being on the first two or last two letters of the noun (34ms, $F_1(1, 39) = 5.85$, $MSE = 3,268$, $p = .02$; $F_2(1, 45) = 4.32$, $MSE = 5,253$, $p = .043$).

finding of these experiments, which are the focus of our discussion, is that the length of the adjective produced a reversed effect on the gaze duration on the subsequent word (i.e., there were longer gaze durations on the noun following shorter adjectives). In order to facilitate comprehension of the discussion below, Table 7 summarizes the reversed length effects found in the three experiments; it also indicates the standard errors of the effects to help evaluate which differences between experiments are likely to be of importance.

Before launching into an attempt to quantitatively model the reversed-length-effect data, we briefly summarize what we consider to be the plausible mechanisms for it and then review the pattern of data over the three experiments to arrive at a preliminary evaluation. (Unless otherwise stated, the measure being discussed is the gaze duration on the noun.) Above, we have proposed five mechanisms, the first two related to biases in saccade programming, the second two related to neighborhood properties of the short adjectives, and the last one related to the quality of parafoveal preview information about the noun. The five mechanisms were the following: (1) inadvertently fixating the noun—intending to fixate the short adjective but overshooting it—leading to lengthened gaze durations on the noun due to continuing processing of the adjective; (2) being in a bad location on the noun—landing near the beginning of the noun in the short-adjective condition; (3) lengthening of later stages of lexical processing of the adjective due to neighborhood properties of the short adjectives—thus lessening the amount of processing of the noun before it is fixated; (4) misidentification of short adjectives due to neighborhood properties—leading to needed repair of the error while fixating the noun; (5) getting poorer parafoveal preview information about the noun in the short-adjective condition because the adjective is skipped more often in this condition and hence the fixation prior to fixating the noun is further away from it. In addition, we should point out that Mechanism 4 has two possible consequences that we should distinguish: (4a) consequences directly related to repairing the misidentification and (4b) delayed processing of the noun because processing time is being usurped by repair of the misidentification of the adjective. (These two submechanisms are implemented separately in the modeling.)

The pattern of results makes it quite unlikely that any one of these mechanisms can be the sole cause of the reverse length effect. However, as we concluded earlier from the results of Experiment 3, the third mechanism is unlikely to explain any of the reverse length effect because preview of the noun was denied by the boundary manipulation and, if anything, the reversed length effect increased. It is also clear that the first mechanism, unintended skipping of the adjective, cannot be the entire explanation for the effect, although it is likely to account for the effect in part. That is, if this mechanism were the sole cause of the reverse length effect, the effect should have gone away in the gaze duration analyses conditional on the adjective being fixated. Although the reverse length effect did go away in this conditional analysis in Experiment 1, it did not in either Experiment 2 or 3. (The reverse length effect was somewhat smaller in the conditional analysis in Experiment 3, but it was still significant.) Moreover, the bad landing position hypothesis, although likely a contributor to the reverse length effect, cannot be the sole explanation for it because the effect was still present in our analyses of Experiment 3 conditional on the landing position being good. Thus, although it appears that a significant part of the reverse length effect is due to eye guidance (either the short adjectives being inadvertently skipped or inducing a nonoptimal landing position on the noun), it appears that at least part of the effect is due to some differing lexical property of the long and short adjectives. As indicated above, we feel the most plausible candidate for this differing lexical characteristic is Mechanism 4, differences that might result from differing neighborhood properties of the long and short adjectives. Similarly, Mechanism 5 cannot be the entire explanation for the reversed length effect because the effect was still there in Experiments 2 and 3 even when we removed trials on which the adjective was skipped.

We thus think that the circumstantial evidence that some neighborhood property difference between the long and short adjectives (i.e., Mechanism 4) is a significant component of the reversed length effect is quite strong. Among other things, the pattern of the phenomenon is quite similar to the pattern of interference effects observed in reading both when neighborhood size (N ; Pollatsek et al., 1999) and the number of higher frequency neighbors were manipulated (Perea & Pollatsek, 1998)—including both increased fixation times on the following word, and, especially, more regressions back to the target word. However, we are far from confident that there is a single predictor variable that accounts for these effects across all three experiments as the pattern of data was not consistent across the three experiments. For example, for go-past time, the reversed length effect was larger for low-frequency adjectives in Experiment 1, but it was larger for high-frequency adjectives in Experiments 2 and 3 (although none of the interactions were significant—see Table 7). Because the high-frequency short adjectives had many fewer high-frequency neighbors than the low-frequency short adjectives, this makes the number of higher frequency neighbors improbable as the sole cause of the reversed length effect. (The long adjectives had virtually no neighbors.) On the other hand, the high-frequency short adjectives did have, on average, a few more neighbors than the lower frequency short adjectives (10.2 vs. 8.7), so that N could be viewed as a reasonable predictor of Experiments 2 and 3, where the reversed length effect was slightly bigger for the high-frequency adjectives. In addition, the correlation over items of the reversed length effect on go-past times in Experiments 2 and 3 with N was about .3 for both high- and low-frequency adjectives, which also suggests that N (the number of neighbors) may be a reasonable predictor of at least part of the effect.

There are many aspects of neighborhood effects that are far from well-understood. One example is that several recent analyses (e.g., Davis, 2005) indicate that only counting the types of neighbors that the standard N indices deal with (i.e., substitution neighbors) is likely to be inadequate to assess how confusable a word is with other words. For example, there is evidence that transposition neighbors (e.g., *trial* and *trail*) behave like substitution neighbors (Johnson, Perea, & Rayner, (2007). In addition, there is also the likelihood that addition–deletion neighbors (e.g., *tail* and *trail*) may have neighborhood properties, and the issue of whether phonological properties of neighbors are important has not been fully explored in reading. We should emphasize, however, that by all of these measures, the short adjectives have many more “neighbors” that they can be confused with than the long adjectives. (In our discussion below, we use neighborhood size as shorthand for some property related to a word having many words that are orthographically or phonologically similar to it.) Another issue is why the reversed length effect was appreciably bigger in Experiments 2 and 3 than in Experiment 1. Intuitively, however, it seems reasonable that the display changes in Experiments 2 and 3, by delaying processing, make it harder to correct a temporarily misencoded word.

All of the discussion to this point, however, involves qualitative arguments, and it is instructive to determine whether these mechanisms can quantitatively account for the reverse length effect, especially using reasonable parameters for various word processing times. Accordingly, we introduce the E-Z Reader model, which we think is the simplest current reading model, to test the adequacy of these mechanisms more carefully. However, because previously published versions of the model (Pollatsek et al., 2006c; Reichle et al., 1998, 2003) do not have any mechanism to deal with misidentification of words and subsequent repair, we completed the simulations that are described below using a new version of our model, E-Z Reader 10 (Reichle, McConnell, & Warren, 2007), which was specifically developed to deal with such effects. Because E-Z Reader 10 is a direct extension of its predecessor, we first describe E-Z Reader 9 and then discuss how it was augmented in the development of E-Z Reader 10.

Simulations With the E-Z Reader Model

The E-Z Reader model is actually a family of computational models that has been developed to explain eye-movement behavior during reading. There are two core assumptions about word encoding that are common to all versions of the model that also make the model different from other current models of eye-movement control (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Feng, 2006; Inhoff, Eiter, & Radach, 2005; McDonald, Carpenter, & Shillcock, 2005; Yang, 2006).³ The first assumption is that lexical processing occurs serially, so that only one word—whatever word is being attended—is lexically processed at a time. The second is that the completion of an early stage of lexical processing on one word provides the signal to the oculomotor system to begin programming an eye movement to the next word. Figure 4 is a schematic diagram of the model. The light gray boxes indicate the components of E-Z Reader 9 (as described in Pollatsek et al., 2006c) and the dark gray boxes show the two additional components that were added to produce E-Z Reader 10 (Reichle, McConnell, et al., 2007). Each of the model's components are briefly described below (for a more complete description of the model, see Reichle, McConnell, et al., 2007).

The first component is a preattentive stage of visual processing (labeled V in Figure 4), during which information from the entire visual field is propagated in parallel from the retina to the brain, although the quality of the featural information decreases rapidly as one gets further from the fixation point. The low-spatial frequency information that is available from this stage of processing is used for selecting the target for the next saccade and the high-spatial frequency information is the raw material for further lexical processing. (Based on recent estimates of the eye-mind lag, we assume that V takes 50 ms to complete.)

Most central to the interpretation of the present experiments, lexical processing is done in the word-identification system in two “stages”: L_1 and L_2 (see Figure 4). The earlier L_1 stage corresponds to the initial processing of a word (sometimes referred to as a familiarity check) and completion of this stage triggers a program to saccade to the next word. The rationale for this assumption is that completion of the L_1 stage of the word currently attended to indicates that word identification is sufficiently advanced so that it is safe to move on. The later L_2 stage corresponds to the point in lexical processing when the word has been identified sufficiently so that it can be integrated into the sentence and discourse structure; the completion of L_2 also triggers an attention shift to the next word, so that L_1 of that word can begin.⁴ Crucially, the time to complete both stages is modulated by factors that influence the difficulty of lexical processing, such as a word's frequency, its length, and its predictability from prior text. The theoretical significance of this assumption is as follows. Because L_1 is the trigger for the eye movement and is a function of the difficulty of processing a word, the model predicts that fixation durations on a target word are influenced by variables such as its frequency and length. In contrast, the duration of L_2 is largely irrelevant to the fixation time on the target word but has implications for the fixation time on the next word. In particular, because the time to program a saccade is independent of the difficulty of a word, the time that the reader has to process the next word before actually fixating it is determined by L_2 : The longer L_2 is, the later attention switches to the next word, and hence the less lexical processing is done before the word is fixated. This allows the model to explain how the difficulty associated with processing one word can “spill over” onto the next (Henderson & Ferreira, 1990; Rayner & Duffy, 1986).

The model that has been described thus far (i.e., the components that are represented by the light gray boxes in Figure 4) corresponds to E-Z Reader 9 (Pollatsek et al., 2006c); E-Z Reader

³For a recent survey of existing models of eye-movement control during reading, see the 2006, vol. 7 special issue of *Cognitive Systems Research* (Reichle, 2006).

10 also contains these components and only differs from its predecessor in two important ways (corresponding to the two components represented by the dark gray boxes). The first is the inclusion of an explicit attention-shifting stage (A in Figure 4). The assumption here is that the completion of lexical processing of word n causes attention to shift to word $n+1$. In E-Z Reader 9, this attention shift was only implicit in the model, with the time required to shift attention from word n assumed to be part of the time that is required to complete lexical processing of word n (Pollatsek, Reichle, & Rayner, 2006b). In E-Z Reader 10, the shifting of attention is a discrete, explicit process that requires time to complete. For simplicity, the mean time to complete A, $t(A)$, was assumed to be 20 ms, with the actual time during any simulated attention shift being sampled from a gamma distribution having a standard deviation equal to .22 of its mean, in a manner identical to what is assumed about $t(L_1)$ and $t(L_2)$ (see Footnote 4). Although 20 ms might seem too short to warrant inclusion, our assumption about the duration of this process is consistent with empirical estimates, which suggest that it takes anywhere from 4 ms to 33.3 ms to shift attention per degree of visual angle (Eriksen & Schultz, 1977; Jolicoeur, Ullman, & Mackay, 1983; Posner, 1978; Shulman, Remington, & McLean, 1979; Tsal, 1983). (The assumption that it takes 20 ms to shift attention is thus consistent with these estimates because, in eye-tracking reading experiments like the ones reported in this article, each degree of visual angle typically corresponds to 3–4 character spaces.) Although the attention-shift assumption does not improve the model's ability to fit data, it was added to address criticisms that instantaneous attention shifts are implausible (Inhoff et al., 2005; Radach, Deubel, & Heller, 2003) and because it plays an important functional role in the second new assumption of E-Z Reader 10, which is described next.

The second way in which E-Z Reader 10 differs from its predecessor is that it includes a component that corresponds to the postlexical integration ("I" in Figure 4) of a word's meaning into whatever representation of the sentence has been constructed by the reader up to that point in time. As Figure 4 indicates, I is initiated for a given word by the completion of L_2 on that same word; that is, the reader attempts to integrate the meaning of a word as soon as it is available. This integration takes some time (on average, 50 ms) to complete, with the actual time being sampled from a gamma distribution (see Footnote 4). The integration of word n may begin while the eyes are on word n or word $n + 1$, but will usually be completed after the eyes have moved to word $n + 1$, with the integration happening in parallel with the on-going lexical processing of word $n + 1$. If word n is integrated before the meaning of word $n + 1$ has been accessed, then nothing happens. In such cases, lexical processing of word $n + 1$ continues,

⁴This is not to say that L_1 and L_2 should necessarily be viewed as corresponding to discrete stages of lexical processing; rather, they should be viewed as corresponding to the amounts of lexical processing that are sufficient for the initiation of a saccade to the next word (in the case of L_1) and for the initiation of an attention shift to the next word (in the case of L_2). In the model, the mean time (in ms) to complete L_1 for a given word, $t(L_1)$ is a function of both its frequency of occurrence in text and its within-sentence (cloze task) predictability, as follows:

$$t(L_1) = \alpha_1 - \ln(\text{frequency}) \alpha_2 - \text{predictability} \alpha_3 \quad (1)$$

where α_1 , α_2 , and α_3 are free parameters that modulate the rate of lexical processing. A word's length affects $t(L_1)$ indirectly, in that limited visual acuity slows the processing of long words or words that are located far from the center of vision (where visual acuity is greatest). The mean time to complete L_2 , $t(L_2)$, is a fixed proportion of $t(L_1)$, as follows:

$$t(L_2) = t(L_1) \Delta \quad (2)$$

where Δ is a free parameter. The actual times required to complete $t(L_1)$ and $t(L_2)$ for any given Monte Carlo simulation trial is determined by sampling random deviates from gamma distributions with means defined by Equations 1 and 2 and standard deviations equal to .22 of the means. In E-Z Reader 9, the values of the aforementioned parameters were $\alpha_1 = 122$, $\alpha_2 = 4$, $\alpha_3 = 10$, and $\Delta = .5$; in E-Z Reader 10, the values were $\alpha_1 = 104$, $\alpha_2 = 2$, $\alpha_3 = 10$, and $\Delta = .5$. For a complete, more formal, description of the two versions of the model and their assumptions and interpretations of their parameters and their values, see Pollatsek et al. (2006c) and Reichle et al. (2007).

and the meaning of word $n + 1$ is then integrated as soon as it becomes available. However, if the meaning of word n is not integrated before the meaning of word $n + 1$ has been accessed, then the integration stage will affect eye movements. But how? The two answers that we entertained in our simulations are that such cases can result in (a) the redirection of both attention and the eyes back to the source of integration difficulty, word n ; and (b) the slowing of lexical processing of word $n + 1$, possibly due to “interference” that results from the difficulty associated with integrating the meaning of word n . Of course, other scenarios are also possible; for example, the very rapid failure to integrate the meaning of word n (e.g., if it’s meaning is semantically anomalous; Rayner, Warren, Juhasz, & Liversedge, 2004) might cancel a saccade that would otherwise continue to move the eyes forward, resulting in a pause. Such scenarios are not examined in the present article; we instead focus on only the former two cases in order to evaluate their efficacy in accounting for our experimental results.

As one might guess, our assumptions regarding the I stage of processing are extremely simplistic and are not meant to be a “deep” model of whatever postlexical linguistic processing is necessary to actually understand sentences (e.g., we make no attempt to describe how a word’s meaning is actually incorporated into the meaning of a sentence). The assumptions are instead meant to function as a “placeholder” for such a model, so that the framework of E-Z Reader can be used to test hypotheses about how such high-level processing might interact with lexical processing and/or saccadic programming and thereby influence readers’ eye movements. Thus, our goal in adding the assumptions about I was to develop a computational framework for thinking about how the type of higher level processing that operates in the “background” of on-going lexical processing might occasionally intervene to influence the normal, forward progression of eye movements during reading. Simulations using E-Z Reader 10 (reported in Reichle, McConnell, et al., 2007) indicate that, with mean durations of 50 ms or less, the I stage has only negligible effects on simulated eye movements, but that with longer durations of I (durations that would correspond to the type of processing difficulty that might be associated with integrating a word’s meaning) there are pauses and/or regressions back to the source of integration difficulty. The assumptions about postlexical integration in E-Z Reader 10 thus seemed like an appropriate starting point for examining how the lexical processing and/or integration difficulties that might have been associated with the misidentification of short adjectives could have affected the patterns of eye movements that were observed in our experiments.

As already stated, the remaining assumptions of E-Z Reader 10 are identical to those of its predecessor, and are related to the programming and executing of saccades.⁵ These assumptions are discussed here only briefly because they are not central to the concerns of the current experiments. (For a more complete description of both the model and the phenomena that it accounts for, see Pollatsek et al., 2006c.) As Figure 4 indicates, saccades are programmed in two stages: a labile stage (M_1) that can be canceled by the initiation of subsequent saccadic programs, followed by a nonlabile stage (M_2) that is not subject to cancellation. The inclusion of two stages of saccadic programming is supported by independent research (Becker & Jürgens, 1979) and allows the model to explain word skipping: Assuming that the eyes are initially fixated on word $n - 1$, the completion of L_1 on word n will cause the oculomotor

⁵With only two exceptions, all of the parameter values that were used in the simulations that are reported in this article were the same as those used in previously published simulations (Pollatsek et al., 2006c). These exceptions are as follows: First, the parameters that determine the amount by which launch-site fixation durations modulate the systematic part of saccadic error were adjusted so as to attenuate the amount of saccadic error. (In previously published simulations, $\Omega_1 = 7.3$ and $\Omega_2 = 3$; in the current simulations, $\Omega_1 = 7$ and $\Omega_2 = 3.5$.) This change was necessary because preliminary simulations indicated that the model, with its default parameter values, overpredicted the amount of simulated adjective skipping in Experiment 2. Second, the parameter (R) that determines how much time must pass to initiate a “corrective” saccade following a misplaced fixation (i.e., initial fixations that are located far from the optimal viewing position) was reduced under the assumption that the feedback that is necessary to determine if a saccade is misplaced is based on an efference copy of the motor program (Carpenter, 2000) and not visual information. (In previously published simulations, $R = 117$ ms; in our current simulations, $R = 0$ ms, thus being effectively eliminated.)

system to initiate a program to move the eyes to word $n + 1$, which will in turn cancel any pending labile program that would otherwise move the eyes to word n , causing word n to be skipped. Finally, saccades are assumed to take a fixed amount of time (25 ms) to complete, and we assume that the target for all saccades is the middle of the targeted word. Crucially, as we have indicated above, the model not only assumes that where the saccade lands is subject to random (i.e., Gaussian) motor error, but that there is systematic motor error: Short saccades tend to overshoot their targets and long saccades tend to undershoot their targets. This is consistent with a large body of research (e.g., McConkie et al., 1988; O'Regan, 1990; Rayner et al., 1996).

Our goal was to come up with a coherent explanation of the reverse length effect that would not be at variance with the other effects we observed. Our first step was to see whether a saccadic programming-error explanation would indeed give a good account of the data. As mentioned earlier, there is bias in saccadic programming in reading such that saccades tend to overshoot near targets. Thus, readers may tend to skip the short adjectives even when they intend to fixate them; and, if so, some of the time spent fixating on the subsequent noun would be spent attending to and processing the adjective. However, it is clearly of interest to see whether a model that instantiates such an assumption can in fact account for our data.

To do this, we ran a series of simulations that were designed to test the plausibility of our explanations of the observed results and, in particular, the reverse length effect. The method that we used in completing the simulations was the one that we have used elsewhere (Pollatsek et al., 2006a, 2006b, 2006c; Rayner et al., 2004, 2007; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006): Rather than using the actual sentence materials that were used in our experiments, we instead used a corpus of 48 sentences that were used by Schilling, Rayner, and Chumbley (1998) to examine word-frequency effects in reading, manipulating the frequencies, lengths, and predictabilities of pairs of adjacent words in these sentences to determine how, in our experiments, these manipulations affected the patterns of first-fixation durations and gaze durations that were observed on the adjectives and nouns, as well as the observed probabilities of skipping the adjectives and of making regressions from the nouns.

Our rationale for adopting this method is threefold. First, it would have been prohibitively expensive to use our actual experimental materials in our simulations because doing so would have first made it necessary, among other things, to collect cloze-task predictability norms for all of the words in our sentences. Second, in order to evaluate the effects of our manipulations using our actual experimental materials, it would have also been necessary to first find the values of our model's free parameters that would allow it to optimally simulate our participants' overall eye-movement behavior; this requirement would have also been prohibitively expensive (for a discussion of how best-fitting model parameters are selected, see the appendix of Reichle et al., 1998). Because the Schilling et al. corpus has often been used with the E-Z Reader model, our decision to use this corpus in the present simulation effectively avoids these two costs. Finally, our decision about which dependent measures to simulate was based on our a priori belief that our capacity to explain our data largely depended upon our capacity to explain these key results.

Our general approach was to introduce assumptions and to select parameter values so as to determine if they would be sufficient to affect the simulated eye movements in the manner that is congruent with our hypotheses about the reverse word-length effect. Our simulations can thus be viewed as a form of hypothesis testing (Seidenberg & Plaut, 2006), where the framework of E-Z Reader 10 was used to determine if one or more of the mechanisms that we postulated earlier are sufficient to reproduce the effect. The fact that we were able to introduce assumptions and change parameter values "by hand" is not a weakness of our approach, but is instead a strength because it demonstrates the conceptual transparency of our model (Rayner,

Pollatsek, & Reichle, 2003) and how it can be used as a heuristic to examine how a variety of variables (e.g., lexical ambiguity: Reichle, Rayner, & Pollatsek, 2007) affect readers' eye movements. With this caveat in mind, we now turn to the simulation results.

Each simulation that is reported below was based on 1,000 statistical subjects and was completed (unless otherwise noted) using all of the model's default parameter values (Pollatsek et al., 2006c). In the simulations, the properties of the adjectives and nouns were set to values equal to those used in the experiments. The frequencies of the high- and low-frequency adjectives were thus set equal to 197 and 5 per million, respectively, and the frequencies of the nouns were set equal to 64 per million. The predictabilities of all of the target words were set equal to zero, which is probably not quite right, but is a reasonable approximation. The lengths of the adjectives varied from 7–9 letters in the long conditions and from 3–4 letters in the short conditions; the lengths of the nouns varied from 7–10 letters in all four conditions. Finally, to allow for more direct comparisons between the observed and simulated data, we also relaxed the constraint in our simulations that Monte Carlo trials containing interword regressions due to saccadic error (i.e., regressions caused by intended refixation saccades launched from the ends of words towards the centers of the words that overshoot their targets) would be excluded from our analyses. (This was done because, as we have already stated, one of our goals was to simulate the patterns of regressions back from the nouns that were observed in the various conditions of Experiments 1–3.)

Figure 5 shows the results of our first two simulations, which represent successive attempts to simulate the pattern of results that was observed in Experiment 1. Panel A of the figure shows the mean simulated first-fixation durations, gaze durations, and skipping probabilities for the adjectives, along with the mean observed values of these three dependent measures (taken from Table 1 to make a comparison easier). Panel B shows the mean simulated first-fixation durations, gaze durations, and regression probabilities for the nouns, again with the observed values (from Table 2) for comparison.

Simulation 1 was completed using the “standard” version of E-Z Reader 10 with its default parameter values and is included here as a “baseline” to show how the model fared in handling the observed results without the inclusion of any additional assumptions. The only assumption that was specific to the task being simulated was that, with some low probability ($p = .05$), the meaning of a given word (word n) could not be integrated prior to the completion of lexical processing of the next word (word $n + 1$) and that this resulted in both the eyes and attention being directed back (with equal probability) to one of the two sources of processing difficulty—either to word $n + 1$ (whose meaning was just accessed) or to word n (whose meaning could not be integrated).

Although this last assumption might seem both ad hoc and overly complicated, it was meant to reflect the fact that readers might have difficulty localizing the precise source of their comprehension difficulty and/or that readers might adopt the “strategy” of keeping their eyes and attention in the general vicinity of processing difficulty (i.e., word n , word $n + 1$, or the boundary between these two words) so as to gain additional time to reprocess either or both words. Although a computational account of how the failure to integrate a word's meaning affects on-going lexical processing is clearly beyond the scope of what we intend to accomplish in this article, this assumption (as we shall show below) was sufficient for our purposes, providing an account of our data. Also, it is important to note that this assumption was applied to *all* of the words in *all* of the sentences, not just the target words (i.e., the adjectives and the nouns) that were the focus of our analyses. (In other words, this assumption is not specific to the target words, which, from the perspective of our participants, were probably no different than any of the other words in the sentences.)

A comparison of the observed and simulated means indicates that the model does reasonably well, especially when one considers that no attempt was made to adjust any of the model's parameters for this set of readers. As Panel A of Figure 5 shows, the model does a reasonably good job predicting all three dependent measures for the adjectives, predicting the correct qualitative patterns of gaze durations, first-fixation durations, and skipping probabilities across the four conditions. In addition, the model provides qualitatively accurate predictions for two of the three dependent measures for the nouns, predicting the correct patterns of gaze durations and first-fixation durations. As Panel B of Figure 5 shows, however, the model predicts too few regressions back from the nouns, especially in the short-adjective conditions. Despite this shortcoming, it is important to note that the model does predict some of the reverse length effect that was observed on the nouns, although the predicted effect sizes are a bit too small (i.e., observed = -19 ms vs. predicted = -10 ms in the low-frequency condition; observed = -8 vs. predicted = -7 in the high-frequency condition).

Overall, the results of Simulation 1 support the hypothesis that part of the reverse length effect that was observed in our experiments was due to the fact that the short adjectives were skipped more often than long adjectives and that the increased rate of skipping in the short conditions resulted in more time being spent processing the adjective while fixating the nouns, thereby inflating the gaze durations on the nouns. This hypothesis is consistent with four facts. The first is that the model skipped short adjectives (mean probability = .22) much more often than long adjectives (mean probability = .02). The second is that the model showed a slight tendency to move its eyes further into the nouns following saccades from long adjectives (mean landing position = 3.54 letters from the beginning of the target word) than short adjectives (3.38) because fewer of the fixations on the nouns following long adjectives were due to skipping. Both of these findings are broadly consistent with what was observed in Experiment 1 and suggest that the short adjectives may have been skipped more often due to saccadic error, thus making it more likely that the participants would have to finish processing the short adjective when fixated on the noun. This explanation is directly supported by the third fact about the simulation: An analysis of which word the model was attending to when its eyes landed on the noun (i.e., whether it was processing the adjective or noun) indicated that the model was still processing the adjective much more often when it was short (25.0% of trials) than when it was long (0.1%). Finally, we examined the predicted gaze durations on the nouns conditional upon the adjectives being fixated exactly once, so as to eliminate trials involving adjective skipping. The results of this conditional analysis (shown in Figure 5, Panel B) indicate that the elimination of trials involving skipping was sufficient to completely remove the reverse word-length effect, in a manner that was consistent with what was observed in Experiment 1.

As already mentioned, with one notable exception (the probability of making regressions from the nouns following short adjectives), the model did a fairly good job predicting the overall observed pattern of results. One hypothesis about why the participants in Experiment 1 made so many regressions from the nouns is that, because the short adjectives had denser orthographic neighborhoods than the long adjectives, the short adjectives may have been misidentified more often than the long adjectives. This misidentification may have caused difficulty with the integration of the meanings of the short adjectives, which in turn may have caused readers to move their eyes back to the source(s) of processing difficulty.

To examine the hypothesis about the possible effects of unequal orthographic neighborhood density on postlexical integration, we ran a second simulation (Simulation 2 in Figure 5) in which, with some probability ($p = .13$), the meanings of the short adjectives (word n) would not be integrated prior to the completion of lexical 1), resulting in both the eyes access (L_2) of the nouns (word $n +$ and attention being directed back to one of the two sources (with equal probability) of processing difficulty (i.e., the adjectives or the nouns). As in the previous simulation, we also assumed that readers would, with some lower probability ($p = .05$), also

have difficulty integrating the meanings of other words in the sentences. Thus, the only difference between Simulations 1 and 2 is that, in the latter, we assumed that readers would have some additional difficulty integrating the meanings of the short adjectives. The results of this simulation are also shown in Figure 5.

As one can see by comparing the results of Simulation 2 to those of Simulation 1 and the observed means, our added assumption that the misidentification of the short adjectives (due to their neighborhood properties) sometimes caused problems with the integration of those words was sufficient for the model to predict the full pattern of results that was observed in Experiment 1. That is, with this assumption, the model now does a fairly good job of predicting the three dependent measures on both the adjectives and the nouns (including regressions) in all four conditions. Perhaps more interesting, however, is that the model now does a better job simulating the effect of adjective length, with robust reverse length effects being evident in both the low-frequency (observed = -19 ms vs. predicted = -17 ms) and the high-frequency (observed = -8 ms vs. predicted = -13 ms) conditions. The results of Simulations 1 and 2 thus suggest that two factors associated with the adjectives (i.e., whether or not they were skipped and difficulty with their integration—perhaps due to the relative density of their orthographic neighborhoods) are likely to be necessary to fully explain the reverse length effects that were observed in our experiments. Finally, to more explicitly test the hypothesis that differential skipping of the adjectives contributed to the observed reverse word-length effect, we recalculated the gaze durations generated in Simulation 2, excluding any trials where the adjectives were skipped. The results of this analysis (also shown in Figure 5) supported our hypothesis; the sizable reverse length effects that were observed in Simulation 2 reverted to the standard length effect, with gaze durations on the nouns being 11 ms and 14 ms longer following long adjectives in the low- and high-frequency conditions, respectively. (The observed effects were 2 ms and 13 ms in the low- and high-frequency conditions, respectively.)

Keeping in mind that these simulations are illustrative in that they did not employ the actual materials in the experiment, we can still draw several conclusions from the modeling. The first is that the inadvertent skipping hypothesis can produce a noticeable reverse length effect – but only about half as big as the observed effect. The second is that the reverse length effect can be accounted for by a combination of increased skipping of the short adjectives and additional processing demands in processing the short adjectives (modeled in the simulation by introducing the assumption that misidentification of the short adjectives sometimes caused difficulty with their subsequent integration).

Given this initial success, we next directed our efforts to accounting for the results of Experiment 2, which employed a gaze-contingent paradigm so that the adjectives were not visible to the participants until their eyes crossed an invisible boundary on the blank space immediately to the left of the adjective. Likewise, the “reader” in the simulation could not start lexical processing of the adjectives until its “eyes” had crossed an invisible boundary to fixate on the adjectives or the blank space immediately before the adjectives. To do this, it was necessary to make one final assumption to handle those rare occasions when the model was processing the word immediately to the left of the adjective while simultaneously programming a saccade towards an adjective, but this saccade fell short of its intended target and landed to the left of the adjective—thereby preventing the display change. (This sequence of events would be a problem because the adjective would remain “invisible” to the model, so that its processing could not “trigger” another eye movement to a new viewing location.) Although there are several ways to avoid this catch-22 situation (e.g., imposing some deadline for initiating progressive saccades; Pollatsek et al., 2006c), these situations occurred very infrequently (in fewer than 6% of the trials reported below) and so we opted to simply eliminate such trials from our analyses. (This solution had little effect on the final simulation results and

was selected because the deadline solution would have required a more specific assumption about the duration of the deadline.)

The first set of simulation results displayed in Figure 6 (Simulation 3) shows the values predicted by the model that was used in Simulation 2, but with a slightly higher probability of misidentifying the short adjectives ($p = .23$ in Simulation 3 vs. $p = .13$ in Simulation 2). This adjustment was necessary to account for the larger number of regressions in Experiment 2 and is consistent with the hypothesis that denying normal parafoveal processing of the adjectives may have made their misidentification more likely, thereby resulting in more frequent problems with the integration of their meanings. As Panels A and B of Figure 6 show, the model did an adequate job accounting for the observed qualitative pattern of results for both the adjectives and nouns. (Observed results are taken from Tables 3 and 4 for Panels A and B, respectively.)

As can be seen in Figure 6, although the model predicted a reverse length effect in both the low- (–45 ms) and high-frequency (–49 ms) conditions, the predicted effect sizes were larger than those that were actually observed in both the low- (–21 ms) and high-frequency (–16 ms) conditions. Moreover, when we recalculated the simulated gaze durations using only those trials where the adjectives were fixated exactly once, the reverse word-length effect disappeared (–3 ms and 0 ms for the low- and high-frequency conditions, respectively). This finding is contrary to what was actually observed in Experiment 2, where the effect sizes in the conditional analyses were –56 ms and –8 ms in the low- and high-frequency conditions, respectively.⁶ This suggests that our assumptions about the loci of the reverse word-length effect are probably not quite correct.

It was at this point that it occurred to us that the misidentification of the short adjectives was likely to also make it more difficult to process the subsequent nouns (Mechanism 4b). To test this idea, we ran a second simulation of Experiment 2 (Simulation 4 in Figure 6) in which we added the assumption that, in those occasions when the integration of an adjective failed, this failure would slow the lexical processing (L_1) of the nouns by some small amount (15 ms for L_1 and 7.5 ms for L_2 ; see Footnote 4). Although the decision to slow lexical processing by this amount is somewhat arbitrary, this delay is quite modest relative to the overall time required to complete lexical processing, and it was sufficient for the model to once again predict a pattern of results that, at least qualitatively, was in close agreement with the observed results. (Although it might seem counter to the spirit of the model to posit that a process in text integration could affect lexical processing, this slow-down could also be viewed as due to a loss of attentional focus similar to Mechanism 4a, but not having any consequences for eye movement control.) Perhaps more important, however, is that this added assumption was sufficient to produce reverse word-length effects (–64 ms and –59 ms in the low- and high-frequency conditions, respectively) that were robust enough to persist in the conditional analyses (–15 ms and –12 ms, respectively). And, although we acknowledge that this assumption made the overall predicted reverse word-length effects too large, it is again worth emphasizing that all of our simulations were completed using a single set of parameter values (see Footnote 5) and that the model's performance can be markedly improved by adjusting the parameters to reduce the amount of saccadic error.⁷ With this disclaimer, we turn to Experiment 3 to see if our assumptions can also account for those results.

Figure 7 shows the results of Simulations 5 and 6, which were two attempts to simulate the results of Experiment 3. Remember that, in this experiment, a boundary paradigm was used to prevent processing of the nouns until after the readers had fixated on, or to the right of, the blank space immediately before the nouns. Our method of implementing the display change

⁶We are skeptical about the size of the reverse word-length effect in the low-frequency condition (–56 ms); this effect is probably spuriously large, and we predict that its true value is probably closer to the size of the effect in the high-frequency condition (–8 ms).

in our simulations of Experiment 3 was exactly the same as in our simulations of Experiment 2, except for the fact that it was parafoveal processing of the nouns—not the adjectives—that was prevented.

In Simulation 5, we used the same assumptions as in Simulation 4 except that (a) the short adjectives were misidentified with a slightly higher probability ($p = .38$) and (b) the resulting difficulty with postlexical integration caused a slightly longer delay (20 ms for L_1 and 10 ms for L_2) in the processing of the nouns. As Panels A and B of Figure 7 indicate, with these assumptions, the model did a reasonably good job predicting the correct (qualitative) patterns of results for both the adjectives and the nouns. (Observed results are taken from Tables 5 and 6 for Panels A and B, respectively.) As Figure 7 also indicates, the model accurately predicts the reverse length effects on the gaze durations: The predicted effects sizes in the low- and high-frequency conditions were -61 and -56 ms, respectively; the corresponding observed effect sizes were -42 and -47 ms. And, although the model failed to predict the small reverse length effects that were observed in the first-fixation durations, these observed effects were not statistically reliable in our analyses. Finally, as can be seen in Figure 7, the elimination of trials in which the adjectives were skipped resulted in robust reverse word-length effects of -41 ms and -38 ms in low- and high-frequency conditions, respectively, which were quite similar to the observed values of -31 ms and -36 ms. Thus, the only major discrepancies between the observed and predicted values were as follows: First, the model tended to underpredict the durations of fixations on the adjectives; and, second, the model tended to underpredict the number of regressions back from the nouns in the long-adjective conditions.

Although there are many possible reasons for the first of these discrepancies, the simplest explanation is that it reflects differences between the participants and/or sentence materials that are not being adequately captured by our model's default parameter settings (see Footnote 5). In other words, the discrepancy could simply be due to the fact that, because the parameter values that were used in our simulation of Experiment 3 were selected to optimize the model's overall goodness-of-fit to the participants used in Schilling et al.'s (1998) experiment, the model is predicting fixations durations that are on average too short. Given the relatively minor nature of the discrepancy, and given that we cannot otherwise justify the assumption that the participants in Experiment 3 reflect a different population of readers than those in Experiments 1 and 2, we opted to forego doing any further simulations to explore the effects of parameter values and to instead focus on resolving the second discrepancy.

As one might guess, there are also many possible reasons for the second of the two discrepancies—that is, explanations for why the model predicted too few regressions in the long-adjective condition. One possibility is simply that the participants in Experiment 3 also occasionally experienced problems associated with misidentifying the long adjectives. To test this idea, we ran a second simulation of Experiment 3 (Simulation 6), using the same assumptions as the previous simulation but increasing the probability of misidentifying the long adjectives from $p = .05$ (which is the “baseline” probability of misidentifying any of the words, except the short adjectives, in all of our simulations) to $p = .18$. As Panel B of Figure 7 shows, this modification produced the desired result: The number of regressions in the long-adjective conditions increased to more closely resemble the level of regressions observed in Experiment 3. Although

⁷We ran another simulation of Experiment 2 using different parameter values ($\Omega_1 = 6$ and $\Omega_2 = 4$) to reduce the amount by which launch-site fixation durations modulate the systematic component of saccadic error. The results of this simulation more closely resembled the observed patterns of results than either Simulation 3 or 4. Most notably, although the predicted reverse word-length effects in gaze durations were much more modest in size (-29 ms and -27 ms in the low- and high-frequency conditions, respectively), they were nonetheless evident in the conditional analyses (-7 ms and -6 ms, respectively). This suggests that, in Simulations 3 and 4, too much of the simulated effect was due to skipping and that not enough was due to problems resulting from the misencoding of the adjectives. Given the relatively minor nature of the discrepancy and given that we cannot otherwise justify the assumption that the participants in Experiment 2 reflect a different population of readers than those in Experiments 1 and 3, we opted to be conservative and report the (slightly) less accurate results of Simulations 3 and 4.

this modification reduced the absolute sizes of the reverse word-length effects (−46 ms and −41 ms in the low- and high-frequency conditions, respectively), these values are actually more similar to those that were actually observed (−42 ms and −47 ms, respectively). Furthermore, the size of the reverse word-length effect for gaze durations in the conditional analyses (i.e., conditional upon the adjectives being fixated exactly once) were still present (−26 ms and −24 ms in the low- and high-frequency conditions, respectively) and were again similar in size to the observed effects (−31 ms and −36 ms, respectively).

Conclusions

In summary, we think the following conclusions can be drawn from the modeling. First, part of the reverse length effect (about 10–15 ms) is attributable to inadvertent skipping of the adjective and this component can be removed (as in Experiment 1) or attenuated (as in Experiments 2 and 3) by analyzing the trials in which the adjective was not skipped. Second, some factor related to differences in the processing of the short versus long adjectives is needed for a full explanation. In the simulations that were reported above, we assumed that, plausibly because of some property related to their denser orthographic neighborhoods, the short adjectives were misidentified more often than the long adjectives and that this misidentification often resulted in problems with the integration of the meanings of the short adjectives. We further assumed (in all of our simulations) that problems with integration resulted in the redirection of both attention and the eyes back to the source of integration difficulty and that (in our simulations of Experiment 2 and 3) problems with integration caused “interference” or slowing in the lexical processing of the nouns. Given that our assumptions were sufficient to account for the results of our experiments, one might ask whether there is some independent basis (i.e., some reason apart from our stated goal of explaining the results of our experiments) for our assumptions.

As indicated earlier, one possible answer to this question is that our misencoding hypothesis also is consistent with the data of Perea and Pollatsek (1998) and Pollatsek et al. (1999), in which the inhibitory effects of both having a higher frequency neighbor and having a larger neighborhood size had quite late effects, including an increase in the number of regressions back to the target word; such effects are most plausibly due to misencoding of the word, which requires a reanalysis. We think that the misencoding hypothesis also naturally explains why the reversed length effect increased in Experiment 3. That is, it seems natural to posit that encoding of the adjective involves excitation of several lexical “candidates” for the word, but that excitation of the “losing” candidates decreases after one of the candidates wins. Thus, denying a preview of the noun would delay encoding of the noun, and thus delay the point in time when reanalysis of the adjective would begin, given that it was incorrectly encoded, and this additional decay of the lexical representation of the correct adjective should increase the reanalysis time needed to instantiate the correct lexical item.

Finally, two questions that we wish to mention are related to the fact that our qualitative arguments have assumed serial processing of words and quantitative tests of effects have employed the E-Z Reader model that assumes such serial processing. The first is whether our conclusions depend strongly on whether that assumption is valid. This issue is difficult to resolve in general, as there could be a radically different conception of reading in which the conclusions would be different. However, if one examines other current models of reading that have parallel processing assumptions, we think that they would be forced to make essentially the same conclusions. In particular, SWIFT (Engbert et al., 2005), which is one such model that has been extensively tested in simulations, has similar assumptions about saccade targeting, and thus we suspect that it would make predictions similar to those of the E-Z Reader model with respect to the effects of mistargeted saccades. Similarly, the SWIFT model (which assumes parallel processing of more than one word at a time) would posit that lengthening the

adjective would, if anything, interfere with parafoveal processing of the noun and hence would have to posit (as we did) some other factor or factors that make the short adjectives more difficult to process in order to explain the reverse length effect.

The second question is whether models that assume parallel processing could indeed explain our data and the reverse length effect in particular. This question is difficult to answer because these models are quite complex and have many more posited processes than E-Z Reader. However, we think that any success that they would have in explaining the data would be independent of, and perhaps in spite of, the parallel processing assumption. That is, in Experiment 3, a preview of the noun was denied so that it was logically impossible to process both adjective and noun while fixating the adjective; yet, a reverse length effect was observed as in the other two experiments. Thus, the effect cannot rely on parallel processing of the two words. (A similar but more complex argument applies to the data of Experiment 2.) More generally, it should be pointed out that, to date, these models have not attempted to model the data of boundary experiments, whereas the E-Z Reader has done so in detail (Pollatsek et al., 2006c). We think that the parallel-processing assumptions of such models would run into trouble when they attempted to do so—especially to be able to predict that reading proceeds quite normally under these conditions with (a) virtually no effect on the word prior to the boundary manipulation and (b) only something like a 30–50 ms increase in processing the word that had no preview.

Acknowledgments

This research was supported by National Institute of Health Grant HD26765 and by Department of Education Institute of Education Sciences Grant R305G020006. The first experiment was completed as part of the requirements of an undergraduate honors thesis by the fourth author. We would like to thank Sarah Brown for her help with the data collection and analyses of Experiment 3. We would like to thank Françoise Vitu for her helpful suggestions on the original submitted manuscript.

References

- Becker W, Jürgens R. Analysis of the saccadic system by means of double step stimuli. *Vision Research* 1979;19:967–983. [PubMed: 532123]
- Brysbaert, M.; Drieghe, D.; Vitu, F. Word skimming: Implications for theories of eye movement control in reading. In: Underwood, G., editor. *Cognitive processes in eye guidance*. Oxford, England: Oxford University Press; 2005. p. 53–78.
- Carpenter RHS. The neural control of looking. *Current Biology* 2000;10:R291–R293. [PubMed: 10801426]
- Coltheart M. The MRC Psycholinguistics Database. *Quarterly Journal of Experimental Psychology* 1981;33A:497–505.
- Davis CJ. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods* 2005;37:65–70. [PubMed: 16097345]
- Drieghe D, Rayner K, Pollatsek A. Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology*. 2007
- Engbert R, Nuthmann A, Richter E, Kliegl R. SWIFT: A dynamical model of saccade generation during reading. *Psychological Review* 2005;112:777–813. [PubMed: 16262468]
- Eriksen CW, Schultz DW. Retinal locus and acuity in visual information processing. *Bulletin of the Psychonomic Society* 1977;9:81–84.
- Feng G. Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research* 2006;7:70–95.
- Francis, W.; Kučera, H. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin; 1982.

- Henderson JM, Ferreira F. Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning Memory and Cognition* 1990;16:417–429.
- Inhoff AW, Eiter BM, Radach R. Time course of linguistic information extraction from consecutive words during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance* 2005;31:979–995. [PubMed: 16262493]
- Johnson RL, Perea M, Rayner K. Transposed letter effects in reading: Evidence from eye movements and parafoveal preview benefit. *Journal of Experimental Psychology: Human Performance and Perception* 2007;33:209–229.
- Jolicoeur P, Ullman S, Mackay MF. Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition* 1983;14:129–140.
- Kennedy A, Pynte J. Parafoveal-on-foveal effects in normal reading. *Vision Research* 2005;45:153–168. [PubMed: 15581917]
- Kennison SM, Clifton C. Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1995;21:68–81.
- Kliegl R, Nuthmann A, Engbert R. Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General* 2006;135:12–35. [PubMed: 16478314]
- McConkie GW, Kerr PW, Reddix MD, Zola D. Eye movement control during reading: I. The location of initial eye fixations in words. *Vision Research* 1988;28:1107–1118. [PubMed: 3257013]
- McDonald SA, Carpenter RHS, Shillcock RC. An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review* 2005;112:814–840. [PubMed: 16262469]
- O'Regan, JK. Eye movements and reading. In: Kowler, E., editor. *Eye movements and their role in visual and cognitive processes*. Amsterdam: Elsevier; 1990. p. 395–453.
- Paap KR, Newsome SL, McDonald JE, Schvaneveldt RW. An activation-verification model for letter and word recognition: The word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance* 1982;10:573–594.
- Perea M, Pollatsek A. The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance* 1998;24:767–779. [PubMed: 9627415]
- Pollatsek A, Perea M, Binder K. The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance* 1999;25:1142–1158. [PubMed: 10464947]
- Pollatsek A, Reichle ED, Rayner K. Attention to one word at a time in reading is still a viable hypothesis: A rejoinder to Inhoff, Radach, and Eiter. *Journal of Experimental Psychology: Human Perception and Performance* 2006a;32:1496–1500. [PubMed: 17154789]
- Pollatsek A, Reichle ED, Rayner K. Serial processing is consistent with the time course of linguistic information extraction from consecutive words during eye fixations in reading: A response to Inhoff, Eiter, and Radach (2005). *Journal of Experimental Psychology: Human Perception and Performance* 2006b;32:1485–1489. [PubMed: 17154787]
- Pollatsek A, Reichle ED, Rayner K. Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology* 2006c;52:1–52. [PubMed: 16289074]
- Posner, MI. *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum; 1978.
- Pynte J, Kennedy A. An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research* 2006;46:3786–3801. [PubMed: 16938333]
- Radach R, Deubel H, Heller D. Attention, saccade programming, and the time of eye-movement control. *Behavioral and Brain Sciences* 2003;26:497–498.
- Rayner K. The perceptual span and peripheral cues in reading. *Cognitive Psychology* 1975;7:65–81.
- Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 1998;124:372–422. [PubMed: 9849112]

- Rayner K, Ashby J, Pollatsek A, Reichle E. 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* 2004;30:720–732. [PubMed: 15301620]
- Rayner K, Duffy SA. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 1986;14:191–201.
- Rayner K, McConkie GW. What guides a reader's eye movements? *Vision Research* 1976;16:829–837. [PubMed: 960610]
- Rayner K, Pollatsek A, Drieghe D, Slattery TJ, Reichle ED. Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert. *Journal of Experimental Psychology: General* 2007;136:520–529. [PubMed: 17696697]
- Rayner K, Pollatsek A, Reichle ED. Eye movements in reading: Models and data. *Behavioral and Brain Sciences* 2003;26:507–518.
- Rayner K, Reichle ED, Stroud MJ, Williams CC, Pollatsek A. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging* 2006;21:448–465. [PubMed: 16953709]
- Rayner K, Sereno SC, Morris RK, Schmauder AR, Clifton C. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes* 1989;4(Special issue):21–49.
- Rayner K, Sereno SC, Raney GE. Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance* 1996;22:1188–1200. [PubMed: 8865619]
- Rayner K, Warren T, Juhasz BJ, Liversedge SP. The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning Memory and Cognition* 2004;30:1290–1301.
- Reichle ED. Models of eye-movement control in reading [Special issue]. *Cognitive Systems Research* 2006;7(1)
- Reichle ED, McConnell K, Warren T. Using E-Z Reader to model effects of higher-level language processing on eye movements during reading. 2007Manuscript submitted for publication
- Reichle ED, Pollatsek A, Fisher DL, Rayner K. Toward a model of eye movement control in reading. *Psychological Review* 1998;105:125–157. [PubMed: 9450374]
- Reichle, E.; Rayner, K.; Pollatsek, A. Modeling the effects of lexical ambiguity on eye movements during reading. In: Van Gompel, RPG.; Fischer, MF.; Murray, WS.; Hill, RL., editors. *Eye movements: A window on mind and brain*. Oxford: Elsevier; 2007. p. 271-292.
- Reichle ED, Rayner K, Pollatsek A. The E-Z Reader model of eye movement control in reading: Comparison to other models. *Behavioral and Brain Sciences* 2003;26:445–476. [PubMed: 15067951]
- Schilling HEH, Rayner K, Chumbley JI. Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition* 1998;26:1270–1281.
- Schroyens W, Vitu F, Brysbaert M, d'Ydewalle G. Eye movement control during reading: Foveal load and parafoveal processing. *Quarterly Journal of Experimental Psychology* 1999;52A:1021–1046. [PubMed: 10605397]
- Sears CR, Campbell CR, Lupker SJ. Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance* 2006;32:1040–1062. [PubMed: 16846296]
- Seidenberg, MS.; Plaut, DC. Progress in understanding word reading: Data fitting versus theory building. In: Andrews, S., editor. *From inkmarks to ideas: Current issues in lexical processing*. New York: Psychology Press; 2006. p. 25-49.
- Shulman GL, Remington RW, McLean JP. Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance* 1979;15:522–526. [PubMed: 528957]
- Slattery TJ, Pollatsek A, Rayner K. The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition* 2007;35:1283–1292.
- Tsal Y. Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance* 1983;9:523–530. [PubMed: 6224890]
- Vitu F, O'Regan JK, Inhoff AW, Topolski R. Mindless reading: Eye movement characteristics are similar in scanning letter strings and reading text. *Perception & Psychophysics* 1995;57:352–364. [PubMed: 7770326]

Yang S. An oculomotor-based model of eye movements in reading: The competition/activation model. *Cognitive Systems Research* 2006;7:56–69.

Appendix

Appendix

Sentences used in all experiments

Target adjectives appear in brackets in the following order: long-high frequency, long-low frequency, short-high frequency, and short-low frequency. The target noun is in bold following the brackets. Participants read each sentence containing one of the adjectives.

1. The section chief learned that an/a [important, ingenious, good, head] **scientist** just quit yesterday.
2. Michelle told me about the [important, ingenious, good, head] **engineer** who filled the job opening.
3. The college announced that the [important, ingenious, good, head] **professor** received a national award.
4. I read that the [important, ingenious, good, head] **scholar** was awarded a large grant.
5. My roommate left a [strange, bizarre, bad, cute] **message** on the refrigerator.
6. We saw a [strange, bizarre, bad, cute] **painting** at the fancy museum in Paris.
7. Barb read the [strange, bizarre, bad, cute] **passage** aloud to her friends.
8. Deb's boyfriend had a [strange, bizarre, bad, cute] **surprise** waiting at home for her.
9. The CEO fired the [foreign, corrupt, fine, bald] **executive** due to budget cut backs.
10. At the flea market the [foreign, corrupt, fine, bald] **merchant** sold his exotic products.
11. The company's [foreign, corrupt, fine, bald] **partner** just signed a new deal.
12. I heard that the [foreign, corrupt, fine, bald] **colonel** retired from the army.
13. The owner of the town's [private, festive, dark, tidy] **restaurant** just sold it to a competitor.
14. The judge's [private, festive, dark, tidy] **chamber** was decorated by his wife.
15. We walked through the [private, festive, dark, tidy] **entrance** to get to the kitchen.
16. I went to my boss's [private, festive, dark, tidy] **apartment** to drop something off.
17. The audience applauded for the [central, teenage, able, lone] **musician** with great enthusiasm.
18. Chores given to the [central, teenage, able, lone] **servant** included all of the cooking and cleaning.
19. Testimony of the [central, teenage, able, lone] **witness** swayed the jury's verdict to not guilty.
20. The final speech from the [central, teenage, able, lone] **candidate** received a standing ovation.
21. Rejection of the [complete, alternate, real, fake] **proposal** stopped all project funding.

22. The insurance company signed a [complete, alternate, real, fake] **contract** with a new client.
23. The guide pointed out a [complete, alternate, real, fake] **version** of the Magna Carta on the tour.
24. The carpenter's [complete, alternate, real, fake] **estimate** was too expensive.
25. The mechanic worked on the [special, gorgeous, red, tan] **vehicle** until the engine was fixed.
26. Every freshman was led into the [special, gorgeous, red, tan] **library** during orientation.
27. Opening night was held at a [special, gorgeous, red, tan] **theater** in the center of Boston.
28. Brook made a [special, gorgeous, red, tan] **creation** in art class for her mother.
29. The judge summoned the [popular, fabulous, thin, rude] **attorney** to the bench.
30. Jacob watched the [popular, fabulous, thin, rude] **reporter** cover the recent terrorist attacks.
31. The mayor's [popular, fabulous, thin, rude] **daughter** arrived late at the party.
32. All players stopped when the [popular, fabulous, thin, rude] **official** called the penalty.
33. The family watched the town's [available, cautious, old, lazy] **detective** examine the scene.
34. The customer demanded to see the [available, cautious, old, lazy] **manager** after being refused service.
35. The company's [available, cautious, old, lazy] **personnel** were called into the director's office.
36. The skating club's [available, cautious, old, lazy] **director** needed help running the show.
37. The ensemble performed [difficult, gigantic, long, dual] **selections** from Bach's sonatas.
38. The knight began his [difficult, gigantic, long, dual] **mission** by crossing into enemy territory.
39. The professor assigned a [difficult, gigantic, long, dual] **reading** on magnetism for next week.
40. The teacher tried to explain the [difficult, gigantic, long, dual] **definition** to the confused class.
41. The newspaper depicted the [serious, colossal, big, vile] **incident** at the local bar last night.
42. Martha witnessed the [serious, colossal, big, vile] **accident** at the race track.
43. Consequences for the teenager's [serious, colossal, big, vile] **mistake** were determined by his parents.
44. The war prisoner's [serious, colossal, big, vile] **struggle** ended victoriously.
45. The crew could never forget the [original, obnoxious, dead, smug] **captain** of the war ship.

46. The unusual talent of the [original, obnoxious, dead, smug] **composer** was not recognized for many years.
47. Jane was forced to complete the [original, obnoxious, dead, smug] **chairman's** improvement project.
48. Riding the train reminded me of the [original, obnoxious, dead, smug] **operator** who used to drive it.

- A. My roommate left a xleocji message on the refrigerator.
*
- B. My roommate left a strange message on the refrigerator.
*
- C. My roommate left a strange message on the refrigerator.
*

Figure 1.

Boundary-change example. The invisible boundary preceded the interword space before the adjective. The asterisks indicate the position of the eyes (a) before the boundary has been crossed and the display has changed and (b, c) after the boundary has been crossed and the display had changed.

Experiment 1

Experiment 2

Experiment 3

Figure 2.

Landing position data from Experiments 1, 2, and 3. Each graph displays the percentage of fixations on the noun with each landing position in character spaces (position 0 is the space before the noun) as a function of the four types of adjectives preceding it (long high-frequency adjective; long low-frequency adjective; short high-frequency adjective; short low-frequency adjective).

Experiment 1

Experiment 2

Experiment 3

Figure 3.

Landing position data from Experiments 1, 2, and 3 conditional on the adjective not being skipped. Each graph displays the percentage of fixations on the noun with each landing position in character spaces (position 0 is the space before the noun) as a function of the four types of adjectives preceding it (long high-frequency adjective; long low-frequency adjective; short high-frequency adjective; short low-frequency adjective).

Figure 4.

Schematic diagram of the E-Z Reader model's basic architecture. The model consists of seven components: (1) a preattentive stage of visual processing (V), (2) an early (L_1) and (3) late (L_2) stage of lexical processing, (4) labile (M_1) and (5) nonlabile (M_2) stages of saccadic programming, (6) an attention shift (A) from one word to the next, and (7) a postlexical stage of meaning integration (I). The processes that are represented by light gray boxes correspond to the standard version of the model, E-Z Reader 9 (Pollatsek et al., 2006c). The processes represented by the darker gray boxes (i.e., A and I) correspond to components that have been added to produce E-Z Reader 10 (Reichle, McConnell, & Warren, 2007).

Figure 5.

Observed (Obs) and simulated (Sim) results of Experiment 1. Panel A shows the mean gaze durations, first-fixation durations, and skipping probabilities for the adjectives, as a function of adjective length (L = long; S = short) and frequency (HF = high frequency; LF = low frequency); observed values are from Table 1. Panel B shows the mean gaze durations, gaze durations conditional upon the adjectives being fixated exactly once, first-fixation durations, and probabilities of making regressions from the nouns, again as a function of adjective length and frequency; observed values are from Table 2. Simulation 1 was completed using the “standard” version of E-Z Reader 10, with no special assumptions about the task being simulated (i.e., we assumed that all words had some probability, $p = .05$, of being subject to

postlexical integration difficulty). Simulation 2 was completed using the added assumption that short adjectives had some probability ($p = .13$) of being misidentified, resulting in postlexical integration difficulty.

Figure 6.

Observed (Obs) and simulated (Sim) results of Experiment 2. Panel A shows the mean gaze durations, first-fixation durations, and skipping probabilities for the adjectives, as a function of adjective length (L = long; S = short) and frequency (HF = high frequency; LF = low frequency); observed values are from Table 3. Panel B shows the mean gaze durations, gaze durations conditional upon the adjectives being fixated exactly once, first-fixation durations, and probabilities of making regressions from the nouns, again as a function of adjective length and frequency; observed values are from Table 4. Simulation 3 was completed using the added assumption that short adjectives had some probability ($p = .23$) of being misidentified, resulting in postlexical integration difficulty. Simulation 4 included the additional assumption that the misidentification of adjectives resulted in interference, slowing lexical processing of the nouns by an additional 15 ms.

Figure 7.

Observed (Obs) and simulated (Sim) results of Experiment 3. Panel A shows the mean gaze durations, first-fixation durations, and skipping probabilities for the adjectives, as a function of adjective length (L = long; S = short) and frequency (HF = high frequency; LF = low frequency); observed values are from Table 5. Panel B shows the mean gaze durations, gaze durations conditional upon the adjectives being fixated exactly once, first-fixation durations, and probabilities of making regressions from the nouns, again as a function of adjective length and frequency; observed values are from Table 6. Simulation 5 was completed using assumptions that short adjectives had some probability ($p = .38$) of being misidentified, resulting in postlexical integration difficulty, and that this resulted in interference, slowing lexical processing of the nouns by an additional 20 ms. Simulation 6 included the additional assumption that this type of processing difficulty also occasionally ($p = .18$) occurred with the long adjectives.

Table 1

Experiment 1: Fixation Measures on the Adjective

Condition	Dependent measure				Total time	Probability of skipping adjective
	Gaze duration	First fixation duration	Single fixation duration			
Long high-frequency adjective	292	262	268		314	.027
Long low-frequency adjective	336	273	277		384	.015
Short high-frequency adjective	273	261	261		302	.268
Short low-frequency adjective	270	261	263		295	.225

Note. Fixation times are in milliseconds.

Table 2

Experiment 1: Fixation Measures on the Noun

Condition	Dependent measure					Probability of skipping noun	Go-past time	Probability of regressing from noun
	Gaze duration	First fixation duration	Single fixation duration	Total time				
Long high-frequency adjective	311 (317)	281	285	338		.009	326	.038
Long low-frequency adjective	317 (314)	280	292	348		.012	353	.081
Short high-frequency adjective	319 (304)	261	272	359		.029	332	.068
Short low-frequency adjective	336 (312)	268	275	370		.015	371	.089

Note. Fixation times are in milliseconds. The values in parentheses are for gaze durations on the noun conditional on there being exactly one fixation on the adjective.

Table 3

Experiment 2: Fixation Measures on the Adjective

Condition	Dependent measure				Total time	Probability of skipping adjective
	Gaze duration	First fixation duration	Single fixation duration			
Long high-frequency adjective	337	289	303		360	.011
Long low-frequency adjective	377	299	321		417	.007
Short high-frequency adjective	297	274	279		323	.141
Short low-frequency adjective	325	298	309		341	.111

Note. Fixation times are in milliseconds.

Table 4

Experiment 2: Fixation Measures on the Noun

Condition	Dependent measure					Probability of skipping noun	Go-past time	Probability of regressing from noun
	Gaze duration	First fixation duration	Single fixation duration	Total time				
Long high-frequency adjective	314 (313)	289	297	330		.029	331	.048
Long low-frequency adjective	323 (312)	290	301	349		.023	359	.063
Short high-frequency adjective	330 (321)	274	287	374		.025	385	.144
Short low-frequency adjective	344 (368)	287	302	383		.027	384	.088

Note. Fixation times are in milliseconds. The values in parentheses are for gaze durations on the noun conditional on there being exactly one fixation on the adjective.

Table 5

Experiment 3: Fixation Measures on the Adjective

Condition	Dependent measure				Total time	Probability of skipping adjective
	Gaze duration	First fixation duration	Single fixation duration			
Long high-frequency adjective	337	288	302		415	.023
Long low-frequency adjective	395	309	331		479	.025
Short high-frequency adjective	299	287	292		364	.213
Short low-frequency adjective	329	304	316		380	.199

Note. Fixation times are in milliseconds.

Table 6

Experiment 3: Fixation Measures on the Noun

Condition	Dependent measure					Probability of skipping noun	Go-past time	Probability of regressing from noun
	Gaze duration	First fixation duration	Single fixation duration	Total time				
Long high-frequency adjective	356 (360)	316	330	407		.005	398	.111
Long low-frequency adjective	382 (378)	332	352	433		.007	425	.109
Short high-frequency adjective	403 (396)	322	351	479		.007	456	.179
Short low-frequency adjective	424 (409)	333	356	488		.003	469	.129

Note. Fixation times are in milliseconds.. The values in parentheses are for gaze durations on the noun conditional on there being exactly one fixation on the adjective.

Table 7

Adjective Length Effects, in ms, for Gaze Duration on the Noun, Gaze Duration on the Noun Conditional on a Single Fixation on the Adjective, and Go-Past Time on the Noun

Experiment	Condition	Gaze duration	Conditional gaze duration	Go-past time
1	Overall	-14 (-25, -2)	8 (-6, 22)	-12 (-29, 5)
1	High-frequency adjective	-9 (-24, 6)	13 (-4., 29)	-6 (-31, 19)
1	Low-frequency adjective	-18 (-39, 2)	2 (-20, 25)	-18 (-45, 9)
2	Overall	-19 (-36, -1)	-32 (-54, -10)	-40 (-66, -14)
2	High-frequency adjective	-16 (-38, 6)	-8 (-32, 17)	-54 (-90, -19)
2	Low-frequency adjective	-21 (-44, 2)	-56 (-89, -23)	-25 (-60, 6)
3	Overall	-45 (-60, -30)	-33 (-49, -18)	-51 (-69, -33)
3	High-frequency adjective	-47 (-70, -26)	-36 (-60, -12)	-58 (-78, -38)
3	Low-frequency adjective	-42 (-63, -22)	-31 (-52, -9)	-44 (-73, -15)

Note. The first number in each cell is the adjective length effect (long adjective–short adjective) on the noun for each dependent measure. The numbers in parentheses are the 95% confidence interval for the length effect.