



The curious case of spillover: Does it tell us much about saccade timing in reading?

Roger W. Remington^{1,2,3} · Jennifer S. Burt¹ · Stefanie I. Becker¹

Published online: 2 July 2018
© The Psychonomic Society, Inc. 2018

Abstract

In completing daily activities, the eyes make a series of saccades by gazing at stimuli in succession. The duration of gaze on each stimulus has been used to infer how the initiation of a saccade is timed relative to the underlying mental processing. In reading, gaze dwells longer on a word that occurs infrequently in English text (low frequency) than on a more frequent word (high frequency), but also on the following word, which is referred to as *spillover*. Accounts of spillover attribute it to mechanisms of lexical access. A low-frequency word n is assumed to delay the onset of cognitive processing of word $n+1$ more than it delays the saccade to $n+1$, leaving more processing to be done on $n+1$ once it is fixated. We tested this assumption by having participants perform a series of speeded lexical decisions on a linear array of letter strings spaced 5° apart, using low- and high-frequency words to vary the lexical difficulty. Lexical decision adds a response selection stage that is absent in reading, which should eliminate differential effects on saccades and cognitive processing. Nonetheless, we found the typical pattern of lengthened gaze duration and spillover for low-frequency words, with effects that were consistent in magnitude with those seen in studies of reading. These data challenge existing accounts of spillover and argue against the idea that reading has a unique interaction with oculomotor control. Instead, the similarity of our gaze patterns to those of reading suggests a common pattern of saccade initiation across tasks.

Keywords Eye movements · Reading · Eye movement mechanisms

Skilled behavior is characterized by smooth, seemingly effortless transitions between items within or across tasks. When we read, look at a picture, type, or sight-read music, we make a series of saccadic eye movements that shift our gaze rapidly from one item or location to the next. The fluidity that develops with skill bespeaks sophisticated coordination between the initiation of a saccade and the underlying cognitive processes occurring while gazing at a stimulus. This coordination has been studied extensively, primarily by examining the duration that the eyes gaze on each item before

making a saccade to the next. A general finding across many tasks is that the eyes gaze longer on items that require more cognitive processing. In reading, for example, gaze duration is longer for words that occur less frequently in written text (low-frequency words) than for those occurring more frequently (high-frequency words), which has been attributed to the greater difficulty in lexical access to low-frequency words. The sensitivity of gaze duration to item difficulty has played a key role in theories positing a direct linkage in which the eyes shift at the completion of some fixed stage of lexical processing (Pollatsek, Reichle, & Rayner, 2006b; Rayner, 2009; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) or when the progress of lexical processing exceeds a threshold (Reilly & Radach, 2003, 2006). The same changes in gaze duration, however, can also be accommodated by opposing theories in which saccade initiation is controlled by an interval-timing mechanism within the saccade system that triggers a saccade when the timer interval has expired (Engbert, Nuthmann, Richter, & Kliegl, 2005; Hooge & Erkelens, 1996, 1998; Nuthmann, Smith, Engbert, & Henderson, 2010; Remington, Lewis, & Wu, 2006).

✉ Roger W. Remington

- ¹ School of Psychology, University of Queensland, Brisbane, Queensland, Australia
- ² Center for Cognitive Sciences, University of Minnesota, Minneapolis, MN, USA
- ³ Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Rayner and Duffy (1986) observed that, in addition to lengthening its own gaze duration, a low-frequency word lengthened gaze on the *following* word, a phenomenon referred to variously as *spillover* (Reichle et al., 2003; Reilly & Radach, 2006) or *lag effects* (Engbert et al., 2005). The size of the spillover effect is typically only about 25% of the gaze duration effect, but the effect is found consistently in reading. Spillover is intriguing because it indexes an interitem interaction that appears related to the linkage between the state of ongoing cognitive processing and saccade initiation. Indeed, Rayner and Duffy speculated that this linkage occurs because a low-frequency word n takes longer to integrate with the following word $n+1$. According to this *integration account*, spillover effects are due to higher-order semantic processes in the construction of meaning from individual words.

Contrary to Rayner and Duffy's (1986) integration account, more recent quantitative models of reading have downplayed the role of semantics, treating spillover largely as an epiphenomenon arising from interaction of low-level mechanisms involved in lexical access. The common theme in these models is that spillover occurs because a low-frequency word lengthens the time required for completion of lexical processing more than it does that for programming the saccade. The details of how this occurs are specific to each model. In the E-Z Reader model, for example, spillover is attributed to reduced parafoveal preview of the upcoming word (see Reichle et al., 2003). Saccade programming begins at the completion of the first stage of lexical access, $L1$, and attention is shifted at the end of the second stage, $L2$. For a high-frequency word, $L2$ is finished well before saccade programming, allowing attention to shift to the next word and begin processing (parafoveal preview). A low-frequency word is presumed to lengthen both $L1$ and $L2$, thus adding an additional delay on the attention shift relative to the saccade. Since preview of word $n+1$ depends on attention having been shifted, lengthening $L2$ means that attention is shifted later, closer to the time of the saccade, reducing parafoveal preview. This leaves more processing to be done during gaze on word $n+1$ than for a high-frequency word n .

In the connectionist Glenmore model (Reilly & Radach, 2003, 2006), saccade programming for word $n+1$ begins once lexical activation for word n has passed a criterion, which occurs prior to the asymptotic level of information accrual that marks the completion of lexical processing. Lexical information accumulates more slowly for a low- than for a high-frequency word, delaying both the saccade to word $n+1$ and the completion of lexical processing. However, the lexical processing of $n+1$ cannot begin until the processing of word n has reached asymptote. The slower rate of evidence accumulation for the low-frequency word means that it takes longer to reach asymptote after saccade programming has begun. Analogous to E-Z Reader, this model imposes an additional delay to the onset of processing that does not affect the timing

of the saccade. This extra delay on the onset of cognitive processing reduces parafoveal preview of word $n+1$ and postpones the start of lexical processing on the newly fixated word, lengthening its fixation duration.

In the SWIFT model, saccades are generated at intervals determined by a random walk, with no direct coupling to ongoing lexical processing. However, SWIFT incorporates an inhibitory mechanism that will delay the saccade when the interval is insufficient to complete cognitive processing. Because feedback from lexical processing is slow relative to that from saccade programming, the inhibitory signal can arrive during the nonlabile stage of saccade programming, too late to cancel the saccade. To deal with this, SWIFT includes a time delay associated with the inhibitory signal that extends into the fixation on word $n+1$, blocking its lexical processing in order to allow for possible unfinished lexical access for word n , lengthening the gaze duration on word $n+1$.

According to these accounts, spillover is not a general saccade phenomenon, but is instead derived from mechanisms of lexical access in reading. Rayner and Duffy's (1986) integration account similarly links spillover to reading through the construction of meaning from individual words. These accounts are consistent with the broader claim that "how the cognitive system interacts with the oculomotor system, differs as a function of the task" (Rayner, 2009, p. 1459). Though vague as to the exact nature of the interaction, the claim suggests a special connection between lexical access and oculomotor control that is absent in other tasks. If the accounts above are correct, spillover may be a phenomenon unique to reading. In support of this, a visual search study by Williams and Pollatsek (2007) revealed no spillover when participants searched for a circle among Landolt Cs arrayed in linear clusters, simulating the linear structure of text in reading. Gaze durations were elevated for more difficult clusters (small gap sizes), but this had no effect on the gaze duration on the next cluster.

Spillover-like effects, however, have been found in nonlexical tasks. Hooge, Vlaskamp, and Over (2007) had participants make speeded responses to Cs embedded in a linear list of Os, varying the gap size in order to vary difficulty. They created contexts biased toward easy or difficulty discriminations and found that fixation durations on an easy stimulus were longer when they were embedded in a context of primarily hard discriminations. Remington, Wu, and Pashler (2011) found contextual effects very similar to those of Hooge et al. by using stimulus-response compatibility to vary item difficulty. In a version of the overlapping-tasks paradigm (Pashler, 1984; Remington et al., 2006; Wu & Remington, 2004; Wu, Remington, & Pashler, 2004), participants made speeded manual button presses to indicate the number of filled cells in each of a series of five 2×2 matrices arrayed linearly (horizontally) across the screen. The digits filling each cell could be either compatible with the number of filled cells (e.g., 2, 2 = respond

“2”) or incompatible (e.g., 3, 3 = respond “2”). Gaze durations were elevated for a compatible matrix in blocks of primarily incompatible matrices. Remington et al. (2011) argued that spillover could not be due to restricted parafoveal preview (see the E-Z Reader model, above): The matrices were separated by 5°, so the effects of crowding at that retinal eccentricity would eliminate any useful parafoveal preview of the relevant visual information (Bouma, 1970; Pelli & Tillman, 2008). The researchers also argued against the kinds of delays caused by unfinished processing, as in the Glenmore and SWIFT models, citing, in part, the finding that spillover to matrix $n+1$ was seen even on trials on which the response to matrix n had been made before the saccade, so that no obvious unfinished processing was left. The overall eye fixation patterns with respect to item difficulty in Remington et al. (2011) were very similar to those from lexical manipulations in studies of reading, despite differences in the stimuli and task. Because the stimuli used by Remington et al. (2011) did not involve meaningful words, we can also exclude higher-order semantic effects as being responsible for the observed spillover. Remington et al. (2011) raised the possibility that spillover could be a more general adaptation of the saccade system to increased demands, similar to the adaptive timers for saccade initiation in SWIFT and CRISP (Nuthmann et al., 2010).

The tasks and stimuli in Remington et al. (2011) and Hooge et al. (2007) admittedly differed significantly from the reading tasks typically used to assess spillover effects—most markedly, perhaps, in that no lexical access was involved. This is an important qualification, because SWIFT, Glenmore, and E-Z Reader all specifically attribute spillover to interactions with the lexical processing of adjacent words. Moreover, in reading, spillover occurs as a function of the frequency of the previous word, not just in a context of low-frequency words. As such, it is difficult to use these studies to directly address current accounts of spillover. In the present study, we sought to test the specific lexical mechanisms postulated by E-Z Reader, Glenmore, and SWIFT by creating conditions that would require lexical access, but in which, according to the models, a low-frequency word should not produce the kind of asymmetric delays to saccade onset time and cognitive processing that are presumed to cause spillover. In a lexical decision task, participants made a series of speeded button presses to indicate whether each of a series of 5 four-letter strings arrayed linearly across the screen was a word or a nonword. The stimuli were responded to from left to right. The centers of the letter strings were spaced approximately 5° apart, far enough that each letter string was fixated in turn and within the critical band for crowding, which effectively eliminated parafoveal preview (see Reichle et al., 2003). Lexical difficulty was manipulated by varying normative word frequency. Previous studies using the sequence paradigm have shown no evidence that unfinished processing on item n interferes with processing while one is fixating item $n+1$, which has

been taken as evidence that the saccade from item n to $n+1$ is not made until response selection is completed (Remington et al., 2011). Therefore, as we describe below, the response demands of the lexical decision should cause the saccade to be made, and processing of word $n+1$ to begin, at or near the end of response selection, well after the completion of lexical processing. Response selection is itself conditioned on the completion of lexical processing, so that lexical access per se should play little if any role in the onset of the saccade or in the lexical processing for word $n+1$. Thus, we should see no spillover.

Logic of the present experiment A common and key assumption of E-Z Reader, Glenmore, and SWIFT is that spillover is a temporally localized effect that occurs at the lexical access stage of processing, in which the difficulty of lexical access that is associated with a low-frequency word delays the onset of the saccade to word $n+1$, but also produces a further delay in the onset of lexical processing on word $n+1$. In reading, the saccade to word $n+1$ and the onset of its lexical processing are assumed to immediately follow the lexical processing of word n . If spillover is directly tied to mechanisms of lexical access, then it should be possible to eliminate (or greatly reduce) it by delaying the execution of the saccade, and so the ensuing processing of word $n+1$, until the completion of a subsequent stage of processing. For example, in the lexical decision task, the completion of lexical access leads to a word/nonword decision that is then mapped onto a key press response. A low-frequency word would be expected to prolong lexical access, slowing both the decision and response selection stages. If the saccade to the following item is initiated near the end of response selection (Remington et al., 2011; Williams & Pollatsek, 2007), then it becomes difficult to attribute spillover to the processes involved in gating lexical access, as is proposed by E-Z Reader, Glenmore, or SWIFT, since the lexical access must be complete in order for the word/nonword judgment to be made. The result should be the elimination or dramatic reduction of the spillover effect, as compared to reading. Because the task required individual responses to each word/nonword, no integration of the word meanings was required. Hence, if spillover effects were still be observed in the current task, we could conclude that they were not due to higher-order semantic integration effects or to task-specific interactions between lexical processing and the oculomotor system.

Method

Participants made speeded lexical decisions to 5 four-letter strings arrayed horizontally across the computer screen. Adjacent word pairs in Positions 1–2, 2–3, 3–4, and 4–5 had a factorial variation of word frequency, high (H) versus low (L), with the frequency combinations HH, HL, LH, and LL.

All strings were four letters long, and the nonwords were orthographically acceptable and pronounceable in English.

Participants

Sixteen introductory psychology students participated for course credit. All were native speakers of English.

Apparatus

An Intel Duo 2 CPU 2.4-GHz computer running the Presentation software (Neurobehavioral Systems, Berkeley, CA) displayed the stimuli on a 17-in. FP92E color monitor, at a resolution of $1,280 \times 1,024$ pixels and a refresh rate of 75 Hz. For eye tracking, a video-based infrared eye tracker (EyeLink 1000, SR Research, Ontario, Canada) was used (500 Hz), which stabilized the participants' head with a chin rest and forehead support at a distance of 62 cm from the screen.

Experimental design and stimulus materials

The targets were 192 high-frequency and 192 low-frequency four-letter words selected from the British National Corpus (Kilgarriff, 1995). The minimum frequency in the corpus for high-frequency words was 75 per million, with a mean of 277. For the low-frequency words, the frequency range was 1.5 to 12 per million (mean = 5.2). The orthographic neighborhood sizes—that is, the number of words differing from the word by one letter, as in *make* for *male*—were matched over frequency sets in order to ensure that frequency was not confounded with orthographic typicality. The mean number of neighbors was 7.4, and the mean summed frequency of the neighbors was 697 per million. An additional 256 medium-frequency four-letter filler words were selected in the frequency range of 15 to 49 per million. In addition, 730 four-letter pronounceable nonwords with neighbor statistics similar to those of the high- and low-frequency sets were selected from the English Lexicon Project (Balota et al., 2007).

Assignment of the high- and low-frequency words to word pairs and to position in the sequence was counterbalanced over eight lists. Each word had the same position within a pair over the lists, occurring four times with a word of the same frequency type and four times with a word of the other frequency type. Over the four lists within each frequency condition, the pair was rotated through the adjacent positions in the sequence. For example, the high-frequency word *line* occurred as the second pair member after the low-frequency word *cult* and the high-frequency word *need*, and over the eight lists, the pairs *cult line* and *need line* occurred in Positions 1 and 2, 2 and 3, 3 and 4, and 4 and 5. The nonwords and medium-frequency filler words were distributed through the other positions in the sequence, with two different word/

nonword sequences being used equally often for each pair position. An additional 48 trials were composed of filler words and nonwords in a variety of sequences containing up to four consecutive items of the same lexical status. These filler trials equated the total numbers of words and nonwords and ensured that the lexical status of the last one or two items in the test sequences was not predictable from the lexical status of the first three or four items. Each list had a total of 240 test trials, composed of 600 words and 600 nonwords. The trial sequence was randomized. The remaining nonwords and filler words were used to construct ten practice trials.

The stimulus displays consisted of a black fixation cross (0.3×0.3 cm) to the left of the display (3.65 cm from the left monitor frame) and five black words or nonwords (Arial 9-point) that were presented equidistantly on an imaginary horizontal line through the center. The four-letter strings measured approximately 1×0.3 cm (single letter: 0.3×0.3 cm), and the center-to-center distance between the nearest items was 6.2 cm. Prior to the start of the trial, the letters were masked with four black Xs (XXXX; 0.85×0.35 cm), and all stimuli were presented against a light gray background (RGB: 150, 150, 150).

Procedure

Prior to the experiment, eye movement recording for each participant was calibrated with a 9-point randomized calibration procedure. We tested only participants whose eye movements could be successfully calibrated. Participants were instructed to respond sequentially to the five letter strings in the display from left to right, pressing the arrow-down key to report a word and the arrow-right key to respond to a nonword. Participants received no instructions regarding their eye movements, except with regard to a fixation control that was implemented with the mask display: participants were instructed that they had to fixate on the fixation cross on the left side of the screen when the mask display appeared to start the trial. When participants had been fixating within 1.5 cm of the center of the cross for 500 ms (within a time window of 3,000 ms), the masks disappeared and revealed the five letter strings. When the eye tracker failed to detect a 500-ms fixation on the cross, participants were calibrated anew, and the next trial started again with the mask display. The letter display remained on screen until five responses had been recorded. Immediately afterward, a feedback display was presented showing the word “Correct!” when all five responses had been correct, or “Wrong!” when one or more responses had been incorrect. The feedback display was presented for 500 ms, following by a blank gray screen for 250 ms, and the next trial again started with the presentation of the mask display.

Results

All effects were tested at $\alpha = .05$. Item analyses were not conducted, given the facts that a large set of words was used, words were cycled through conditions, and the frequency effect examined here is well established for response latencies in the lexical decision task.

In keeping with previous studies of the effects of word frequency and spillover (e.g., Rayner & Duffy, 1986), we focused our analysis on *gaze*, or the sum of all contiguous fixation durations on an item before making a saccade to the next. Gaze duration reflects the decision on when to transition to the next item and gives us an opportunity to assess any processing conflicts that would generate spillover from word n onto word $n+1$. With this in mind, we also analyzed *release-hand span* (RHS), defined as the interval from the time that gaze has shifted to word $n+1$ to the manual response to word n . The RHS interval overlaps unfinished processing on word n with the lexical and cognitive processing of word $n+1$, making it sensitive to resource conflicts that could underlie lengthened $n+1$ gaze durations.

Before proceeding to an analysis of spillover in word pairs, it was important to confirm that our overall effects were characteristic of those from previous studies in this paradigm. Those studies focused primarily on the effects of item difficulty on the *interresponse interval* (IRI), defined as the time between successive manual responses, and *eye-hand span* (EHS), defined as the total time from first fixating a word until its response (Pashler, 1994; Remington et al., 2011). For comparison with previous sequence studies, Fig. 1 plots IRI, gaze, and EHS across items for both words and nonwords. An analysis of variance (ANOVA) with the factors lexical status (word, nonword) and position in sequence (1–5) showed that the qualitative pattern of data was representative of the data from other studies using this paradigm with nonlexical difficulty manipulations (Remington et al., 2006; Remington et al., 2011; Wu et al., 2004). IRI and EHS were elevated for the first item, converging to a steady state by Item 2. This elevation has been attributed to a strategy of deferring the response to item n until completion of some amount of the processing on item $n+1$ (Remington et al., 2011). For gaze, an ANOVA revealed significant main effects of lexical status (words faster than nonwords), $F(1, 15) = 34.46$, $MSE = 2,466$, $\eta_p^2 = .70$; position in sequence (1–5), $F(4, 60) = 4.82$, $MSE = 7,471$, $\eta_p^2 = .24$; and their interaction, $F(4, 60) = 4.43$, $MSE = 1,304$, $\eta_p^2 = .22$. For IRI, an ANOVA also revealed main effects of lexical status $F(1, 15) = 18.73$, $MSE = 3,329$, $\eta_p^2 = .56$, and position, $F(4, 60) = 164.38$, $MSE = 23,921$, $\eta_p^2 = .92$, but no interaction. Follow-up comparisons revealed that the IRI was faster at Position 5 than at Position 4, $F(1, 15) = 7.89$, an end-of-sequence effect, since without an additional item following, the manual response can be made directly after the completion of response selection on Item 5. For EHS, an ANOVA

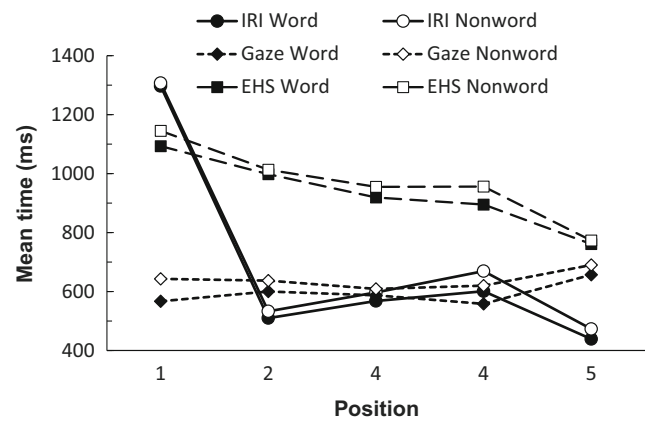


Fig. 1 Interresponse interval (IRI), gaze, and eye-hand span (EHS) are plotted as a function of position in the sequence. Open symbols refer to nonwords, filled symbols to words

revealed significant effects of lexical status, $F(1, 15) = 19.74$, $MSE = 2,516$, $\eta_p^2 = .57$, and position, $F(4, 60) = 38.36$, $MSE = 13,724$, $\eta_p^2 = .72$, as well as their interaction, $F(4, 60) = 4.54$, $MSE = 843$, $\eta_p^2 = .23$.

To assess spillover, we conducted separate analyses of gaze and RHS within word pairs as a function of the frequency of the words in the first and second positions. To avoid complications related to the first and last items, we focused on the frequency effects only for the steady-state Positions 2–3 and 3–4. Prior to the analysis we excluded fixations and responses on filler trials or items for which the lexical decision was incorrect, as well as for trials on which there was an eyetracking problem or a regressive saccade (1.8% of trials). There were occasional outliers in the gaze data, so we also excluded times in excess of 1,800 ms or in excess of three *SDs* above or below a participant’s mean for each item type—that is, for words versus nonwords in the sequence analyses, and for high- versus low-frequency words in the word pair analyses. This resulted in the loss of 26% of the trials in the pair data, with an additional loss of 2.3% of the gaze data as a result of culling extreme times. For the RHS data, we excluded times less than 11 ms and in excess of 1,000 ms (1.6% of trials)

As we noted, the effects of word frequency were assessed on the data for word pairs in Positions 2–3 and 3–4. Gaze and RHS were collapsed over list positions but were recorded separately for position within pairs. A within-subjects ANOVA on lexical decision error rates, shown in Table 1, with the factors pair member (Word 1 vs. Word 2) and word frequency (high vs. low) revealed only a main effect of word frequency, with error rates higher for low- than for high-frequency words, $F(1, 15) = 19.08$, $MSE = 37$, $\eta_p^2 = .56$.

The data for gaze are plotted in Fig. 2. A within-subjects ANOVA with the factors pair member (Word 1 vs. Word 2), Word 1 frequency (high vs. low), and Word 2 frequency (high vs. low) revealed a main effect of Word 1 frequency, $F(1, 15) = 28.70$, $MSE = 2,310$, $\eta_p^2 = .66$, and no interaction of Word 1

Table 1 Mean error percentages for high- and low-frequency words as a function of position in the word pair (first vs. second)

Word Frequency	Position in Pair	
	First	Second
High	2.54	2.67
Low	10.22	8.33

frequency with pair member. A low-frequency Word 1 elevated gaze for Word 1 by 58 ms, with a spillover effect on Word 2 of 33 ms. The main effect of Word 2 frequency was also significant, $F(1, 15) = 8.96$, $MSE = 8,650$, $\eta_p^2 = .37$, as was the Word 2 Frequency \times Pair Member interaction, $F(1, 15) = 13.73$, $MSE = 48.25$, $\eta_p^2 = .48$. A low-frequency Word 2 elevated gaze by 95 ms, as compared to a nonsignificant 4-ms effect on Word 1. The latter result is as expected, because the spatial separation of words prevented preview of the next word. No other effects were significant.

The mean RHS was 325 ms, indicating that participants moved their eyes to Word 2 on average 325 ms before making the response to Word 1. A Pair Member \times Word 1 Frequency \times Word 2 Frequency ANOVA on RHS revealed a main effect of pair member, with Word 2 RHS being shorter than Word 1 RHS, $F(1, 15) = 13.72$, $MSE = 7,599$, $\eta_p^2 = .48$. As in the gaze data, the main effect of Word 1 frequency was significant for RHS for Word 1 (32 ms), $F(1, 15) = 17.75$, $MSE = 937$, $\eta_p^2 = .54$, with a spillover effect on Word 2 (13 ms), but no significant Pair Member \times Word 1 Frequency interaction. A follow-up test showed the spillover effect to be significant, $F(1, 16) = 5.28$. The only other significant effect on RHS was a Pair Member \times Word 2 Frequency interaction, $F(1, 15) = 5.33$,

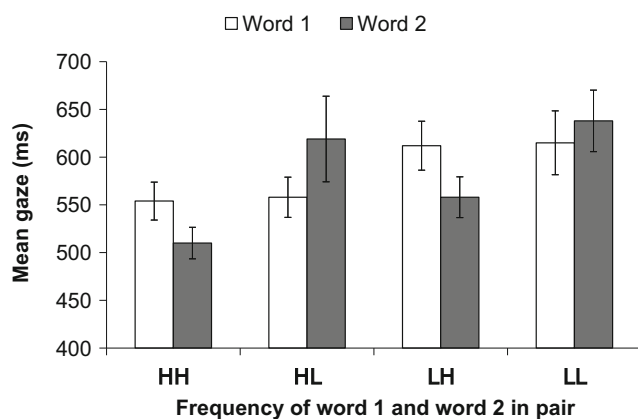


Fig. 2 Mean gaze durations on Word 1 and Word 2 as a function of the frequency of each word in the pair. The first letter in each pair refers to the frequency of Word 1, the second to Word 2. Thus, LH = low-frequency Word 1 paired with a high-frequency Word 2. Error bars depict one standard error above and below the mean

$MSE = 749$, $\eta_p^2 = .02$, reflecting a Word 2 frequency effect on Word 2 (15 ms) and a small reverse effect on Word 1 (–7 ms). Simple-effects analysis revealed that neither effect was significant.

Conclusions

The clear effects of lexical category (word, nonword) on response timing and errors confirm the typical findings in the lexical decision task (Forbach, Stanners, & Hochhaus, 1974). The effects of lexical difficulty (word frequency) yielded generally additive effects on IRI, gaze, and EHS across items, in keeping with the difficulty manipulations of nonlexical items in previous studies using the overlapping-tasks paradigm (Remington et al., 2011; Wu et al., 2004; Wu, Remington, & Pashler, 2007). The word frequency effect averaged across Words 1 and 2 was 76 ms; the mean gaze duration on high-frequency words was 545 ms, as compared with 621 ms for low-frequency words.

Significant spillover was evident in the increase in Word 2 gaze with a low-frequency Word 1, consistent with previous findings in text reading (Rayner & Duffy, 1986), as well as in difficult visual search (Hooge et al., 2007) and the “number Stroop” (Remington et al., 2011). These data patterns suggest that saccades were made after lexical processing was completed. The mean gaze duration averaged across word frequencies and positions was 583 ms, substantially longer than the 200–250 ms typically observed in reading (Pollatsek, Reichle, & Rayner, 2006a, b; Rayner, Slowiaczek, Clifton, & Bertera, 1983). This supports our contention that the interplay between lexical processing of words n and $n+1$ that is proposed to underlie spillover in E-Z Reader, SWIFT, and Glenmore cannot readily account for the spillover seen here, since at least 250 ms elapsed from the completion of lexical processing on word n to the initiation of lexical processing on word $n+1$. Neither the unfinished lexical processing presumed in Glenmore nor SWIFT’s compensatory delay of word $n+1$ is needed to modulate $n+1$ processing after such a delay. Given that these purported delays produce only about a 30-ms spillover effect, they could not have produced the effects we see after a prolonged response selection stage. We have previously noted that the results of Remington et al. (2011) cast doubt on the role of parafoveal preview in spillover, given that spillover was observed with a 5° separation of items, which is well within the critical band for crowding (Bouma, 1970; Pelli & Tillman, 2008), even for cases in which the stimulus is the intended target of an impending saccade (Harrison, Mattingley, & Remington, 2013). The present study also showed spillover with the same separation, producing further evidence against the parafoveal preview account of spillover in E-Z Reader.

The E-Z Reader account rests on a delay in shifting spatial attention to word $n+1$. It is possible that a low-frequency word

might affect attention in other ways that disrupt the saccade timing by altering the landing position on word $n+1$, leading to delays in $n+1$ processing.¹ To test this, we subtracted the x -coordinate, in screen pixels, of the center of each word from the x -coordinate of the first fixation on the word. A negative deviation score would reflect a bias toward the front of the word, a positive score a deviation toward the end. The mean deviation for high-frequency words was 5.32 pixels, as compared to 3.24 pixels for low-frequency words. A paired t test showed no significant effect, $t(15) = 0.50$, $p > .6$. Thus, we found no indication that that spillover effects in the present study were due to systematic imprecision in saccades or shifts in landing position.

Proponents of SWIFT, Glenmore, or E-Z Reader could argue that their mechanisms do indeed provide an accurate account of spillover in reading and that what we have observed is a different phenomenon. That is, the addition of a response selection stage may have masked the direct lexical-processing interactions between adjacent words that goes on in reading, while introducing a different source of spillover. We have no direct evidence against such a claim. Two considerations, however, weigh against the argument for separate sources. First, our sequence task retains key features of reading, including the linear structure of text and the necessity to complete lexical access. Second, lexical access has shown no evidence of being altered by the added response requirement of our lexical decision task as compared with reading. Our 76-ms frequency effect is within the normal range of frequency effects in reading, as is our 33-ms spillover effect (see, e.g., Rayner & Duffy, 1986; Reichle et al., 2003). Given that all three models localize spillover to changes in lexical processing, the observation that lexical access shows no sign of being altered suggests that our task provides a fair test of the underlying assumptions of all three models. To maintain that spillover in reading is fundamentally distinct from what we have observed would require that two quite distinct sources produce surprisingly similar quantitative effects.

The alternative explanation is that the spillover seen here and in reading is a characteristic of the dynamics of saccadic eye movements more generally. Spillover is the result of a saccadic system that uses past context to estimate the timing of the execution of a saccade in order to most efficiently process a list of items in rapid succession, be they words, matrices, speeded perceptual judgments, or scenes. The utility of using estimates from recent history becomes clear when one considers the tight time constraints on saccade execution in reading. Pollatsek, Reichle, and Rayner (2006a) reported that the latency to initiate a saccade is 150–175 ms (Rayner et al., 1983), with saccade durations in the vicinity of 25 ms. Given evidence that typical word fixations are 200–250 ms in length, and the plausible assumption that word

identification requires at least 150 ms (Pollatsek et al., 2006a), little time would be available for a reader to make online, real-time decisions about the timing of the next eye movement on the basis of the progress identifying the currently fixated word. There is precedent for adaptive saccade control in mixed models of saccade generation in reading (SWIFT) and scene analysis (e.g., CRISP; Engbert et al., 2005; Nuthmann et al., 2010). The ICAT model of Trukenbrod and Engbert (2014), for example, attempts to provide a model of saccade dynamics that encompasses a range of tasks, including scene analysis, visual search, and reading. It incorporates an adaptive timer for saccade generation whose interval is determined in part by past history. Although ICAT does not explicitly address spillover, its adaptive, history-sensitive timer would provide the mechanism needed to account for spillover across a broad range of tasks.

It is not necessary, however, to invoke a dedicated interval saccade timer, since other adaptive mechanisms could also produce spillover. As in the Glenmore model, the saccade can be viewed as a response made when evidence accumulates to a threshold. Such trial-by-trial adjustments of threshold in order to avoid errors and minimize response times are a feature of evidence accumulation models. The adaptive account we propose attributes spillover to continuous tuning of the cognitive system, whose purpose is to increase efficiency and decrease cognitive demands. This is in clear contrast to existing models, in which spillover emerges from the complex interaction of specialized mechanisms in the interplay of lexical access and saccade programming.

Our results challenge the accounts of spillover in E-Z Reader, SWIFT, and Glenmore, but they are neutral with respect to the more central claims of each. In fact, our conclusion that spillover is a more general adaptive adjustment unrelated to reading would remove the need for those models to provide a specialized account of it. In this regard, it is worth noting that spillover does not play a central role in any of their accounts of eye movements in reading—seeming, instead, more like an odd phenomenon that must be accounted for and that each model creatively finds ways for its mechanisms to produce.

One could object that we have not entirely ruled out a localized effect that spreads from lexical access all the way through to response selection. It could be argued that the effects of word frequency are not confined to lexical access but also affect the decision and response selection stages (e.g., Balota & Chumbley, 1984; McCann, Remington, & Van Selst, 2000). Indeed, the processes of lexical access, decision-making, and response selection may not be separate at all, but instead reflect one continuous accumulation of evidence leading to a response. Nonetheless, this accumulation must take into account the logical dependence of both decision and response selection on evidence accruing earlier. Even if our detailed account were considered overly simplistic, our more important points would remain valid: that (1) saccade

¹ We thank Reinhold Kliegl for this suggestion.

dynamics in reading reflect a general mechanism for saccade initiation across tasks, (2) saccade initiation reflects a readiness to process new input, and (3) spillover is not a phenomenon specific to reading.

Author note This work was supported by Australian Research Council Discovery Grant DP130101001 to R.W.R., and by Australian Research Council Discovery Grant DP170102559 to S.I.B.

References

- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357. <https://doi.org/10.1037/0096-1523.10.3.340>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., .. Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. <https://doi.org/10.3758/BF03193014>
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813. <https://doi.org/10.1037/0033-295x.112.4.777>
- Forbach, G.B., Stanners, R.F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*, 337. <https://doi.org/10.3758/BF03209005>
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, *33*, 2927–2933. <https://doi.org/10.1523/JNEUROSCI.4172-12.2013>
- Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception & Psychophysics*, *58*, 969–976. <https://doi.org/10.3758/BF03206825>
- Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, *38*, 1295–1302.
- Hooge, I. T. C., Vlaskamp, B. N. S., & Over, E. A. B. (2007). Chapter 27—Saccadic search: On the duration of a fixation. In R. P. G. V. Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements* (pp. 581–595). Amsterdam, The Netherlands: Elsevier. <https://doi.org/10.1016/B978-008044980-7/50029-X>
- Kilgarriff, A. (1995). BNC database and word frequency lists (website). Retrieved from <https://www.kilgarriff.co.uk/bnc-readme.html>
- McCann, R. S., Remington, R. W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1352–1370. <https://doi.org/10.1037/0096-1523.26.4.1352>
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*, 382–405. <https://doi.org/10.1037/A0018924>
- Pashler, H. (1984). Processing stages in overlapping tasks—Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358–377. <https://doi.org/10.1037/0096-1523.10.3.358>
- Pashler, H. (1994). Overlapping mental operations in serial performance with preview. *Quarterly Journal of Experimental Psychology*, *47A*, 161–191. <https://doi.org/10.1080/14640749408401148>
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*, 1129–1135. [10.1038/nn.2187](https://doi.org/10.1038/nn.2187)
- Pollatsek, A., Reichle, E. D., & Rayner, K. (2006a). 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*, *32*, 1485–1489.
- Pollatsek, A., Reichle, E. D., & Rayner, K. (2006b). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56. <https://doi.org/10.1016/j.cogpsych.2005.06.001>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201. <https://doi.org/10.3758/BF03197692>
- Rayner, K., Slowiaczek, M. L., Clifton, C., Jr., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 912–922. <https://doi.org/10.1037/0096-1523.9.6.912>
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157. <https://doi.org/10.1037/0033-295X.105.1.125>
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–476, disc. 477–526. <https://doi.org/10.1017/S0140525X03000104>
- Reilly, R. G., & Radach, R. (2003). Foundations of an interactive activation model of eye movement control in reading. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognition and applied aspects of eye movement research* (pp. 429–455). Amsterdam, The Netherlands: Elsevier.
- Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*, 34–55. <https://doi.org/10.1016/j.cogsys.2005.07.006>
- Remington, R. W., Lewis, R., & Wu, S.-C. (2006, April). *Scheduling mental operations in a multiple-response sequence: Modeling the effects of a strategy to minimize variance in the timing of saccades*. Paper presented at the 7th International Conference on Cognitive Modeling, Trieste, Italy.
- Remington, R. W., Wu, S.-C., & Pashler, H. (2011). What determines saccade timing in sequences of coordinated eye and hand movements? *Psychonomic Bulletin & Review*, *18*, 538–543. <https://doi.org/10.3758/s13423-011-0066-0>
- Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*, 907–934. <https://doi.org/10.3758/s13423-013-0575-0>
- Williams, C. C., & Pollatsek, A. (2007). Searching for an O in an array of Cs: Eye movements track moment-to-moment processing in visual search. *Perception & Psychophysics*, *69*, 372–381.
- Wu, S.-C., & Remington, R. W. (2004). Coordination of component mental operations in a multiple-response task. In A.T. Duchowski & R. Vertegaal (Eds.), *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications* (pp. 63–70). New York, NY: ACM Press.
- Wu, S.-C., Remington, R. W., & Pashler, H. (2004). Coordination of component mental operations in sequences of discrete responses. In K. Forbus (Ed.), *Proceeding of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 1476–1481). Mahwah, NJ: Erlbaum.
- Wu, S.-C., Remington, R. W., & Pashler, H. (2007, July). *Scheduling of eye movements and manual responses in performing a sequence of choice responses: Empirical data and model*. Paper presented at the 8th International Conference on Cognitive Modeling, Ann Arbor, MI.