BEHAVIORAL AND BRAIN SCIENCES (2003) **26,** 445–526
Printed in the United States of America

The E-Z Reader model of eye-movement control in reading: Comparisons to other models

Erik D. Reichle
Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260
reichle@pitt.edu

Keith Rayner
Department of Psychology, University of Massachusetts, Amherst, Amherst, MA 01003
rayner@psych.umass.edu

Alexander Pollatsek
Department of Psychology, University of Massachusetts, Amherst, Amherst, MA 01003
pollatsek@psych.umass.edu

Abstract: The E-Z Reader model (Reichle et al. 1998; 1999) provides a theoretical framework for understanding how word identification, visual processing, attention, and oculomotor control jointly determine when and where the eyes move during reading. In this article, we first review what is known about eye movements during reading. Then we provide an updated version of the model (E-Z Reader 7) and describe how it accounts for basic findings about eye movement control in reading. We then review several alternative models of eye movement control in reading, discussing both their core assumptions and their theoretical scope. On the basis of this discussion, we conclude that E-Z Reader provides the most comprehensive account of eye movement control during reading. Finally, we provide a brief overview of what is known about the neural systems that support the various components of reading, and suggest how the cognitive constructs of our model might map onto this neural architecture.

Key words: attention; eye-movement control; E-Z Reader; fixations; lexical access; models; reading; regressions; saccades

1. Introduction

Reading is a complex skill that involves the orchestration of many different stages of information processing. As the eyes move across the printed page, the visual features of the text are converted into orthographic and phonological patterns, which are then used to guide further language processing so that the content of the text can be understood. In this target article, we will compare different models that try to account for how eye movements are controlled in reading. We will not review all of the models that have been proposed to explain various aspects of reading. Instead, we will discuss only those models that have attempted to explain the interface between vision and low-level aspects of language processing; that is, models that specify some combination of the following components of reading: eye-movement control, visuospatial attention, and/or the visual processing of words.[1] Not surprisingly, we will argue that the model that we implemented, E-Z Reader[2] (Reichle et al. 1998; 1999), does a better job of accounting for a wide range of data than does its competitors. However, we will also point out some shortcomings of the model.

The remainder of this article will be organized into five major sections. First, in section 2, we will briefly review some important findings regarding eye movements in reading; within this section we will describe some findings that we believe a model of eye-movement control should be able to accommodate. Second, in section 3, we will provide an overview of the E-Z Reader model, including an updating of the model (E-Z Reader 7). Third, in section 4, we will provide an overview of other models of eye-movement control in reading (including discussions of the pros and cons of the models compared to E-Z Reader). Fourth, in section 5 we will discuss future directions and ways that we intend to extend the E-Z Reader model. In this section, we will also discuss a possible mapping between model components and neurophysiological mechanisms. Finally, we will provide some concluding comments in section 6.

2. Eye movements in reading

Any discussion of models of eye-movement control must begin with a brief overview of eye movements during reading. In this section, we will describe what is known about eye movements during reading as background material. The following topics will be discussed: (1) saccades and fixations; (2) visual acuity; (3) saccade latency; (4) the acquisition of information during eye fixations; (5) perceptual span; (6) parafoveal preview effects; (7) regressions; (8) eye-movement control (where to fixate next and when to

 0140-525X/03 $12.50

move the eyes); (9) measures of processing time. It is not our intention to provide a complete and comprehensive review of each of these topics, as our primary purpose in this article is to compare different models of eye movement control in reading. The interested reader is invited to consult Rayner (1998) for a more complete review of each of the nine topics discussed in this section.

2.1. Saccades and fixations

Contrary to our subjective impression, the eyes do not move smoothly across the printed page during reading. Instead, the eyes make short and rapid movements, called *saccades* (Erdmann & Dodge 1898; Huey 1908), that typically move them forward about 6–9 character spaces, although there is considerable variability (Rayner 1978; 1998). Since the distribution of saccade sizes, measured in number of character spaces, is largely independent of visual angle when the number of character spaces is held constant (Morrison & Rayner 1981; O'Regan 1983), virtually all studies of reading use number of character spaces as the appropriate metric. Saccades take 20–50 msec to complete, depending upon the length of the movement, and virtually no visual information is extracted during eye movements (Ishida & Ikeda 1989; Wolverton & Zola 1983). Between saccades, the eyes remain stationary for brief periods of time (typically 200–250 msec) called *fixations* (Erdmann & Dodge 1898; Huey 1908). Because visual information is only extracted from the printed page during fixations, reading is similar to a slide show in which short segments of text are displayed for approximately a quarter of a second. It is important to note that there is considerable variability in both saccade length and fixation duration. Some saccades only move the eyes a single character, whereas others are as large as 15–20 characters (although such long saccades typically follow *regressions,* or backward movements to previous parts of the text, and place the eyes beyond the place from which the regression was initiated). Likewise, some fixations are shorter than 100 msec and others are longer than 400 msec (Rayner 1978; 1998). Much of this variability apparently is related to the ease or difficulty involved in processing the currently fixated text.

Erik D. Reichle is Assistant Professor of Psychology at the University of Pittsburgh, and a Research Scientist at the Learning Research and Development Center. He is the author of several papers on eye-movement control in reading, and is on the Editorial Board of the journal *Psychological Review* (beginning 2004). His research interests include word identification, attention, and eye-movement control.

Keith Rayner is Distinguished University Professor of Psychology at the University of Massachusetts, Amherst. He has published extensively in the areas of: eye movements, reading, language processing, scene perception, and visual search. He is the author of *The Psychology of Reading* (with A. Pollatsek) and has edited five others volumes. He is the former Editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition* (1990–1995) and current Editor of *Psychological Review* (beginning 2004).

Alexander Pollatsek is Emeritus Professor of Psychology at the University of Massachusetts, Amherst. He is co-author of the book *The Psychology of Reading* (Rayner & Pollatsek 1989) and has served on the Editorial Board of numerous journals. His research deals with reading, eye movements, phonological processing, scene perception, and decision making, and he has published widely in these areas.

2.2. Visual acuity

One of the reasons that the eyes are constantly moving in reading is that there are severe limits to how much visual information can be processed during a fixation. Visual acuity is maximal in the center of the retina and rapidly decreases towards the periphery, and fine visual discriminations can only be made within the *fovea,* or central 2° of vision. As a result, the visual features that make up individual letters can be encoded only from a very narrow window of vision. The practical significance of this limitation is that it is necessary to fixate most words so that they can be identified. Indeed, there is considerable evidence that a word becomes increasingly difficult to identify as the angular disparity between the fovea and the retinal image of a word increases (Rayner & Bertera 1979; Rayner & Morrison 1981). Explaining how the reader deals with this limited acuity is one constraint on any model of eye movements.

2.3. Saccade latency

A second kind of constraint on any model of reading stems from the "race" between the processes identifying words and the need to plan a saccade early enough in a fixation so that reading can carry on at about 300 words per minute. On the one hand, experiments in which subjects move their eyes to visual targets indicate that the *saccadic latency,* or the time needed to plan and execute a saccade, is approximately 180–250 msec (Becker & Jürgens 1979; Rayner et al. 1983b). This suggests that the decision to make a saccade is often made within the first 100 msec of a fixation. However, this is seemingly at odds with the intuitively appealing idea that word recognition is a major contributor to driving eye movements during reading because most estimates indicate that lexical access requires 100–300 msec to complete (Rayner & Pollatsek 1989; Schilling et al. 1998; Sereno et al. 1998;). It is thus not immediately obvious how the identification of one word can be the signal to begin planning a saccade to the next. Indeed, early theories of eye movements in reading (Bouma & de Voogd 1974; Kolers 1976) posited that word identification was too slow to be the engine driving eye movements.

2.4. The acquisition of information during reading

During saccades, vision is suppressed so that the information needed for reading is acquired only during fixations (Ishida & Ikeda 1989; Wolverton & Zola 1983). Furthermore, reading proceeds quite smoothly if text is available for processing for only the first 50–60 msec of a fixation prior to the onset of a masking pattern (Ishida & Ikeda 1989; Rayner et al. 1981). This does not mean that words are identified within 50 msec, but rather, that the information that is needed for reading gets into the processing system within 50–60 msec.

2.5. Perceptual span

One solution to the quandary of how word identification can be a signal to move the eyes is that words can be partially processed in the *parafovea,* or region of the retina that extends five degrees on either side of the fovea. McConkie and Rayner (1975) demonstrated the importance of parafoveal processing using an *eye-contingent display change technique,* called the *moving-window paradigm,* which is illustrated in Figure 1. In this paradigm, the letters outside of a "window" spanning a given number of character spaces is distorted in some way (e.g., replaced with Xs). By varying the size of the window and making its location contingent upon where the reader is looking, it is possible to determine the *perceptual span,* or region from which useful visual information can be encoded. With alphabetic text (like English), readers can progress at a more-or-less normal rate when the window extends 14–15 character spaces to the right (McConkie & Rayner 1975; Rayner 1986; Rayner & Bertera 1979; Rayner et al. 1982; Den Buurman et al. 1981) and 3–4 character spaces to the left of the fixation point (McConkie & Rayner 1976; Rayner et al. 1980). However, word encoding probably does not extend more than 7–8 characters to the right of fixation (McConkie & Zola 1984; Rayner et al. 1982; Underwood & McConkie 1985); beyond this distance, only low-spatial frequency information about letter shape (e.g., descenders vs. ascenders) and word length is extracted from the page. The left-right asymmetry reflects covert attention and is language specific; with Hebrew text (which is read from right to left), the perceptual span extends asymmetrically to the left of fixation (Pollatsek et al. 1981).

Four other points about the perceptual span are relevant. First, the perceptual span does not extend below the line that is currently being read (Inhoff & Briihl 1991; Inhoff & Topolski 1992; Pollatsek et al. 1993); readers focus their attention on the line that they are currently reading. Second, studies using various eye-contingent display change techniques have revealed that the size of the span is fairly constant for readers of similar alphabetic orthographies (such as English, French, and Dutch; see Rayner [1998] for further details). Third, characteristics of the writing system influence not only the asymmetry of the span, but also the overall size of the perceptual span. Thus, the span is smaller for Hebrew than for English (Pollatsek et al. 1981) because Hebrew is a more densely packed language than English. And it is much smaller for writing systems like Japanese (Ikeda & Saida 1978; Osaka 1992) and Chinese (Inhoff & Liu 1998) that have ideographic components and hence are even more densely packed than Hebrew. Fourth, the perceptual span is not hardwired, but rather seems to be attention-based. The fact that there is an asymmetry due to the direction of the writing system is consistent with the span being attention-based. In fact, Pollatsek et al. (1982) found that the perceptual span of Israeli readers who were bilingual in Hebrew and English had opposite asymmetries when reading the two languages. Furthermore, Rayner (1986) found that the span was smaller for beginning readers than skilled readers and that the span got smaller when children with four years of reading experience were given text that was too difficult for them. Analogous to this finding, Henderson and Ferreira (1990; see also Inhoff et al. 1989; Kennison & Clifton 1995; Schroyens et al. 1999) found that the span got smaller when the fixated word was difficult to process. Finally, Balota et al. (1985) found that readers obtained more information to the right of fixation when the upcoming word was highly predictable from the preceding text.

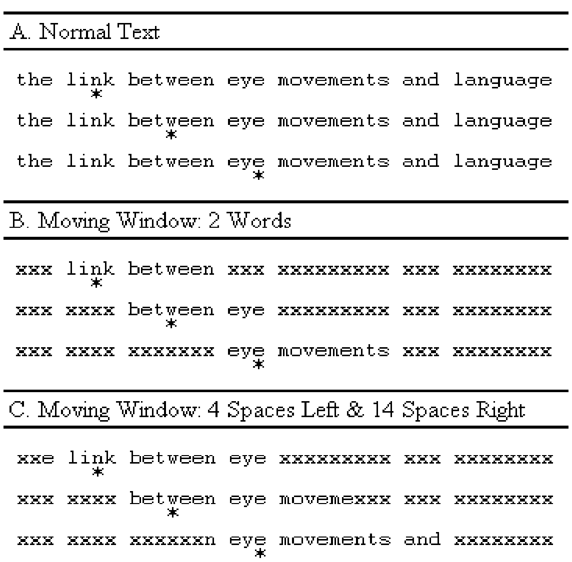

Figure 1. The moving-window paradigm. Panel A shows the positions of three successive fixations (indicated by the asterisks) in a normal line of text. Panels B and C illustrate how a "window" of normal text is displayed contingent upon where the eyes are currently looking. Panel B shows a two-word moving window; that is, both the fixated word and the word to the right of fixation are displayed normally, and all of the letters in the remaining words are replaced by Xs. In Panel C, the window extends four character spaces to the left of fixation and 14 character spaces to the right of fixation.

2.6. Parafoveal preview effects

Consistent with the findings of the last section, it has been demonstrated that orthographic (Balota et al. 1985; Binder et al. 1999; Rayner 1975) and phonological (Pollatsek et al. 1992) processing of a word can begin prior to the word being fixated. These results indicate that, during normal reading, the *parafoveal preview* of a word can reduce the duration of the subsequent fixation on the word, which is one measure of the time needed for identification (Schilling et al. 1998). Surprisingly, neither semantic (Altarriba et al. 2001; Rayner et al. 1986) nor morphological (Kambe, in press; Lima 1987; Lima & Inhoff 1985) information extracted from the parafovea appears to be of any benefit when the word is later fixated.[3] Furthermore, parafoveal preview benefit is not due to retention of visual featural information, as the case of all the letters can change from fixation to fixation with virtually no disruption to the reading process (McConkie & Zola 1979; Rayner et al. 1980). Instead, the source of the preview benefit seems to be due to abstract letter codes and phonological codes (see Rayner [1998] for a review). However, parafoveal information can produce word skipping (i.e., the word is not fixated) because words that can be identified in the parafovea do not have to be fixated and can therefore be skipped. Many experiments (Balota et al. 1985; Binder et al. 1999; Ehrlich & Rayner 1981; Rayner et al. 2001; Rayner & Well 1996; Schustack et al. 1987) have demonstrated that predictable words are skipped more than unpredictable words and that short function words (like "the") are skipped more than content words (O'Regan 1979; 1980; Gautier et al. 2000). When words are skipped, there is some evidence suggesting that the durations of the fixations preceding and following the skip are inflated (Pollatsek et al. 1986; Reichle et al. 1998).[4]

2.7. Regressions

One indicator of the inherent difficulty of reading (even for skilled readers) is that 10–15 percent of the saccades move the eyes back to previous parts of the text. These backward movements, called *regressions,* are thought to result both from problems with linguistic processing and from oculomotor error. The hypothesis that regressions can be caused by difficulties in linguistic processing is perhaps most clearly supported by the finding that regressions can be induced with structurally difficult "garden path" sentences; because such sentences often lead to incorrect syntactic analyses, readers often make regressions back to the point of difficulty and then re-interpret the sentence (Frazier & Rayner 1982). The idea that regressions are sometimes due to simple motor error is supported by the finding that, when the eyes fixate near the end of a word, they often move back a few character spaces (O'Regan 1990). This presumably happens because the eyes overshot their intended target (near the middle of the word) and a second fixation location affords a better place from which to see the word. This interpretation is consistent with the finding that identification is most rapid if a word is fixated just to the left of its center, on the *optimal viewing position* (Clark & O'Regan 1999; O'Regan 1990; 1992b; O'Regan et al. 1984).

2.8. Eye-movement control

Numerous studies have attempted to determine the characteristics of the mechanisms that control eye movements during reading. There are two different activities that must be explained: (1) What determines where the reader decides to look next? and (2) What determines when the reader moves his/her eyes (either forward or backward in the text)? Although there is not total consensus on these issues, there is some evidence to suggest that decisions about where to fixate next and when to move the eyes are made somewhat independently (Rayner & McConkie 1976; Rayner & Pollatsek 1981). The earliest unambiguous demonstration that the duration of the current fixation and the length of the next saccade are computed online was provided by Rayner and Pollatsek (1981). They varied physical aspects of the text randomly from fixation to fixation and found that the behavior of the eyes mirrored what was seen on a fixation. In their first experiment, they used the moving window paradigm described above and varied the size of the window randomly from fixation to fixation, and found that saccade length varied accordingly. Thus, if the window on the current fixation was small, the eyes only moved a few characters, while if it was large, the eyes moved further. In their second experiment, they delayed the onset of text in the fovea via a mask that appeared at the beginning of a fixation (with the time the mask was on varying randomly from fixation to fixation) and found that fixation durations were adjusted accordingly. In addition, the manipulations affected saccade length and fixation duration independently; in the first experiment, saccade length was affected, but fixation duration was not, whereas in the second experiment, fixation duration was affected, but saccade length was not. Thus, while the decisions about where to fixate next and when to move the eyes may sometimes overlap (see Rayner et al. 2000), there is reason to believe the two decisions are made somewhat independently.

2.8.1. Where to fixate next. Decisions about where to fixate next seem to be determined largely by low-level visual cues in the text, such as word length and the spaces between words. Five types of results are consistent with this claim. First, saccade length is influenced by the length of the fixated word and the word to the right of fixation (Blanchard et al. 1989; O'Regan 1979; 1980; Rayner 1979; Rayner & Morris 1992). Second, when readers do not have information about where the spaces are between upcoming words, saccade length decreases and reading is slowed considerably (McConkie & Rayner 1975; Morris et al. 1990; Pollatsek & Rayner 1982; Rayner et al. 1998a). Third, although there is some variability in where the eyes land on a word, readers tend to make their first fixation about halfway between the beginning and the middle of the word (Rayner 1979; McConkie et al. 1988; 1989; 1991; Vitu 1991b). Recently, Deutsch and Rayner (1999) demonstrated that the typical landing position in Hebrew words is likewise between the beginning (i.e., right-most end) and middle of a word. Rayner (1979) originally labeled this prototypical location the *preferred viewing location.* This position where the eyes typically land in a word is different from the *optimal viewing location,* which is the location in the word at which recognition time is minimized. According to O'Regan and Levy-Schoen (1987), the optimal viewing position is a bit to the right of the preferred viewing location, closer to the center of the word. Fourth, while contextual constraint influences skipping, in that highly predictable words are skipped more than unpredictable words (Balota et al. 1985; Ehrlich & Rayner 1981), contextual constraint has little influence on where the eyes land in a word (Rayner et al. 2001).[5] Finally, the landing position on a word is modulated by the *launch site* (McConkie et al. 1988; Radach & Kempe 1993; Radach & McConkie 1998; Rayner et al. 1996) because the landing position varies as a function of the distance from the prior fixation. As the launch site moves further from the target word, the distribution of landing positions shifts to the left and becomes more variable (see Fig. 2).

2.8.2. When to move the eyes. The ease or difficulty associated with processing a word primarily influences when the eyes move. Although a case can be made that low-level non-linguistic factors can also influence the decision about when to move the eyes, the bulk of the evidence suggests that linguistic properties of words are the major determiner of when to move. A very robust finding is that readers look longer at low-frequency words than at high-frequency words (Altarriba et al. 1996; Henderson & Ferreira 1990; 1993; Hyönä & Olson 1995; Inhoff & Rayner 1986; Just & Carpenter 1980; Kennison & Clifton 1995; Lavigne et al. 2000; Raney & Rayner 1995; Rayner 1977; Rayner & Duffy 1986; Rayner & Fischer 1996; Rayner & Raney 1996; Rayner et al. 1996; 1998a; Sereno 1992; Vitu 1991b; Vitu et al. 2001). There are three additional points with respect to this finding that are relevant. First, there is a *spillover* effect associated with fixating a low-frequency word; that is, fixation time on the next word is inflated (Rayner & Duffy

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

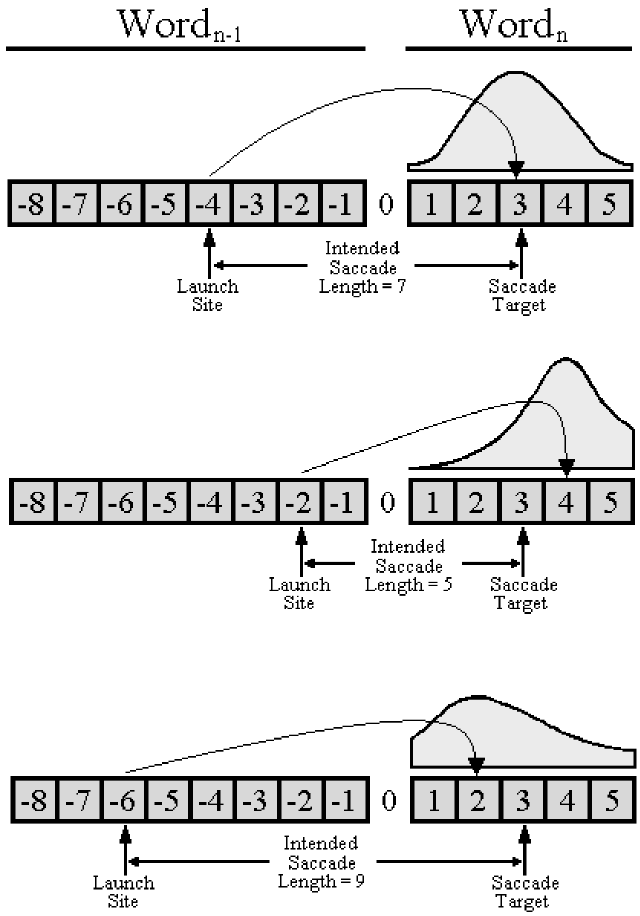

Figure 2. Landing site distribution as a function of the saccade length between the launch site (word$_{n-1}$) and intended saccade target (word$_n$). In all three panels, the launch site and target words are depicted by rectangles, with character spaces represented by numbers (as per convention, the space to the left of word$_n$ is denoted by a zero). The landing site distributions are approximately Gaussian in shape. Although the distributions are centered near the middle of the saccade targets, the oculomotor system is biased towards making saccades approximately seven character spaces in length. This bias results in a systematic range error; that is, the eyes tend to overshoot close targets and undershoot more distant targets. For example, in the middle panel, the intended saccade target is five character spaces from the launch site, so that (on average) the eyes overshoot their intended target, thereby causing the landing site distribution to shift towards the end of word$_n$. In the bottom panel, the opposite happens: The eyes undershoot their target, causing the landing site distribution to shift towards the beginning of word$_n$.

1986). Second, although the duration of the first fixation on a word is influenced by the frequency of that word, the duration of the prior fixation is not (Carpenter & Just 1980; Henderson & Ferreira 1993; Rayner et al. 1998a). Third, high-frequency words are skipped more than low-frequency words, particularly when they are short and the reader is fixated close to the beginning of the word (O'Regan 1979; Rayner et al. 1996).

A second important finding is that there is a predictability effect on fixation time in addition to a frequency effect. Words that are highly predictable from the preceding context are fixated for less time than are words that are not so constrained (Altarriba et al. 1996; Balota et al. 1985; Binder et al. 1999; Ehrlich & Rayner 1981; Inhoff 1984; Lavigne et al. 2000; Rayner & Well 1996; Rayner et al. 2001; Schustack et al. 1987; Zola 1984). Generally, the strongest effects of predictability are not as large as those of the strongest frequency effects. Also, as we noted above, predictability has a strong effect on word skipping: Words that are highly predictable from the prior context are skipped more often than words that are not so constrained.

2.9. Measures of processing time

To investigate the components of reading, researchers typically have subjects read sentences or passages of text while an eye tracker interfaced with a computer records the locations and durations of individual fixations. Because an average college-level reader can read approximately 300 words per minute (Rayner & Pollatsek 1989), this technique produces a staggering amount of data. Accordingly, the data are usually reduced to *word-based measures,* which are across-subject averages that reflect how often and for how long individual words are fixated. A number of word-based measures are standard (Inhoff & Radach 1998; Liversedge & Findlay 2000; Rayner 1998; Rayner et al. 1989; Starr & Rayner 2001). The first is *gaze duration,* which is defined as the sum of all fixations on a word, excluding any fixations after the eyes have left the word (i.e., including only *refixations* before the eyes move on to another word). Gaze duration is usually averaged only over words that are not skipped during the initial encounter (or *first pass*) through that region of text. Two other common measures are *first-fixation duration* and *single-fixation duration.* The former is the duration of the first fixation on a word (again conditional on the word being fixated during the first pass through the text), while the latter is the average fixation duration on words that are fixated exactly once during the first pass. These indices are typically reported along with indices of how often a word was fixated, which reflect the probability of a word being skipped, fixated once, and fixated more than once before moving to another word. Often, the *total time* (the sum of all fixations on the word, including regressions back to the word) is also reported.

The word-based measures provide a complete record of where and when fixations occurred. These two aspects (where vs. when) also provide a useful framework for organizing a discussion of reading models because much of the controversy surrounding reading concerns the determinants of where and how long the eyes remain fixated. The models that have been developed to explain eye-movement control form a continuum, extending from models in which eye movements are determined primarily by oculomotor factors (*oculomotor models*) to those in which eye movements are guided by some form of cognitive control (*processing models*). Prior to comparing different models, we will discuss our model, E-Z Reader (Pollatsek et al. 1999b; Rayner et al. 1998c; 2000; Reichle et al. 1998; 1999; Reichle & Rayner 2001) in some detail. We will also provide an updated version of the model (E-Z Reader 7).

3. E-Z Reader

E-Z Reader is a processing model, and extends the earlier work of Morrison (1984). Morrison drew much of the inspiration for his model from the work of Becker and Jür-

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

gens (1979) and McConkie (1979). McConkie (1979) suggested that, during reading, visual attention progressed across a line of text until the limitations of the visual system made it difficult to extract further lexical information; once this point of difficulty has been established, attention shifts and an eye movement is programmed and subsequently initiated, sending the eyes to the problematic location. Although elegantly simple, the model was soon discarded due to problems in defining and explaining what the point of difficulty was, how it might be computed, and whether it could be computed soon enough to be of any use in skilled reading (Rayner & Pollatsek 1989).

The limitations inherent in McConkie's (1979) early model of eye-movement control led Morrison (1984) to propose a model in which the movement of the eyes was a function of successful processing. According to Morrison, the identification of word$_n$ (i.e., the word that is currently being fixated) causes the attention "spotlight" (Posner 1980) to move to word$_{n+1}$, which in turn causes the oculomotor system to begin programming a saccade to word$_{n+1}$. If the program finishes before word$_{n+1}$ is identified, then the saccade will be executed and the eyes will move to word$_{n+1}$. However, if word$_{n+1}$ is identified before the program finishes, the saccade to word$_{n+1}$ may be cancelled. Cancellation can occur some of the time when attention shifts to word$_{n+2}$ while word$_n$ is fixated. In this case, the oculomotor system begins programming a saccade to word$_{n+2}$, which overrides the program to move the eyes to word$_{n+1}$ if the new program interrupts the old program soon enough. Thus, according to Morrison, attention moves serially, from word to word, whereas saccades can be programmed in parallel.

Morrison's (1984) assumption about the parallel programming of saccades followed Becker and Jürgens' (1979) demonstration that saccadic programming is completed in two stages: an initial, labile stage that is subject to cancellation, and an ensuing, non-labile stage in which the program cannot be cancelled. Their results suggested that if the oculomotor system begins programming a saccade while another saccadic program is in its labile stage of development, then the first program is aborted. However, if the second program is initiated while the first saccadic program is in its non-labile stage, then both saccades will be executed, which typically results in a very short fixation between the two saccades.

With these simple assumptions, Morrison (1984) was able to provide an elegant account of both frequency effects and parafoveal preview effects: Because short frequent words are more easily identified in the parafovea than long infrequent words, the former tend to be fixated for less time (and skipped more often) than the latter. Despite its successes, however, Morrison's model cannot explain refixations because the strictly serial attention shifts mean that each word is either fixated exactly once or is skipped.

More fundamentally, however, because Morrison's model posits both that processing of words is strictly serial and that attention shifting is time-locked to word identification, the model is unable to handle some simple and robust phenomena in reading. The first, as we noted above, is that one often gets "spillover" effects due to word frequency (e.g., Rayner & Duffy 1986). That is, lower-frequency words often not only cause longer fixations on that word (word$_n$), but also lengthen either gaze durations and/or first fixations on the succeeding word (word$_{n+1}$). According to Morrison's model, this shouldn't happen because attention doesn't shift until word$_n$ has been processed. Because parafoveal processing on word$_{n+1}$ begins after this attention shift, the amount of information extracted from word$_{n+1}$ before it is fixated will only be a function of how long it takes to program and execute the saccade, and will not vary as a function of the frequency of word$_n$. As a result, Morrison's model predicts no delayed effects of word frequency (or any other delayed effects of word processing difficulty). A related phenomenon (Henderson & Ferreira 1990; Kennison & Clifton 1995) is that the benefit gained through parafoveal preview decreases as foveal processing becomes more difficult (e.g., because the fixated word is lower frequency). By essentially the same argument as above, Morrison's model predicts that this shouldn't happen because parafoveal preview time is only a function of the latency of moving the eyes after covert attention has shifted.

There are at least three ways to circumvent the limitations of Morrison's (1984) model. The first is to add the assumption that if word identification is not completed by a processing deadline, attention does not shift to the next word, but instead remains on the current word, resulting in a refixation (Henderson & Ferreira 1990; Sereno 1992). This leads to the prediction (which has not been supported; Rayner et al. 1996; Schilling et al. 1998) that the first of two fixations should be longer than single fixations because the former reflect cases in which the processing deadline must have been reached. The second solution is to simply assume that difficulties with higher-order linguistic processing somehow cause the eyes to remain on the current word (Pollatsek & Rayner 1990; Rayner & Pollatsek 1989). Unfortunately, how this happens has not been well specified. Finally, a third way to avoid the shortcomings of Morrison's proposal is to assume that word identification is completed in two stages. This last approach is instantiated by E-Z Reader, which is discussed next.

3.1. Overview of the E-Z Reader model

E-Z Reader, like other processing models, makes the basic assumption that ongoing cognitive (i.e., linguistic) processing influences eye movements during reading. Because the model was not intended to be a deep explanation of language processing, it does not account for the many effects of higher-level linguistic processing on eye movements (for reviews, see Rayner 1998; Rayner & Sereno 1994; Rayner et al. 1989). Although this is clearly a limitation, it should also be noted that many of these effects typically occur when the reader is having difficulty understanding the text that is being read, such as when a reader makes a regression to re-interpret a syntactically ambiguous "garden path" sentence (Frazier & Rayner 1982). The model can therefore be viewed as the "default" reading process. That is, we view the process of identifying words to be the forward "driving engine" in reading, as the process of knitting the words into larger units of syntax or meaning would be too slow (whether successful or not) to be a signal to decide how and when to move the eyes forward for skilled readers. Thus, we posit that higher-order processes intervene in eye-movement control only when "something is wrong" and either send a signal to stop moving forward or a signal to execute a regression. Hence, we view E-Z Reader as an explanation of what happens during reading when higher-level linguistic processing is running smoothly and doesn't

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

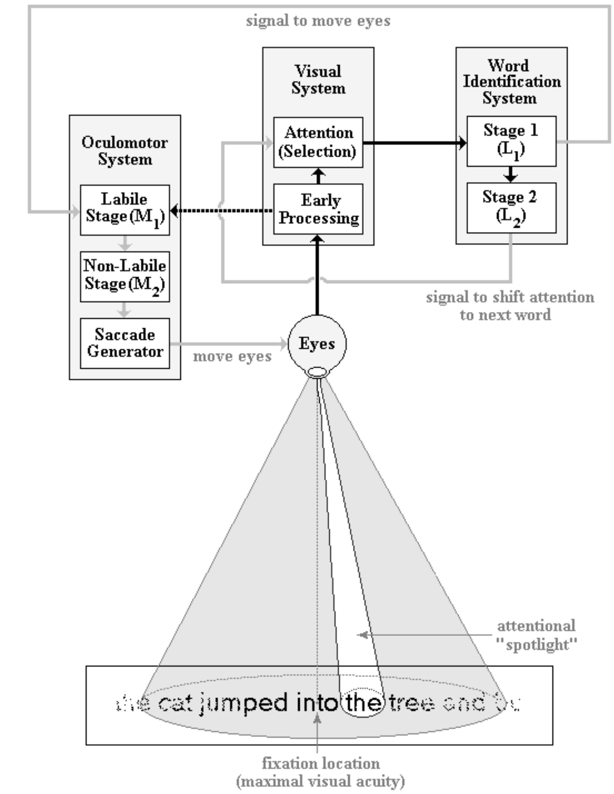

Figure 3. A schematic diagram of *E-Z Reader 7*. Visual features on the printed page are projected from the retina to an early stage of visual processing, which then proceeds at a rate that is modulated by visual acuity limitations. The low-spatial frequency information (e.g., word boundaries) is used by the oculomotor system to select the targets of upcoming saccades. High-spatial information is passed on to the word identification system, which, though attentional selection, allows individual words to be identified by the word identification system. The first stage of lexical processing (L_1) signals the oculomotor systems to begin programming a saccade to the next word. The completion of the second stage of word identification (L_2) causes attention to shift to the next word. Saccadic programming is thus decoupled from the shifts of attention. Saccadic programming is completed in two stages: The first, labile stage (M_1) can be cancelled by the initiation of subsequent programs; the second, non-labile stage (M_2) is not subject to cancellation. Saccades are executed immediately after the non-labile stage of saccadic programming has been completed. Black lines represent the flow of visual information, with the dashed line representing the low-spatial frequency information that is used by the oculomotor system to select the target locations of upcoming saccades. The gray lines represent signals that are propagated among the various components of the model (e.g., the signal to shift attention).

intervene. One implication of this is that the model currently does not explain inter-word regressions.

Like its immediate predecessors (see Reichle et al. 1998; 1999), E-Z Reader 7 consists of a small number of perceptual-motor and cognitive processes that determine when and where the eyes move during reading. Figure 3 is a schematic diagram showing the flow of control among these processes. As is evident in the figure, the central assumptions of the model are that: (1) a stage of word identification is the signal to move the eyes; and (2) attention is allocated from one word to the next in a strictly serial fashion. Notice, however, that both visual encoding limitations and oculomotor constraints also play central roles in the moment-by-moment control of eye movements during reading. In the discussion that follows, we will describe the specific assumptions of our model and how they are related to four major cognitive and perceptual-motor systems: visual processing, word identification, attention, and oculomotor control.

3.1.1. (Early) visual processing. Visual features from the printed page are projected from the retina to the visual cortex so that the objects on the page (i.e., the individual words) can be identified. The earliest stages of visual processing are thought to be pre-attentive in that the features that make up individual words are not fully integrated into perceptual wholes (Lamme & Roelfsema 2000; Wolfe & Bennett 1996). This processing is not instantaneous, with neural transmission from retina to brain taking approximately 90 msec to complete.

In our model, the preceding ideas are formalized by including the early processing stage in the visual system, which, though pre-attentive, is subject to visual acuity constraints (see Fig. 3). The duration of this early visual processing stage, $t(V)$, is a free parameter that corresponds to the base time needed for neural transmission to propagate from the retina to those cortical and subcortical areas that mediate early visual processing. To keep this assumption psychologically plausible, the value of $t(V)$ was set equal to 90 msec. However, because the rate of this early stage of processing is modulated by visual acuity, the rate at which a word is encoded is inversely proportional to both its length and its mean distance from the point of fixation. More specifically, during each fixation, the amount of early visual processing (in msec) that is completed on each word in the visual field is determined by:

$$\text{visual processing} = t/(\varepsilon^{\Sigma|i_{letter} - i_{fixation}|/N}) \qquad (1)$$

In Equation 1, t is the duration of the fixation (in msec), N is the number of letters in a word being processed, and ε (= 1.08) is a free parameter[6] that modulates the effects of the spatial disparity between each word's letters and the fixation location (i.e., the center of the fovea). Thus, the time needed to encode a word increases as the distance between its center and the fovea increases. Moreover, the time needed to encode a word also increases with its length because the individual letters of long words will (on average) be further away from the point of fixation than will the individual letters of short words.[7] One interesting implication of this equation is that the early visual processing of a word will be most rapid if the word is fixated near its center because a fixation on a word's center will minimize the mean spatial deviations between the fixation and each of the word's letters. This property is also consistent with evidence that word identification is most rapid if the word is fixated near its center (or optimal viewing position; O'Regan 1990; 1992b; O'Regan & Lévy-Schoen 1987; Vitu et al. 1990) and provides one explanation for why the eyes are seemingly directed towards this location during reading (see Shillcock et al. 2000). It also allows the model to account for length effects (i.e., the finding that long words take longer to identify than short words; Just & Carpenter 1980).

Early visual processing is important for two other reasons. First, it is necessary to obtain the word-boundary information that is needed to program saccades to upcoming words. This is denoted in Figure 3 by the dashed arrow that extends from early visual processing to the labile stage of saccadic programming. This arrow represents the flow of low-spatial frequency information that is acquired in the visual periphery (e.g., word boundaries, the presence/absence of ascenders and descenders, etc.). The oculomotor system uses this information to program saccades to upcoming words. Second, early visual processing provides the information that is subsequently used by higher-level visual areas to focus the attention "spotlight" and identify individual words. Word identification (which is discussed in the next section) must therefore wait until the early visual encoding of that word has been completed.

3.1.2. Word identification. The process of identifying a word begins as soon as attention is focused on that word. This identification process is then completed in two stages, reflecting early and late stages of lexical processing. The first stage corresponds to being at (or at least close to) the identification of the orthographic form of the word. We assume that this is not full lexical access, as the phonological and semantic forms of the word are not yet fully activated. We labeled this process the "familiarity check" (i.e., f) in earlier versions of the model, but in E-Z Reader 7 it is simply referred to as the first stage of lexical access (i.e., L_1).

The second stage of word identification involves the identification of a word's phonological and/or semantic forms so as to enable additional linguistic processing. This second stage, therefore, more or less corresponds to what is typically thought to be "lexical access." In prior versions of our model, this stage of word identification was called the "completion of lexical access" (i.e., *lc*). To avoid confusion, however, we will simply refer to this process as the second stage of lexical access (i.e., L_2) in E-Z Reader 7.

The distinction between early and late stages of lexical processing has precedent in the literature; indeed, our distinction was partly motivated by the *activation-verification model* of lexical access (Paap et al. 1982). The two models are broadly consistent if one conceptualizes the first stage of lexical access as a "quick and dirty" assessment of whether or not word identification is imminent, and the second stage as being the actual act of identification. As indicated in Figure 3, this distinction is also important because the two stages of lexical processing play unique functional roles: The completion of the first stage of lexical access causes the oculomotor system to begin programming the next saccade, while the completion of the second stage causes the "spotlight" of attention to shift to the next word. Thus, in E-Z Reader, saccadic programming is de-coupled from the shifting of attention.

As with earlier versions of our model, the time (in msec) required to complete the first stage of lexical access on a word, $t(L_1)$, is a linear function of the natural logarithm of the word's normative frequency of occurrence in printed text and its predictability within a given sentence context. The mathematical statement of this relationship is given by Equation 2:

$$t(L_1) = [\beta_1 - \beta_2 \ln(\text{frequency})]\,(1 - \theta\ \text{predictability}) \quad (2)$$

In Equation 2, β_1 and β_2 (= 228 and 10 msec, respectively) are free parameters that control how a word's normative frequency (number of occurrences per million, as tabulated by Francis & Kučera [1982]) affect lexical processing time. This time is also modulated by the right-hand term, in which the free parameter θ (= 0.5) attenuates the degree to which a word's predictability in a specific sentence context (as estimated using cloze task probabilities) attenuates the lexical processing time.[8] In all of the simulations reported below, the actual times needed to complete the first stage of lexical processing was found by sampling from gamma distributions having means equal to $t(L_1)$ and standard deviations equal to 0.18 of their means.

The completion of the first stage of lexical processing of a word has two immediate consequences in the model: (1) it cues the oculomotor system to begin programming a saccade to the next word (the details of how the oculomotor system does this will be discussed in detail below), and (2) it initiates further processing of the word. Because all (or at least most) of the orthographic coding has been completed in L_1, the time required to complete the second stage of lexical processing, L_2, is more influenced by a word's predictability. This distinction is reflected in Equation 3:

$$t(L_2) = \Delta[\beta_1 - \beta_2 \ln(\text{frequency})]\,(1 - \text{predictability}) \quad (3)$$

As in Equation 2, the free parameters β_1 and β_2 control the degree to which a word's frequency of occurrence affects the time necessary to process the word, but this quantity is attenuated by the free parameter $\Delta(= 0.5)$. Note that, in contrast to L_1, a word's predictability fully affects L_2; that is, words that can be predicted with complete certainty within a given sentence context will require no time in this second stage (i.e., if predictability = 1, then $t(L_2) = 0$ msec). Such cases reflect the situation when top-down information has already fully activated the semantic and phonological codes given reasonable corroborating input from orthography. As was the case with the first stage of lexical processing, the actual process durations were sampled from gamma distributions.

Finally, it should be mentioned that – by adding the early visual processing stage to E-Z Reader 7 – the minimal time needed to identify words in the model is very plausible. Given the parameter values reported above, for example, the mean time needed to identify the word "the" (the most frequent word in English text) when it is centrally fixated and in a completely predictable context is 148 msec, while the time needed to identify the lowest frequency words in completely unpredictable contexts is 432 msec. In contrast, E-Z Reader 6 predicted minimal and maximal mean word identification times of 16 and 278 msec, respectively. E-Z Reader 7 thus predicts word identification latencies that are much more in line with the best available estimates: 150–300 msec (Rayner & Pollatsek 1989).

3.1.3. Attention. A central, and perhaps the most contentious, assumption of E-Z Reader is that covert shifts of attention occur serially, from one word to the next, as each word is identified in turn and then integrated into the discourse representation. By "attention," though, we do not mean spatial orientation; rather, we refer to the process of integrating features that allows individual words to be identified. The separation between these two types of attention has considerable precedence in the literature (LaBerge 1990). For example, Treisman (1969) distinguished between *input selection,* or spatial orientation, and *analyzer selection,* or feature integration. This distinction is important because spatial orientation shifts towards the targets of upcoming saccades (Hoffman & Subramaniam 1995; however, see Stelmach et al. 1997), which in E-Z Reader occur whenever the oculomotor system uses the low-spatial frequency information provided by the visual processing stage

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

to program a saccade (see the dashed line in Fig. 3). These shifts in spatial orientation, however, are decoupled from the shifts in attention (i.e., analyzer selection) that precede lexical processing.

Attention is allocated serially during reading because readers need to keep word order straight (Pollatsek & Rayner 1999). By shifting the focus of attention from one word to the next, readers identify and process each word in its correct order. Although the results of several recent experiments (Inhoff et al. 2000b; Kennedy 1998; 2000; Kennedy et al. 2002; Starr & Inhoff, in press) suggest that properties of two words (particularly visual/orthographic properties) can sometimes be encoded in parallel, we suspect that this does not usually occur in normal reading (see Rayner et al. 2003c, for an extended discussion of these issues). The reason for this is that much of the information that is conveyed by language (both written and spoken) is heavily dependent upon word order.

Furthermore, by decoupling eye movements from attention, our model can also explain aspects of eye-movement control that Morrison's (1984) model could not. For example, E-Z Reader can explain why parafoveal preview benefit decreases as foveal processing difficulty increases (Henderson & Ferreira 1990; Kennison & Clifton 1995). If the eyes are on word$_n$, parafoveal processing of word$_{n+1}$ begins, not with completion of the first stage of lexical processing of word$_n$, but after the completion of second stage. Because parafoveal processing of word$_{n+1}$ ends (by definition) with the onset of the saccade to word$_{n+1}$, more time will remain for parafoveal processing of word$_{n+1}$ when word$_n$ is easy to process (e.g., high-frequency). This is depicted in Figure 4: The time required to complete L_1 and L_2 on word$_n$ increases as its normative frequency decreases (see Equations 2 and 3). Because the saccadic latency is not modulated by word frequency, a saccade will (on average) occur 240 msec (i.e., the mean saccadic latency) after the completion of L_1. This means that, with everything else being equal, the amount of time available to process word$_{n+1}$ in the parafovea will increase as the amount of time needed to process word$_n$ decreases.

In the model, the serial-allocation-of-attention assumption is instantiated as follows: The completion of the second stage of lexical processing on word$_n$ causes attention to shift to word$_{n+1}$, at which point the first stage of lexical processing begins on word$_{n+1}$ when pre-processing of word$_{n+1}$ is complete.[9] The identification of one word thus causes the focus of attention to shift so that the word-identification system can begin identifying the next word (see Fig. 3).

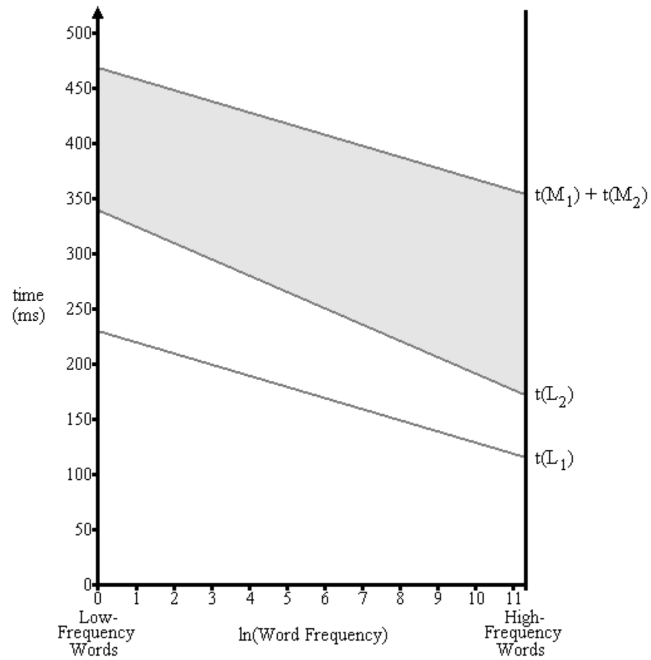

Figure 4. A diagram showing how parafoveal preview benefit is modulated by normative word frequency. The bottom line represents the time required to complete the first stage of lexical processing, $t(L_1)$, as a function of the natural logarithm of word$_n$'s token frequency. The middle line represents the time required to complete the second stage of lexical processing, $t(L_2)$, on word$_n$. Finally, the top line represents the saccadic latency, or time required to initiate a saccade from word$_n$ to word$_{n+1}$. On average, the saccadic latency requires a constant $t(M_1) + t(M_2)$ ms to complete (starting from the point in time when the first stage of lexical processing on word$_n$ has been completed). In E-Z Reader, parafoveal preview begins as soon as word$_n$ has been identified and attention has shifted to word$_{n+1}$. The parafoveal preview is therefore limited to the duration of the interval (depicted by the shaded area in the figure) between $t(L_2)$ and $t(M_1) + t(M_2)$. Notice that, because the relative disparity between $t(L_1)$ and $t(L_2)$ increases as the frequency of word$_n$ decreases, the duration of the parafoveal preview decreases with the frequency of word$_n$.

3.1.4. Oculomotor control. Saccadic programming in E-Z Reader is completed in two stages: an early, labile stage (M_1) that is subject to cancellation by subsequent programs, and a later, non-labile stage (M_2) that is not subject to cancellation. This assumption was motivated by demonstrations that a saccade to a first target can be cancelled by the presentation of a second to-be-fixated target if the second target is presented within approximately 230 msec after the first; after this time, both targets are typically fixated in sequence (Becker & Jürgens 1979). A considerable amount of subsequent research has supported this distinction between labile and non-labile stages of saccadic programming (Leff et al. 2001; McPeek et al. 2000; Molker & Fischer 1999; Vergilino & Beauvillain 2000).

During the first (labile) stage of saccadic programming, the eye-movement system is simply engaged (or made ready) so that it can begin programming an eye movement. The system then computes the distance between the current fixation location and the location of the saccade target (i.e., the *intended saccade length*). Thus, although the target location is represented in terms of spatial coordinates, the saccadic program is represented in terms of a distance metric. This is necessary because the distance that is specified by the saccadic program must ultimately be converted into the appropriate amount of force that has to be exerted (by the extraocular muscles) to execute the actual movement. The labile stage of programming therefore consists of two sub-stages: (1) general system preparation, followed by (2) a location-to-distance transformation, in which the spatial location of the upcoming saccade target is converted into the necessary saccade length. In E-Z Reader, the time needed to complete the labile programming stage is a random deviate that is sampled from a gamma distribution having a mean equal to a free parameter, $t(M_1)$, with each of the two aforementioned sub-stages subsuming half of this time.

An important part of our model is that, when a saccade

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

program is in the labile stage, it is subject to cancellation by a subsequent saccadic program. If the second program is initiated during the system preparation sub-stage of the first program, then whatever amount of preparation has been done to ready the oculomotor system will also be applicable to the second program, so that it will be completed more rapidly than it otherwise would be. If, however, the second program is initiated somewhat later, during the first program's location-to-distance transformation sub-stage, then whatever processing has been done to specify the distance of the first saccade will not apply to the second because the target locations (and hence the distances) of the two saccades are different. This means that the second program will always require a minimal amount of time to finish – the time necessary to convert the spatial location of the saccade target into the intended saccade length.

During the second (non-labile) stage of programming, the command to move the eyes a particular direction and distance is communicated to the motor system. At this point, an intended saccade is obligatory, and cannot be cancelled or modified by subsequent programs. As with the labile stage of programming, the time needed to complete the non-labile stage of programming is sampled from a gamma distribution, with the mean of this distribution being equal to a free parameter, $t(M_2)$. Upon completing the non-labile stage of programming, the saccade is executed immediately.

In E-Z Reader 7, the mean times needed to complete the labile, $t(M_1)$, and non-labile, $t(M_2)$, stages of saccadic programming were set equal to 187 and 53 msec, respectively. To keep the model as simple as possible, the saccade durations were set equal to a fixed value: $t(S) = 25$ msec.[10] Our saccadic-programming parameter values are consistent with estimates from simple saccade latency tasks (Becker & Jürgens 1979; McPeek et al. 2000; Rayner et al. 1983b). It should be noted, however, that these values are in fact estimates of the *minimal* time required to initiate a saccade, often to *pre-specified* targets; in the context of reading text, therefore, the average saccadic latency may be slightly longer in duration than would be suggested by these previous estimates.

Let us examine these assumptions using five key situations in reading. The first situation (shown schematically in Fig. 5A) is the simplest: Word_n is fixated, an eye movement is programmed to word_{n+1}, and no subsequent eye-movement command is made while this program is in its labile stage. The program therefore enters its non-labile stage, and an eye movement is made to word_{n+1}.

Now consider a second situation (Fig. 5B): Word_n is fixated, a program to fixate word_{n+1} is initiated, but while the oculomotor system is being readied, a second program (to move the eyes to word_{n+2}) is initiated. In this case, the program to fixate word_{n+1} is cancelled, and the saccade leaving word_n will move the eyes to word_{n+2} (i.e., word_{n+1} will be skipped). Whatever time elapsed in preparing the oculomotor system to program the first saccade will also be subtracted from the time that would otherwise be necessary to program the second saccade, thereby allowing it to be completed more rapidly than would otherwise be the case. Moreover, because situations like the one just described tend to occur when word_{n+1} is processed rapidly, the model successfully predicts that skipping is more likely to occur whenever word_{n+1} is high frequency, predictable from prior context, and/or short.

Now let's consider a situation (Fig. 5C) that is similar to the one just described: Word_n is fixated, and a program to fixate word_{n+1} is initiated. However, just as the labile stage of this program is about to finish (i.e., the location-to-distance transformation is almost complete), the oculomotor system begins programming a saccade to word_{n+2}. As in the previous situation, the program to fixate word_{n+1} will be cancelled, and the eyes will again go directly from word_n to word_{n+2}. Because the saccade length specified by the second saccade program is different from the length specified by the first, however, the duration of the second program's labile stage will include the time needed to recompute the distance between the location of the current fixation location and that of the new saccade target. The second program's labile stage will therefore be reduced, but only by the amount of time needed for general system preparation; that is, the second program's labile stage will equal the time needed to complete its location-to-distance transformation.

Finally, let us consider the situations depicted in Panels D and E of Figure 5: In both cases, word_n is fixated, the program to fixate word_{n+1} is initiated, and then (after some time) this program goes into its second, non-labile stage. At this time, a second program (to move the eyes to word_{n+2}) is initiated. In both of the situations depicted in Panels D and E, the program to fixate word_{n+1} will run to completion, and the eyes will move from word_n to word_{n+1}. However, in Panel D, the second program does not really benefit (i.e., it requires the full amount of time to be completed) because there was no ongoing labile program when the second program was initiated. Because the first saccade is actually executed while the second program is in its early, system-preparation phase, though, the second program's labile stage does not have to be re-started. In contrast, Panel E shows what happens when the first saccade is executed while the second program is in its location-to-distance transformation phase: Because the eyes are now fixated on word_{n+1}, the relative distance between the location of the current fixation and that of the saccade target (word_{n+2}) must be re-calculated. This means that the location-to-distance transformation has to be re-started, which extends the time needed to complete this part of the second saccade's labile programming.

Our discussion of saccadic programming so far has focused largely on the time needed to program the saccades, and has only addressed the question of where the eyes move at a fairly coarse level (i.e., at the level of individual words). As McConkie and his colleagues demonstrated (1988; 1991), saccades are prone to both systematic and random error. The effects of these sources of error are not negligible, and have been an oft-cited reason for the claim that the control of eye-movements during reading is primarily mediated by fairly low-level visual and oculomotor constraints (e.g., visual acuity limitations, systematic motor error, etc.; see O'Regan 1990; 1992b; O'Regan & Lévy-Schoen 1987; Reilly & O'Regan 1998). It is therefore important to specify how the model handles the effects of saccadic error.

Our assumptions regarding the oculomotor system are based on McConkie et al.'s (1988; 1991) data and analyses. In fact, we more or less directly incorporated their views of saccadic error into our model. In the model, saccades are directed towards the optimal viewing position of the words being targeted. However, these saccades are subject to both systematic and random error, so that, on average, saccades will deviate from their intended targets. More formally, each saccade is the sum of three components:

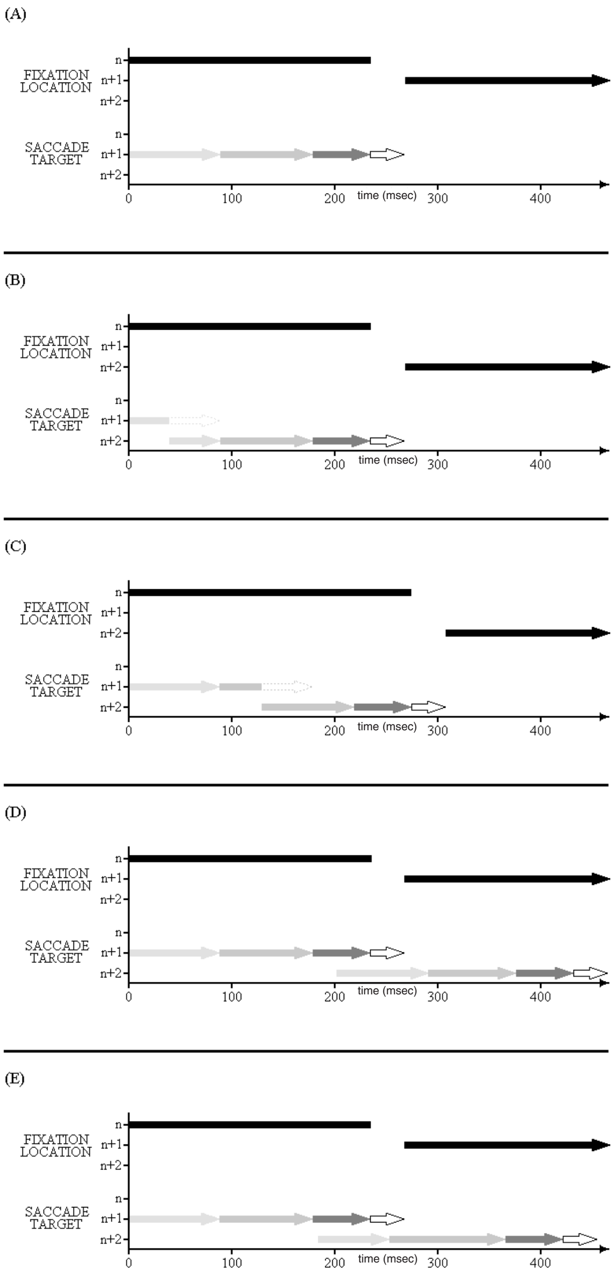

Figure 5. This diagram shows E-Z Reader 7's oculomotor control assumptions and how these assumptions affect saccadic programming in five common situations that occur during reading. In all of the panels, time (in msec) is represented along the horizontal axis, the black horizontal bars indicate the word (*n, n*+1, or *n*+2) that is being fixated at each given point in time, and the arrows represent the various stages of saccadic programs that are being directed towards specific word targets (*n, n*+1, or *n*+2). The light gray arrows represent the general preparation component of the first, labile programming stage, the medium gray arrows represent the location-to-distance transformation phase of the labile programming stage, and the dark gray arrows represent the second, non-labile stage of programming. The white arrows represent the actual saccades. In Panel A, one program follows another, and the eyes move in sequence from word$_n$ to word$_{n+1}$ to word$_{n+2}$. In Panels B and C, a program is initiated while another, labile program is in progress; in these situations, the first program is cancelled, and the eyes move from word$_n$ to word$_{n+2}$ (skipping word$_{n+1}$). Finally, in Panels D and E, the second program is initated while the first program is in its non-labile stage; in these situations, the first program runs to completion, and the eyes move in sequence from word$_n$ to word$_{n+1}$ to word$_{n+2}$.

$$\text{saccade} = \text{intended saccade length} + SRE + RE \qquad (4)$$

In Equation 4, the *intended saccade length* is the distance (in character spaces) between the current fixation (i.e., launch site) and the middle of the word that is the saccade target, and *SRE* and *RE* are the systematic and random error, respectively. The *SRE* emerges from the fact (at least for readers of English) that the oculomotor system "prefers" to make saccades that are seven character spaces in length. Saccades that are intended to be longer than seven character spaces tend to undershoot their targets, whereas saccades that are intended to be shorter than seven character spaces tend to overshoot their targets. The saccades that are executed tend to overshoot (or undershoot) by approximately a half of a character space for each character space that the intended target deviates from the preferred distance. This tendency is modulated by the duration of the launch site fixation, however, with longer fixations (on average) leading to greater saccade accuracy (McConkie et al. 1988; 1991). Both of these tendencies are instantiated in the model using Equation 5:

$$SRE = (\Psi - \text{intended saccade length})\,[\Omega_1 - \ln(\text{fixation duration})/\Omega_2] \qquad (5)$$

In Equation 5, Ψ is a free parameter representing the preferred saccade length: 7 character spaces. The discrepancy between this preferred distance and the length of the intended saccade is scaled by the right-hand term, which is a linear function of the natural logarithm of the launch site fixation duration. (The values of the free parameters Ω_1 and Ω_2 were fixed at 7.3 and 4, respectively.) Equation 5 thus ensures that the saccades that are executed will tend to overshoot (undershoot) their targets by approximately half of a character space for each character space that the intended saccade is less than (more than) seven character spaces. This systematic error is also modulated by the fixation duration on the launch site, so that there is less error following longer fixations.

The final term in Equation 4, *RE,* is the random error component. Consistent with McConkie et al.'s (1988; 1991) interpretations, this error term is normally distributed, with $\eta = 0$ and σ given by Equation 6. This equation stipulates that the size of the random error component increases proportional to the length of the intended saccade as determined by the values of the two free parameters, η_1 and η_2. (The values of these parameters were fixed at 1.2 and 0.15, respectively.)

$$\sigma = \eta_1 + \eta_2\ \text{intended saccade length} \qquad (6)$$

In closing this discussion of oculomotor control, we must revisit the issue of refixations. A key assumption of earlier versions of our model was that the oculomotor system begins programming a saccade to refixate a given word as soon as it is fixated. This saccade then ensues (resulting in a refixation) unless the first stage of lexical processing on that word finishes before the labile stage of programming, in which case the program is cancelled, and the oculomotor system begins programming a movement to the next word. This "horse race" between the initial stages of saccadic programming and lexical processing allowed the model to predict the correct proportion of refixations, but was problematic because it resulted in a non-monotonic relationship between the first-fixation durations and word frequency (i.e., the first-fixation durations on the low-frequency words were too short). This problem reflected an inherent limita-

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

tion of the "horse race" assumption. That is, to predict the correct proportion of refixations, the model's parameter values had to set so that the labile programming of automatic refixations completed before the first stage of lexical processing. As a result, the saccades that moved the eyes off the initial landing site (i.e., the refixation saccades) occurred very rapidly, causing the first of several fixations to be too short. Thus, although longer words had a greater probability of being refixated, in the process, they also had a greater number of first fixations that were too short.

In E-Z Reader 7, we modified our assumption about automatic refixations: Rather than being started by default, upon fixating a given word, a program is instead initiated with a probability that is determined by the length of the word that is to be fixated. (The low-spatial frequency information that is used to determine word length is rapidly available from parafoveal vision; see Fig. 3.) Upon fixating a word, the oculomotor system initiates a labile program to refixate the word with probability, p, given by Equation 7. In Equation 7, $\lambda(= 0.07)$ is a free parameter that modulates how word length affects the probability of making a refixation. The model thus correctly predicts that long words are more often the recipients of multiple fixations than are short words. Similarly, the model also correctly predicts more refixations on low-frequency words than on high-frequency words. This is true because the first stage of lexical processing will complete less rapidly on low-frequency words, and as a result be less likely to cancel any labile refixation programs that happen to be pending. Finally, it should be noted that E-Z Reader 7 – like its predecessors – predicts that a substantial proportion of refixations occur because saccades overshoot and undershoot their intended targets.

$$p = \begin{cases} \text{length}\,\lambda & \text{if}(\text{length}\,\lambda) < 1 \\ 1 & \text{if}(\text{length}\,\lambda) \geq 1 \end{cases} \tag{7}$$

3.2. Simulation results

E-Z Reader 7's performance was evaluated using data from an eye-tracking experiment in which 30 college students read 48 sentences containing 8–14 words each (Schilling et al. 1998). We used the norms of Francis and Kučera (1982) to estimate what the token frequencies of the words were for our readers. (For example, the word "torpedo" is used very infrequently in written text, and as a result occurs only once in the corpus, whereas "the," the most frequently used word, occurs 69,974 times.) Before running the simulations, we completed a separate "cloze-task" experiment in which participants had to guess word$_{n+1}$ when given the sentence up through word$_n$ so as to determine each word's mean predictability within its sentence context. Finally, because regressions are outside of the scope of the model, we did not include data from sentences in which readers made inter-word regressions.

The first simulation examined the model's capacity to predict the means and distributions of several commonly used word-based measures of fixation duration and probability. To do this, we first divided the words into five frequency classes. For each of the frequency classes, we computed the means of the following measures: first-fixation duration, single-fixation duration, and gaze duration; and the probability of making a single fixation, the probability of at least one refixation, and the probability of skipping a word. We also constructed first-fixation and gaze duration distributions. Finally, we ran a simulation using 1,000 statistical subjects to determine how well the model could predict the observed means and distributions. The observed and predicted means are presented in Figure 6, and the observed and predicted distributions are presented in Figure 7.

As can be seen in Figure 6, the model does an excellent job predicting the observed means ($r^2 = 0.94$ for fixation durations; $r^2 = 0.98$ for fixation probabilities). In particular, E-Z Reader 7 – in contrast to its predecessors – correctly predicts the negative monotonic relationship between first-fixation durations and word frequency. This pattern was inherently problematic for earlier versions of the model because the relatively slow lexical processing of low-frequency words rarely finished before the "automatic" program to make a refixation, thereby causing the first of several fixations (and the mean first-fixation durations) on low-frequency words to be too short. E-Z Reader 7 avoids this problem by eliminating the assumption that refixations are automatically programmed upon fixating a word.[11]

Figure 7 shows that the model generated first-fixation and gaze duration distributions that are very similar to those that were observed. In fact, this aspect of the model's performance is considerably better than that of its predecessors. Although we have not quantified this improvement, it

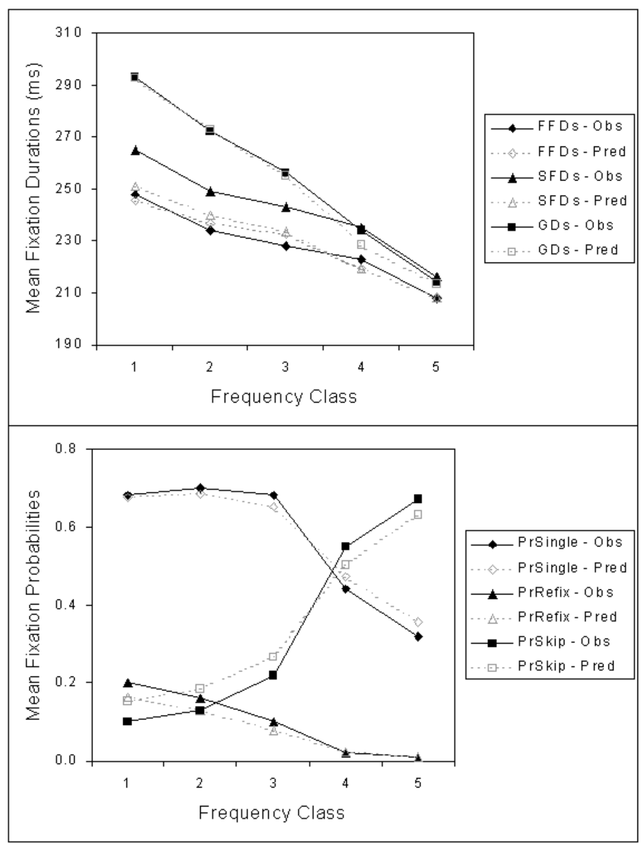

Figure 6. The top panel shows the mean observed and predicted first-fixation (FFD), single-fixation (SFD), and gaze durations (GD) for five frequency classes of words. The bottom panel shows the mean observed and predicted single-fixation (PrSingle), refixation (PrRefix), and skipping probabilities (PrSkip) for five frequency classes of words.

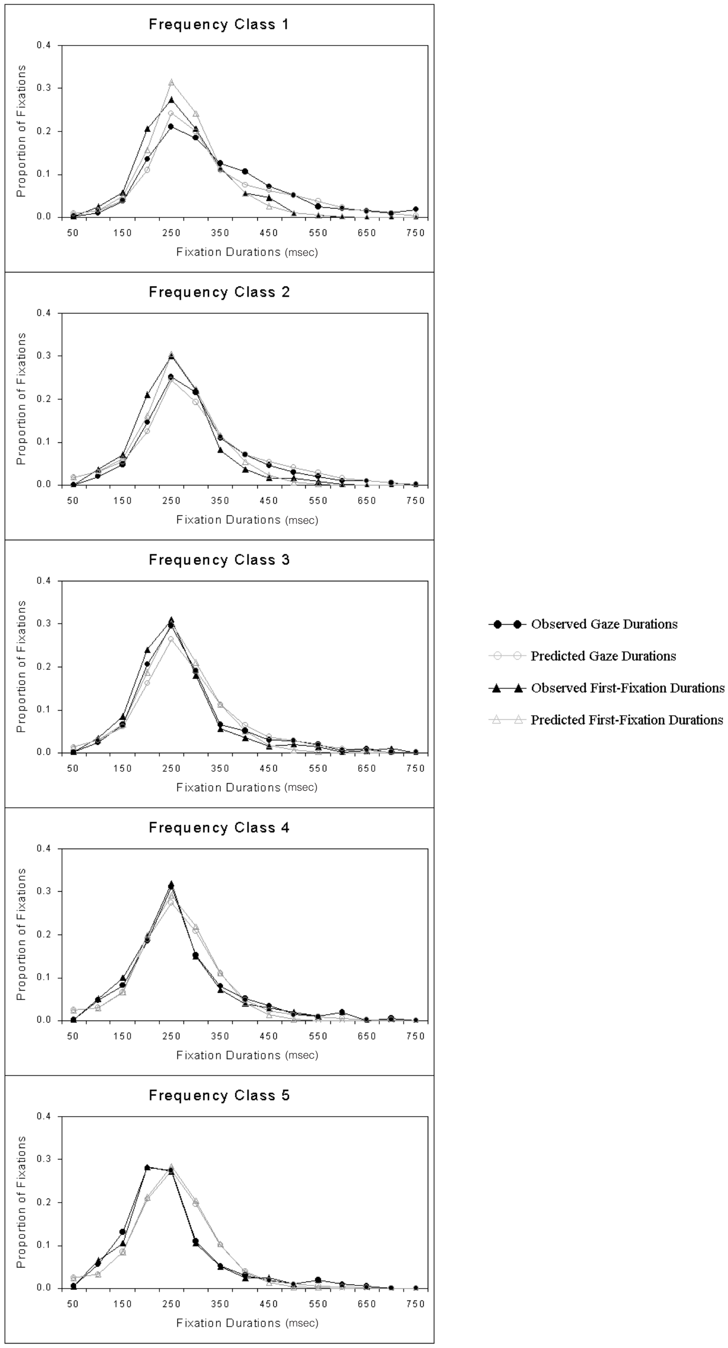

Figure 7. Observed and predicted frequency distributions of first-fixation (FFD) and gaze durations (GD). Each of the five panels shows the distributions for a separate frequency class of words. Each point represents the proportion of fixation durations within a given 50-msec interval (e.g., points above the abscissa labeled "100" represent the proportion of fixation durations between 50 and 100 msec that were observed in the sentence corpus and predicted by E-Z Reader 7).

is clear that the model is no longer over-predicting the amount of variability in the fixation durations (cf. Fig. 7 to Figs. 8 and 9 in Reichle et al. 1998).

Finally, we examined the first-fixation and gaze durations that were predicted for the low- and high-frequency target words that were used by Schilling et al. (1998) to study word-frequency effects during reading. In their experiment, Schilling et al. observed a mean gaze duration difference of 50 msec between the low- and high-frequency target words, as well as a 31-msec frequency effect on the first-fixation durations. E-Z Reader 7 predicted mean gaze and first-fixation duration frequency effects of 54 and 21 msec, respectively. The results of this simulation thus show that the model can handle both the aggregate properties of the Schilling et al. sentences and the frequency effects on specific words. Of course, previous versions of E-Z Reader could also account for a number of other "benchmark" phenomena; in the interest of evaluating the model further, therefore, we completed several additional simulations (each based on 1,000 statistical subjects).

In the first of these simulations, we first replaced the frequency values of all of the Schilling et al. (1998) target

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

words with the mean frequency of the high-frequency targets (141 per million). We then repeated this procedure using the mean frequency of the low-frequency targets (2 per million). In both cases, the mean frequency values were inserted into the same within-sentence word positions as the original targets. The reason for inserting the mean frequency values into the sentence "frames" is that any resulting between-target differences can be attributed entirely to those items. As expected, the model predicted 84- and 44-msec frequency effects on the gaze and first-fixation durations, respectively. More importantly, the model also predicted 30- and 24-msec spillover frequency effects (for gaze and first-fixation durations, respectively) on the words immediately following the targets. These results are consistent with demonstrations that such spillover effects are typically one-third to one-half of the size of frequency effects (Rayner & Duffy 1986; Rayner et al. 1989; Schilling et al. 1998).

The second simulation examined the effects of parafoveal preview. To do this, we calculated the gaze durations on the Schilling et al. (1998) targets both with and without parafoveal processing of these words. The former condition was simulated using the standard (normal) model; to simulate the latter condition, we "lesioned" the model so that the first stage of lexical processing, L_1, on the targets could begin only after the words had been fixated. (Visual pre-processing was allowed.) Typically, the gaze durations on words increase 40–60 msec in the absence of parafoveal preview. Our simulation indicated that, with no parafoveal processing, the model predicted a 26-msec increase in the gaze durations on the target words. Although this prediction is a little smaller than what is typically observed, it is not entirely unreasonable, especially if one considers that the model predicts an additional increase in gaze durations (90 msec) in the complete absence of early visual processing.

The third simulation examined the processing "costs" that are incurred on word$_n$ that are due to: (1) skipping word$_{n-1}$; or (2) skipping word$_{n+1}$. Typically, the gaze duration on word$_n$ will be longer if word$_{n+1}$ is skipped than if word$_{n+1}$ is fixated (Pollatsek et al. 1993; Reichle et al. 1998). Likewise, there is some evidence that the gaze durations on word$_n$ are longer if word$_{n-1}$ is skipped than if word$_{n-1}$ is fixed. To examine these effects, we first calculated the mean

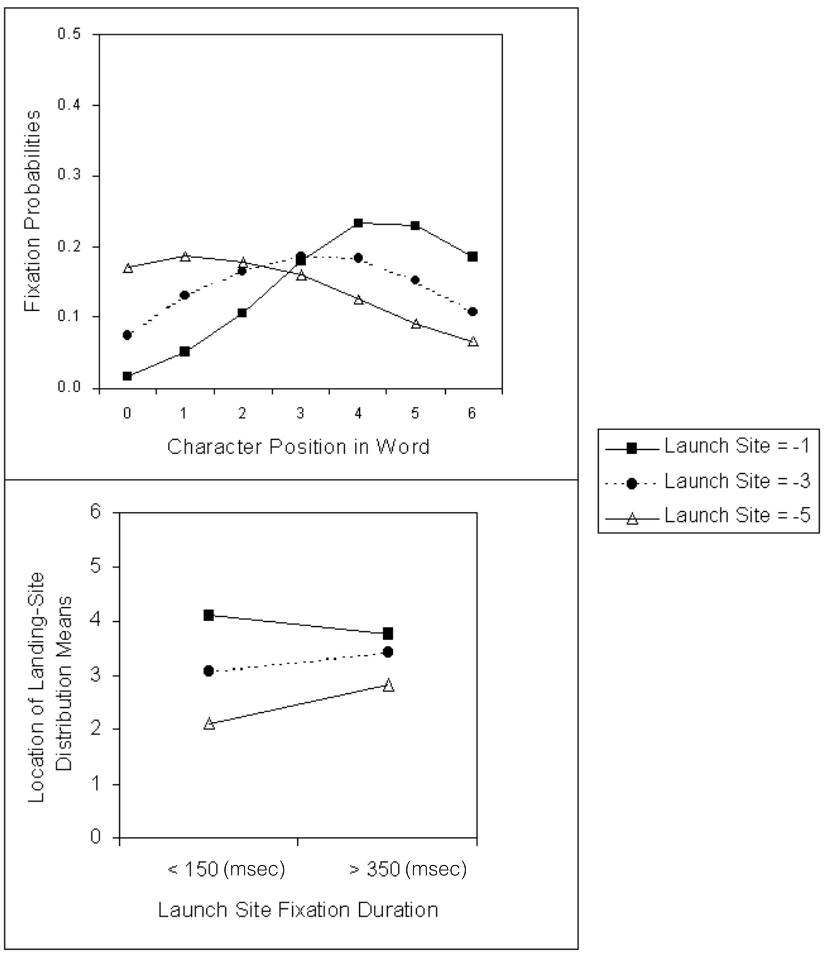

Figure 8. Simulation results showing the details of where the eyes move during reading. The top panel shows the landing site distributions on 6-letter words as a function of saccade length (i.e., the distance between the launch site and the middle of the saccade target). The locations of the launch sites and landing sites are indicated by numbers (in the legend and along the x-axis, respectively) representing ordinal position, from left to right, with the blank space between the two words being zero. The predicted landing sites are similar to those that have been reported elsewhere (e.g., McConkie et al. 1988; cf. Figs. 2 and 8A); that is, the distributions are approximately Gaussian in shape, with means that shift from near the word centers to near their beginnings with increasing saccade length. The bottom panel shows how the predicted systematic range error depicted in the top panel is modulated by the launch site fixation durations. As is evident, the systematic range error is attenuated following longer (above 350 msec) fixations on the launch site words.

gaze duration difference on the Schilling et al. (1998) target words when the following word was skipped versus fixated. The model should predict such an effect because, if $word_{n+1}$ is skipped, then the oculomotor system must modify the program to move the eyes to $word_{n+1}$ so that it instead moves the eyes to $word_{n+2}$, and such modifications take additional time. Indeed, the model predicted a 58-msec effect, which is similar is size to the 38-msec effect observed in the Schilling et al. (1998) corpus. Next, we calculated the mean gaze duration difference on the Schilling et al. targets when the immediately preceding word was skipped as opposed to fixated. Again, the model should handle this effect because, in cases where $word_{n-1}$ is skipped, any parafoveal processing that is done on $word_n$ will be completed from a more distant location than if $word_{n-1}$ is fixated. The model confirmed our predictions; it predicted a 66-msec effect, which again corresponds fairly closely to the 50-msec effect that was observed with the Schilling et al. materials. (E-Z Reader 7 handled these results significantly better than earlier versions of the model.)

The final simulation evaluated the model's capacity to account for the fine-grained details of where the eyes move during reading. This was done by examining the landing site distributions that were generated by E-Z Reader 7 on words of various lengths (again using the Schilling et al. [1998] sentences).[12] Figure 8A shows the landing site distributions that were predicted for 7-letter words. The figure indicates that the predicted landing site distributions closely resemble those reported by McConkie et al. (1988; 1991): (1) the landing sites are normally distributed; (2) the distribution means are located near the middle of the words; and (3) the distributions shift towards the beginnings of the words and become more variable as the distance between the launch sites and landing sites increases. Furthermore, as Figure 8B indicates, the magnitude of this systematic range error (i.e., how much the saccades over/undershoot their intended targets) is modulated by the launch-site fixation duration, so that there is less spread among the landing site distribution, means following longer launch site fixations. Together, the results of this final simulation are inconsistent with Reilly and O'Regan's (1998) claim that models like E-Z Reader cannot explain the patterns of landing site distributions that are normally observed during reading.

Before moving to the alternative models of eye-movement control, it is useful to note that Engbert and Kliegl (2001) sought to evaluate the basic assumption in E-Z Reader that lexical processing is the "engine" driving eye movements during reading. That is, they wanted to know if the time course of saccades is always determined by the time course of lexical processing. To answer this question, they implemented a computational model that, like E-Z Reader, accounts for eye-movement control during reading in terms of a few assumptions about lexical access and saccadic programming. There are two versions of the model, a two-state and a three-state version. The former is quite similar to a simpler version of E-Z Reader (Model 2 in Reichle et al. 1998), but there is only one stage of lexical processing, and it makes somewhat different assumptions about the variability of processes. The three-stage model is similar to the version of E-Z Reader that we are discussing except that, functionally, the first stage of lexical processing is replaced by an all-or-none process. That is, the reader is either assumed to wait until lexical access is completed before programming a saccade, or an "autonomous saccade" (i.e., completely independent of lexical processing) is executed. This all-or-none process (i.e., fully process the word before making a saccade or don't pay any attention to lexical processing) contrasts with E-Z Reader, in which the signal to make the saccade is partial lexical processing of the attended word.

Engbert and Kliegl's (2001) three-state model was first fitted to the same sentences (taken from Schilling et al. 1998) that were used to evaluate E-Z Reader. The model successfully predicted the mean fixation durations and skipping rates for the five frequency classes of words, and in so doing demonstrated that the state transitions can in fact be described using different distributional assumptions (i.e., residence-time dependent probabilities). Because these residence-time dependent probabilities can be implemented as an exact algorithm, whereas sampling from gamma distributions cannot, the model advances our understanding of eye-movement control by providing something like a process model of where the variability is coming from. The introduction of autonomous saccades in the three-state model marginally improved the ability of the model to fit frequency effects on both gaze durations and probability of word skipping. It also allows the model to predict (at least qualitatively) other phenomena that E-Z Reader can predict, such as spillover effects and word-frequency effects on preview benefit. However, it is by no means clear that this improvement can be taken as evidence for the existence of autonomous saccades during reading (as Engbert & Kliegl claim) because our model predicts the same phenomena by positing two stages of lexical processing.

4. Alternative models of eye-movement control

Models of eye-movement control during reading can be compared and contrasted along any number of different dimensions. Historically, the models have most often been classified as being either *oculomotor* or *cognitive/processing;* that is, with respect to whether or not language processing plays a prominent role in guiding the eyes during reading (Reilly & O'Regan 1998). Proponents of the oculomotor models claim that properties of the text (e.g., word length) and operating characteristics of the visual (e.g., acuity) and oculomotor systems (e.g., saccade accuracy) largely determine fixation locations. An auxiliary assumption of this view is that fixation durations are determined largely by where in a word the eyes have fixated. In contrast, proponents of the processing models tend to emphasize the role of language processing in guiding eye movements during reading. According to this view, the decision about how long to fixate is determined by ongoing linguistic processing, whereas the decision about where to fixate is jointly decided by linguistic, visual, and oculomotor factors. Although these two views of eye-movement control in reading have often been treated as completely distinct theoretical "camps," the distinction is one of degree because the actual models vary considerably with respect to how central a role linguistic processing plays in determining the moment-to-moment movements of the eyes through the text.

This fact has been acknowledged in more recent papers. Engbert et al. (2002), for example, have also categorized the existing oculomotor models with respect to their assumptions regarding attention. According to this taxonomy, the models near the cognitive end of the continuum can be further divided into those that assume the serial allocation of attention (i.e., *sequential attention shift*), and those that posit an attention gradient (i.e., *guidance by attentional gradient*). In the sequential-attention-shift models, attention is allocated serially, from one word to the next, whereas in the guidance-by-attentional-gradient models, attention is a gradient, so that more that one word can be attended to (and processed) in parallel. Because the question of how attention is allocated during reading is quite contentious (see Henderson & Ferreira 1990; Inhoff et al. 2000a; 2000b; Kennedy 2000; Murray 1998; Rayner et al. 1998), the models will undoubtedly play a prominent role in guiding future research in an effort to resolve this issue. (How this issue is resolved will also have important ramifications for the models.) Consequently, in the following review, we shall use both of these dimensions in describing existing models of eye-movement control during reading. We shall also use the oculomotor-cognitive dimension to organize our discussion, starting with those models that assign the least significance to linguistic processing.

4.1. Minimal control

In this model, neither fixation durations nor saccade lengths are affected by linguistic or cognitive factors, but are instead affected only by the physical layout of the text (Suppes 1990; 1994). The model consists of a small number of axioms that describe the fixation-duration distributions and a random-walk process that determines where the eyes will move next.

The axioms describing fixation durations are as follows: First, the duration of each fixation is a function of the number of operations (which are never specified) that must be completed during each fixation. Second, the fixation durations are stochastically determined by sampling from an exponential distribution if a single operation must be completed; in cases requiring two operations, the durations are described by the convolution of two independent exponential distributions. Finally, the fixation times are independent of both earlier processing and the current text content. Thus, the model stipulates that variability in fixation durations is not due to variability in the duration of the underlying cognitive processing, but instead reflects the probabilistic nature of the processing.

Saccades are determined by a similar set of rules. First, if the processing within a "region of regard" (which is defined – in the case of reading – by a given word) completes, then the eyes are moved to the next word; otherwise, they remain in the same location. Second, if processing has not finished and the memory for a prior region of regard has decayed, then the eyes are moved back to that prior region. Third, if perceptual processing of the upcoming word has finished from the current location, then the upcoming word is skipped. Finally, the length of each saccade is independent of both earlier processing and the length of prior saccades. (Thus, cognitive processing is posited to affect the locations of fixations.)

Unfortunately, the minimal-control model has only been used to simulate eye movements during an arithmetic task (Suppes 1990; Suppes et al. 1982; 1983), so that it is difficult to evaluate its adequacy with respect to reading. It is clear, however, that the model only makes predictions on the level of individual words, and therefore cannot account for either landing site distributions (McConkie et al. 1988) or the optimal viewing position effects (O'Regan 1990). The model also fails to account for many other factors that are acknowledged by Suppes (1994) to affect eye movements during reading.

4.2. Strategy tactics

This model originated from two observations: First, words are identified most rapidly if they are fixated slightly to the left of center, on the *optimal viewing position;* and second, words are also less likely to be refixated if they are initially viewed from this position (O'Regan 1990; 1992b; O'Regan & Lévy-Schoen 1987). These results led O'Regan to suggest that readers adopt a "strategy" of directing their eyes from word to word in an attempt to fixate each word's optimal viewing position. This reading strategy is "risky" because the saccades often miss their intended targets, so that the words are sometimes viewed from sub-optimal locations. To compensate for this, the reader can also use a "careful" variant of the strategy that includes the following within-word "tactic": If the eyes do not land near the optimal viewing position, then immediately move them to the other end of the word. Using this tactic ensures that every word will either be viewed from its optimal position (in the case of single fixations) or will be viewed from two different locations (in the case of refixations).

Because the within-word tactics are guided by visual factors (e.g., word length), the model predicts that linguistic variables (e.g., word frequency): (1) should only modulate fixation durations when there is a single long fixation or when the fixation is the second of two, and (2) should not modulate refixation probabilities. Unfortunately for the strategy-tactics model, neither of these predictions has been confirmed. Rayner et al. (1996) found that word frequency effects were evident in the first of two fixations (see also Sereno 1992), and that refixations were more likely on low-frequency words than on high-frequency words (with length controlled). In addition, Rayner et al. found that neither fixation durations nor frequency effects on single-fixations varied as a function of landing position,[13] which suggests that the optimal viewing position may be much less important in normal reading than in the identification of single words when they are presented in isolation (see also Vitu et al. 1990). It is worth noting that our current conjecture about refixations (see Equation 8) is similar to that of the "careful" strategy; both assume that the reason for moving the eyes to a second location within a given word is that it affords the reader a better view from which to identify the word.

4.3. Word targeting

This theory was largely motivated by the seminal work of McConkie and his colleagues (McConkie et al. 1988; 1989; 1991; Radach & McConkie 1998). As mentioned previously, they expanded upon the observation that readers typically fixate the preferred viewing location (Rayner 1979), and found that landing site distributions behaved systematically with respect to both saccade length and the launch site fixation duration. These findings led McConkie and his colleagues to conclude the following: First, the landing site distributions (which resembled truncated Gaussian distributions; see Fig. 2) reflect random noise in the oculomotor system, with the missing tails being due to cases in which

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

the eyes undershot or overshot their intended targets. The oculomotor system is also assumed to be "tuned" to make saccades approximately seven character spaces in length, so that longer saccades tend to undershoot their targets, while shorter saccades tend to overshoot their targets. This systematic range error causes the distributions to shift towards the beginnings of words as the launch site becomes more distant from the intended saccade target. With longer launch site fixations, however, the eye-movement system has more time to plan its saccades, which results in more accurate saccades and a reduction in the systematic range error.

The relationships among saccade length, the duration of the launch site fixation, and saccadic accuracy led to the development of precise mathematical descriptions of how these variables affect the landing site distributions during reading (McConkie et al. 1994). Although there have also been attempts to provide similar mathematical descriptions of fixation durations (McConkie et al. 1994; McConkie & Dyre 2000; see also Brysbaert & Vitu 1998), these accounts are little more than precise descriptions of the data, and do not attempt to explain how linguistic processing affects fixation durations during reading. Also, because these descriptions address the "where?" and "when?" questions of eye-movement control independently, they fail to explain why the durations of fixations are related to their spatial locations.

Recently, however, several word-targeting strategies were implemented as computer simulations (Reilly & O'Regan 1998) so that several theoretical assumptions about eye-movement control could be evaluated with respect to how well they handle the findings related to landing-site distributions (McConkie et al. 1988). These simulations included several alternative strategies, including three that might be classified as oculomotor (e.g., *word-by-word, target long words,* and *skip short words*) and at least one in which language processing is important (e.g., *skip high-frequency words*). The results of these simulations indicated that the target-long-words strategy fit the landing-site distributions better than the other strategies, while the language-based strategies fared rather poorly overall. On this basis, Reilly and O'Regan suggested that language-processing models do not provide an adequate account of eye-movement control during reading. As we demonstrated earlier, however, processing models (e.g., E-Z Reader) can generate reasonable-looking landing-site distributions (see Reichle et al. 1999). Our model's successes here are largely due to the fact that McConkie et al. (1988; 1989; 1991) provided such a precise explanation of how visual and oculomotor variables affect eye movements, and that incorporating such an eye-guidance mechanism into our model is fully compatible with our model's other language processing assumptions.

4.4. Push-Pull

Yang and McConkie (2001) recently applied the core assumptions of the Push-Pull theory of saccade generation (Findlay & Walker 1999) to the domain of reading. The name of this model originates from the hypothesis that the timing of saccades is determined by the outcome of competitive ("push-pull") operations that occur among various components of the oculomotor system. These operations are necessary to resolve the ever-present conflict of whether to keep the eyes stationary (i.e., to fixate) or move the eyes to a new location (i.e., to make a saccade). Thus, the key assumption of this model is that the timing of saccades is largely independent of lexical processing (with the exception that processing difficulty can inhibit the oculomotor system from initiating a program). At present, however, the model has not been implemented within a computational framework, so it is difficult to evaluate how well it accounts for the various reading phenomena that have been described in this paper.

4.5. SWIFT

Many of the ideas of the Push-Pull model have been instantiated in the SWIFT (Saccade-generation With Inhibition by Foveal Targets) model (Engbert et al. 2002; Kliegl & Engbert 2003). The model's architecture is shown in Figure 9. If one compares Figure 9 to Figure 3, it is evident that SWIFT and E-Z Reader share several key assumptions: In both models, words are identified in two stages and saccadic programming is completed in two stages. In contrast to E-Z Reader, however, SWIFT assumes that lexical processing is distributed over a four-word attentional gradient (i.e., SWIFT is a guidance-by-attentional-gradient model). Another important difference between the two models is that saccadic programs in SWIFT are initiated autonomously, after a variable (random) time interval, unless this interval is extended because the word being fixated is difficult to process. In contrast to E-Z Reader, therefore, lexical processing in SWIFT is not the engine driving eye movements during reading; instead, saccades are initiated so as to maintain a preferred mean rate of eye movements.

During the first stage of lexical processing, the lexical activity of word_n at time t, $a_n(t)$, increases (i.e., $da_n/dt > 0$) until it reaches some maximum value, L_n. During the second stage of lexical processing, $a_n(t)$ decreases (i.e., $da_n/dt < 0$) until it equals zero. L_n is a function of the word's normative frequency of occurrence in text and its predictability in the local sentence context, as given by:

$$L_n = (1 - \text{predictability}_n)\,[\alpha - \beta \log(\text{frequency}_n)] \tag{8}$$

In Equation 8, α and β are free parameters that modulate the effect of word frequency. (Note the similarity between Equation 8 and the equations that determine lexical processing times in E-Z Reader: Equations 2 and 3.) The lexical activity of word_n reaches its maximum at time t_p. The rate at which a_n approaches L_n is given by:

$$\frac{d\,a_n(t)}{d\,t} = \begin{cases} +f\lambda_k t & \text{if } t < t_p \\ -\lambda_k t & \text{if } t \geq t_p \end{cases} \tag{9}$$

In Equation 9, f and λ_k are parameters that control the rate at which a_n approaches L_n. The parameter f increases the rate of the first stage of lexical processing (relative to the second) so that it is completes more rapidly, and the λ_k parameter adjusts the rate of lexical processing as a function of the distance between the word and the fovea (i.e., the point of fixation). The parameter λ_k has four values (as indexed by the k subscript): One for each of the four words in the attentional gradient. Thus, the word being fixated (word_n) is processed most rapidly, word_{n-1} and word_{n+1} are processed less rapidly, and word_{n+2} is processed least rapidly (i.e., $\lambda_n > \lambda_{n-1} = \lambda_{n+1} > \lambda_{n+2}$). This asymmetry in the attentional gradient reflects the well-known fact that, for readers of English, the perceptual span extends further to the right of fixation (McConkie & Rayner 1975; Rayner 1986; Rayner & Bertera 1979; Rayner et al. 1982).[14]

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

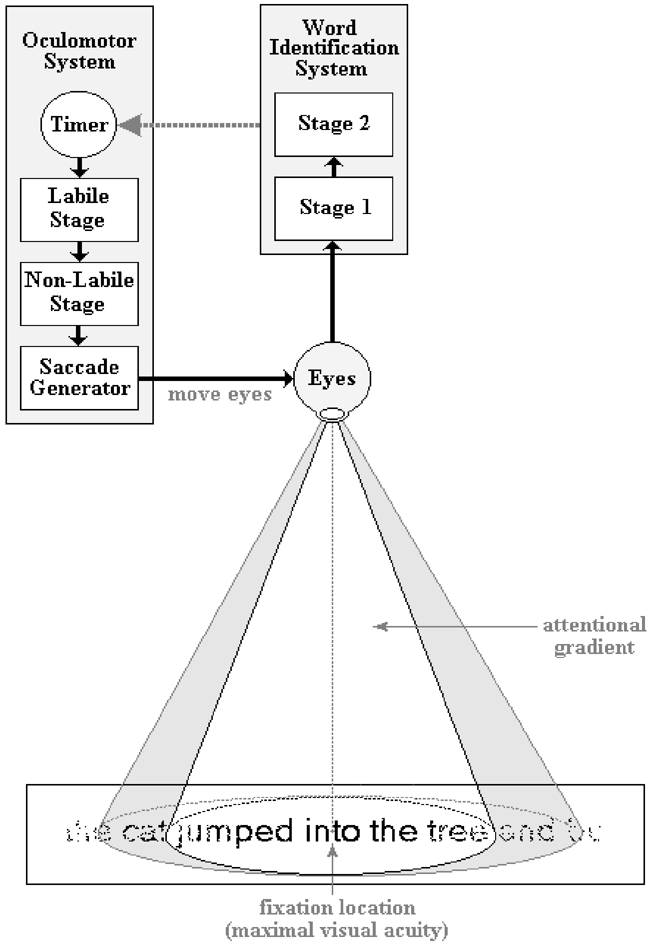

Figure 9. A schematic diagram of the *SWIFT* model (Engbert et al. 2002; Kliegl & Engbert 2002). Lexical processing occurs within a four-word attentional gradient. Saccadic programs are initiated autonomously, by a timing mechanism, so as to maintain a mean rate of eye movements. The dashed gray arrow represents the inhibitory link between the fovea and the oculomotor system. This inhibitory link allows word identification to extend the duration of the current fixation (via increasing the duration of the time interval between saccades) if the word being fixated is difficult to process.

In SWIFT, eye movements are directed towards words that have received intermediate amounts of lexical processing. The conditional probability of a saccade being directed towards word_k at time t if the eyes are currently on word_n is given by Equation 10. In this equation, the subscript m indexes word position within the attentional gradient, which extends two words to the right of the currently fixated word (i.e., word_n). If $\Sigma_m a_m(t) = 0$, then the eyes are directed towards the next word immediately to the right of the attentional gradient that has not been completely processed.

$$\Pr(k, t\_n) = \begin{cases} a_k(t) / \Sigma_m a_m(t) & \text{if } k \leq n+2 \\ 0 & \text{if } k > n+2 \end{cases} \qquad (10)$$

As already mentioned, saccadic programs are initiated so as to maintain a mean rate of eye movements. Saccadic programs are initiated after a random interval, *t*, that is given by Equation 11. In Equation 11, t_s is a random time interval (the value of which is determined by sampling from a gamma distribution) and h is a free parameter that lengthens t_s by an amount proportional to the lexical activity of word_n. The intuition behind Equation 11 is that the model's tendency to relentlessly drive the eyes forward will be held in check if the word identification system is experiencing difficulty processing the word that is currently being fixated. Two points about Equation 11 are noteworthy: First, this inhibition by foveal targets is necessary for the model to account for the frequency effects that are typically observed on first-fixation durations. (The model presumably predicts frequency effects on the other word-based measures because, in natural text, word frequency is negatively correlated with word length, so that longer words tend to be fixated more often – purely by chance – than shorter words.) Second, although this inhibition is necessary to produce normal word frequency effects, it is operational only approximately 15% of the time.

$$t = t_s + h\, a_n \qquad (11)$$

Finally, the initiation of saccadic programs in SWIFT is separated from the selection of saccade targets. Thus, the target of an upcoming saccade is not selected as soon as the program is initiated; instead, there is a lag, so that there is little "cost" in terms of re-programming time if the labile program has to be cancelled. This assumption provides a means of avoiding the problem associated with earlier versions of our model (e.g., E-Z Readers 5 and 6); namely, that our model predicted costs due to skipping that were too large.

SWIFT was applied to the same corpus used to evaluate

E-Z Reader (i.e., the Schilling et al. 1998 sentences). Like our model, SWIFT successfully predicted the mean values for each of the word-based measures. (Engbert et al. [2002] have not, however, examined the predicted distributions.) Although Engbert et al. did not examine their model's performance on the Schilling et al. high- and low-frequency target words, the model would undoubtedly handle the frequency effects on these specific items, too. Furthermore, in contrast to earlier versions of our model (E-Z Readers 5 and 6) but not to the current version, SWIFT predicts costs for skipping upcoming words that are concordant in size with those that have been reported in the literature. As Engbert et al. indicate, this aspect of the model's performance stems from the fact that the timing of the saccadic programs is decoupled from their target selection. This distinction between the two models has been blurred, however, because of our assumption in the current version of the E-Z Reader model that target selection occurs during the later half of the labile saccadic programming stage.

Kliegl and Engbert (2002) have recently examined SWIFT's capacity to simulate the results of a gaze-contingent display experiment reported by Binder et al. (1999) in which parafoveal preview of specific target words was either allowed or denied. The model successfully captured the pattern of effects observed in this experiment: In the absence of parafoveal preview, the target words tended to be fixated longer, skipped less often, and be the recipients of more regressions.

In the final analysis, we agree with Engbert et al. that SWIFT provides a viable alternative – at least as measured with respect to the model's capacity to handle a wide array of phenomena – to the current sequential-attention-shift models, including E-Z Reader. Although the model has not yet been fitted to the landing site distribution data reported by McConkie and his colleagues (McConkie et al. 1988; 1991), we acknowledge that the model could probably account for these effects if it were augmented with assumptions similar to those used by E-Z Reader (i.e., Equations 4 to 6). Nevertheless, we strongly believe that the remaining differences between the two models are far from being merely cosmetic. To reiterate, in E-Z Reader, attention is allocated serially, from one word to the next, with word identification being the "engine" driving the eyes forward. In stark contrast to this, in SWIFT, attention is allocated in parallel, to several (four) words within an attentional "window," with the tempo of the eye movements being largely independent of the moment-to-moment lexical processing (with the only exception being due to the occasional delays in the initiation of saccadic programs due to foveal inhibition by difficult words). We suspect that, in the future, the relative merits of the two sets of assumptions will be measured with respect to how well they handle the many effects of linguistic variables that have been documented in the reading literature (see Rayner 1998). For reasons that we have discussed elsewhere (Pollatsek & Rayner 1999), we believe that the ability to explain such effects will ultimately support our claim that the intrinsic nature of language processing during reading hinges upon word identification: (1) proceeding in a serial fashion, and (2) being the primary determinant of when the eyes move.

4.6. Glenmore

Yet another model inspired by Findlay and Walker (1999) is the Glenmore model[15] of Reilly and Radach (2003). The model's architecture is depicted in Figure 10. As is evident in the figure, Glenmore is a connectionist model (cf. McClelland & Rumelhart 1986; Rumelhart et al. 1986b) that consists of three major components: (1) a saliency map that selects the saccade targets; (2) an interactive-activation network that identifies words; and (3) a saccade generator that initiates and executes eye movements.

Like both the Push-Pull model (Yang & McConkie 2001) and SWIFT model (Engbert et al. 2002), lexical processing is distributed across a gradient. Letter presence/absence is encoded across a series of 30 letter-sized input units, each of which corresponds to a unique spatial location in the visual array. The activation of these units is scaled so that it decreases for units that are farther away from unit 11 (which, in the model, is the center of the fovea). The scaling is done using a gamma distribution function with a mean centered on unit 11, as described by Equation 12. In this equation, i is the position of the input unit, and μ_G and σ_G are parameters that specify the mean and standard deviation of the distribution, respectively. The scaled input unit activation is then propagated (via direct one-to-one connections) to both the letter units of the word-identification system and units of the saccade target saliency map.

$$\text{activation(i)} = \text{Gamma}(\text{i}, \mu_G, \sigma_G) \tag{12}$$

Each letter unit receives activation from (and sends activation to) the word units, so that a given letter sequence can be mapped onto its corresponding lexical representation. The model thus incorporates many of the basic processing principles of the classic *Interactive-Activation Model* of word-identification (McClelland & Rumelhart 1981), such as top-down modulation of letter activation and a "winner-take-all" competition among word units. Letter units also send activation to the saliency units, which also receive activation from the input units. The saliency units form a map, with each unit corresponding to one of the 30 locations specified by the input units. This saliency map is used to select the targets of upcoming saccades; the unit that is most active will be the target of any saccade that is executed.

Activation is propagated to the letter and saliency units in standard fashion; the input to each unit i at time t is given by Equation 13, in which $o_{j,t}$ is the activation that is being propagated to unit i from unit j, and $w_{i,j}$ is the connection weight between unit i and unit j.

$$\text{input}_{i,t} = \text{input}_{i,t-1} + \Sigma_j \, w_{i,j} \, o_{j,t} \tag{13}$$

The accumulation of activation within these two types of units is described by a Gaussian probability density transfer function; that is, the units accumulate activation over time as described by Equation 14. Here, input_i is the net input to unit i (as given by Equation 13), and μ_N and σ_N are parameters that specify the mean and standard deviation of the distribution, respectively.

$$\text{activation}(\text{input}_i, \mu_N, \sigma_N) = \frac{1}{\sqrt{(2\pi\sigma_N)}} e^{-(\text{input}_i - \mu_N)^2/(2\sigma_N^2)} \tag{14}$$

The activation described by Equation 14 is then propagated to the word units using Equation 15. In Equation 15, $L_{j,t}$ is activation from letter unit j (which is divided by word length, n, to nullify the effect of this variable), $W^R_{i,t}$ is the activation from a word unit i to itself (via recurrent connections), and $W^O_{k,t}$ is activation from other word units (via inhibitory connections).

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

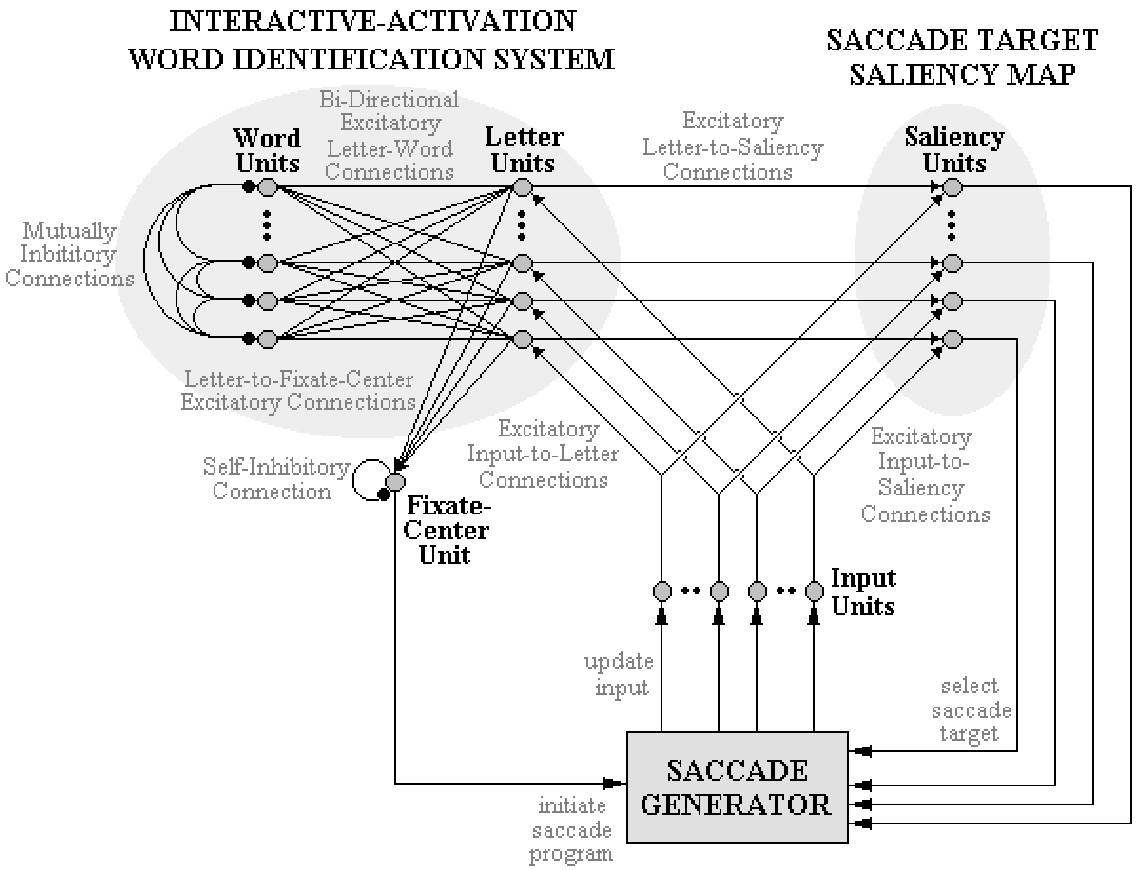

Figure 10. A schematic diagram of the *Glenmore* model (Reilly & Radach 2003). The model has a connectionist architecture and is comprised of three main components: (1) an interactive-activation network that is responsible for identifying words; (2) a saliency map that selects saccade targets; and (3) the saccade generator. Activation of the input units is propagated forward to the letter and saliency units so as to identity and localize the individual letters in the 30-unit input array. Letter activation is then spread to the word units (which provide top-down modulation of the letter units), the saliency units, and a fixate-center unit. A saccade is initiated to the target location that corresponds to the most active saliency unit whenever the activation of the fixate-center unit falls below a certain threshold.

$$\text{input}_{i,t} = \text{input}_{i,t-1} + \frac{\Sigma_j w_{i,j} L_{j,t}}{n} + \Sigma_i w_{i,i} W^R{}_{i,t} - \Sigma_k w_{i,k} W^O{}_{k,t} \quad (15)$$

Word unit activation is accumulated using a sigmoid transfer function, so that the activation of unit *i* is given by Equation 16. Activation therefore ranges continuously over the range 0 to 1 and is equal to 0.5 when the net input (given by Equation 15) equals the free parameter α. The other free parameter, β, controls the steepness of the function, or the rate at which activation goes from zero to one as the net input increases. The role of the word units is to support the letters of words that are presented as visual input. This is critical because the letter units also propagate activation to the fixate-center unit, which is responsible for initiating saccades. When the activation of the fixate-center unit falls below a certain threshold, it signals the saccadic generator to move the eyes to the location specified by the saliency map.

$$\text{activation(i)} = 1 / \{1 + e^{-[(\text{input i} - \alpha)/\beta]}\} \quad (16)$$

The saccades that are generated by Glenmore are subject to both systematic and random error. The landing site distribution mean, μ, is centered (i.e., is equal to zero) on the target word and deviates from the target as described by Equation 17. Likewise, the standard deviation of the landing site, σ, also varies as a function of saccade length, as described by Equation 18. In these equations, the slope (b_1 and b_2) and intercept (m_1 and m_2) parameters modulate the effect of saccade length.

$$\mu = b_1 + m_1 \text{ (saccade length)} \quad (17)$$

$$\sigma = b_2 + m_2 \text{ (saccade length}^3) \quad (18)$$

Finally, each landing site, *x,* is a random deviate that is independently sampled from a Gaussian distribution defined by Equation 19, with μ and σ being defined by Equations 17 and 18, respectively.

$$f(x,\mu,\sigma) = \frac{1}{\sqrt{(2\pi\sigma)}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (19)$$

The Glenmore model has been successfully applied to wide range of eye-movement phenomena. However, instead of fitting their model to a sentence corpus (as we and others have done with the Schilling et al. 1998 sentences), Reilly and Radach (2003) have demonstrated their model's competence by running simulations in which they illustrate key properties of its performance. So far, they have shown that Glenmore successfully predicts many of the findings simulated by our model, including word-frequency effects, spillover effects, and preview effects that are modulated by

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

the difficulty of the fixated word. Moreover, although they did not provide evidence that the model reproduces the types of landing site distributions observed by McConkie et al. (1988; 1991), the model has clearly been designed to account for such effects (see Equations 17–19). Likewise, it remains an open question as to whether the model can predict the costs that have been observed for skipping. Based on these results, therefore, we think that the Glenmore model is very promising, and that – again, if one only considers the model's performance – it provides a viable alternative to the cognitive-based, serial attention models (like E-Z Reader). However, we also believe that the model may be inherently limited in that it makes no provisions for explaining how linguistic variables affect eye movements during reading. As it is currently implemented, for example, the Glenmore model cannot handle predictability effects. We suspect that, given the model's core asumptions (e.g., the gradient of lexical processing), many of the well-documented effects of linguistic processing (see Rayner 1998) may prove to be even more challenging for the model.

4.7. Mr. Chips

This model was proposed as a means to evaluate how an *ideal-observer* (i.e., a reader with perfect lexical knowledge and the well-specified goal of maximizing reading speed) would move his/her eyes (Klitz et al. 2000; Legge et al. 1997). Consequently, the model exemplifies a very different approach to understanding the interrelationships among visual processing, word recognition, and eye-movement control during reading. The model does this using three pieces of information: (1) input from a "retina" that encodes a small number of letters in the fovea and indicates whether letters in the parafovea/periphery are present or absent; (2) knowledge about the relative frequencies with which words occur in text; and (3) knowledge of the likelihood of making a saccadic error of a given size for each given saccade length. These three types of knowledge are depicted in Figure 11.

The Mr. Chips model attempts to use all of the above information that is available from a particular fixation location to identify the next word in text using the fewest saccades possible. To do this, the model calculates the expected uncertainty that is associated with being able to identify a word for saccades of each possible length. It then executes a saccade that minimizes this uncertainty. For example, imagine that the model has the following information about a word: It is five letters long and begins with "abo" (see Fig. 11). The model uses this information in conjunction with its lexical knowledge to calculate conditional probabilities of the letter string being each of the words that satisfy these constraints, using Equation 20:

$$p_i = P_i / \Sigma_j P_j \qquad (20)$$

In Equation 20, p_i is the conditional probability of the letter string being word_i, given the letter information already known ("abo" in the example); P_i is the absolute probability of the letter string being word_i; and the P_js are the absolute probabilities of the letter strings in the "candidate set" (in the example, all of the 5-letter word beginning with "abo"). In the Figure 11 example, the conditional probability that "abo—" is "about" is equal to 0.849.

The conditional probabilities are then used to compute the conditional entropy, or degree of uncertainty, H, that would result from a saccade of length, L, under the assumption that the letter string is word_i using Equation 21. For example, from the current fixation, the entropy associated with the letter string is: $H(0, \text{abo—}) = 0.613$. (Smaller entropy values represent less uncertainty about the identity of a word, so that identification occurs with certainty when the entropy value associated with a letter string equals zero.) A saccade of $L = 1$ would reveal one letter, which, given the model's lexical knowledge, must be either "u" or "v." If the letter is "u," then the conditional probability of the word being "about" is $p = 1$, and the conditional en-

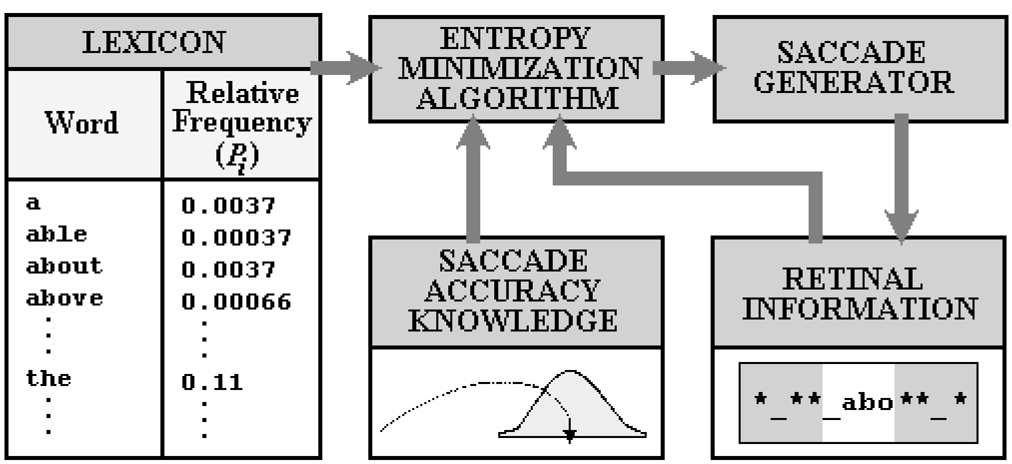

Figure 11. A schematic diagram of the *Mr. Chips* model (Klitz et al. 2000; Legge et al. 1997). The model attempts to compute the saccade length that will minimize the uncertainty about the identity of next unidentified word. It does this using three sources of information: (1) the relative frequencies with which the words in its lexicon occur in text; (2) the accuracy of saccades for each possible saccade length; and (3) visual information from the model's "retina." Visual information is encoded from two regions in the retina: a fovea, in which letters can be identified, and a parafovea, in which letters can be discriminated from blank spaces. (In the Figure, the retina is presented by a rectangle, with the white and gray areas corresponding to the fovea and parafovea, respectively.) The entropy-minimization algorithm computes the saccade length that will minimize the uncertainty of the next unidentified word, and then an error-prone "Saccade Generator" executes the saccade so that the retina can encode additional letter information.

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

tropy would be reduced to: $H(1, \text{about}) = 0$. Likewise, if the letter is "v," then the conditional entropy is reduced to: $H(1, \text{above}) = 0$.

$$H(L, \text{word}_i) = -\Sigma_i \, p_i \log_2 (p_i) \tag{21}$$

After the conditional entropies are calculated for each possible saccade length, Mr. Chips computes a probability-weighted average to determine the expected entropy associated with a saccade of each given length. This is done using Equation 22. In the example, $H(L) = 0$ for saccades of lengths 1 to 5. Because of saccadic error, however, each saccade of intended length, L, has an associated landing-site distribution, $P_L(x)$, which determines the probability of making a saccade of actual length, x. The model uses this knowledge to calculate the entropy associated with each saccade length, L, averaged across all of the possible landing sites. Equation 23 gives the expected uncertainty, H_L, associated with making a saccade of intended length L. Finally, the model makes the saccade that minimizes H_L, and thereby maximizes the probability of identifying the word. In cases where more than one possible saccade yields the same expected entropy, Mr. Chips executes the longest saccade possible so as to maximize reading speed.

$$H(L) = \Sigma_i \, p_i \, H(L, \text{word}_i) \tag{22}$$

$$H_L = \Sigma_i \, P_L \, (x) \, H(x) \tag{23}$$

Because Mr. Chips was developed with the intent of examining the way lexical knowledge and restrictions on visual encoding affect saccade lengths and fixation locations, the model does not address the "when?" question of eye-movement control. Several of the model's emergent properties, however, are consistent with research findings about where the eyes move. For example, the model predicts that the mean saccade length will be around seven character spaces (McConkie et al. 1988) and that saccades will tend to be directed towards the optimal viewing position (O'Regan 1990). The model also predicts parafoveal preview effects because the left-most letters of upcoming words are often identified before the words are actually fixated.

Unfortunately, it does not seem plausible that human readers compute the expected amount of information to be gained from each possible saccade length so as to make the saccade that maximizes this gain. Klitz et al. (2000) acknowledge this fact, and say that their model "is not intended as an exact model of how humans perform a task, but rather establishes an upper bound (i.e., a level of competence) for human performance." Furthermore, the Mr. Chips algorithm is well approximated by the simple heuristic of left-justifying the target word in the high-resolution part of vision, so that, on some level, the model is psychologically plausible.

Moreover, it is important to point out that Mr. Chips, unlike the other models discussed in this article, was developed to investigate how visual impairment might affect eye movements during reading. In this capacity, the model has been successful (Klitz et al. 2000). A comparison of the model's performance to that of a human in a reading task[16] with a simulated *scotoma* (i.e., a blind spot in the visual field) indicated that, in contrast to the model, the human had difficulty integrating information across central scotomas more than a single character-space in size. The human reader appeared to primarily use visual information from one side of the scotoma and to use the visual strategy of moving the eyes in order to place the region of normal vision over all of the character spaces in turn, rather than using lexical knowledge to winnow down the possible identities of letter strings from a single fixation. Although the human reader's natural strategy produced shorter saccades, it markedly increased reading speed over when they tried to execute the Mr. Chips strategy. These analyses, therefore, suggest that, while the seemingly erratic eye movements of readers with scotomas do not allow the maximal amount of information to be extracted from the page during each fixation, they are nevertheless adaptive in that they allow a maximal overall rate of information extraction.

4.8. Attention shift

In the attention-shift model (or ASM), linguistic processing and eye-movement control are loosely coupled (Reilly 1993). As Figure 12 indicates, the model's architecture consists of pair of interacting connectionist networks that are trained using the back-propagation learning algorithm (Rumelhart et al. 1986a). One of these networks is responsible for word identification; the other is responsible for programming saccades. As each word is identified, the lexical-encoding network signals attention to shift to the next word, so that it can be processed. The movement of attention, in turn, causes the saccadic-programming network to begin programming a saccade to the next word. In contrast to E-Z Reader, the ASM does not allocate attention serially, from one word to the next. The attention "spotlight" is instead fixed in size, so that whatever falls within the spotlight will be the focus of attention. This means that, in cases where two or more short words follow in immediate succession, they both may be in the spotlight and can be encoded on a given fixation. The ASM is therefore a guidance-by-attentional-gradient model.

In the ASM, the times needed to complete both lexical access and saccadic programming are determined by the number of cycles that the two networks require to settle into stable activation patterns. As in E-Z Reader, the visual input to the word identification system is affected by retinal acuity limitations. Thus, the activation patterns that represent letter features become more "degraded" (i.e., the activation values of the units representing the letters decrease and are more prone to noise) as they are encoded further from the fovea, especially for letters that share many features with other letters. This degradation allows the model to account for the finding that word identification becomes more difficult as the distance between the word and the fovea increases (Morrison & Rayner 1981).

Although Reilly (1993) does not provide a detailed account of his model's performance, the ASM does simulate a few of the basic phenomena related to eye-movement control in reading. For instance, the model generates mean fixation durations and saccade lengths that are in close agreement to values that have been reported in the literature. In contrast to E-Z Reader, however, the ASM has not fitted to the various word-based measures, nor has it been shown to generate means and distributions for the different frequency classes of words. Nonetheless, because the amount of training that the word-recognition module receives on each word is proportional to each word's frequency of occurrence, the model does predict that low-frequency words are fixated longer than high-frequency words. Moreover, because two successive short words are

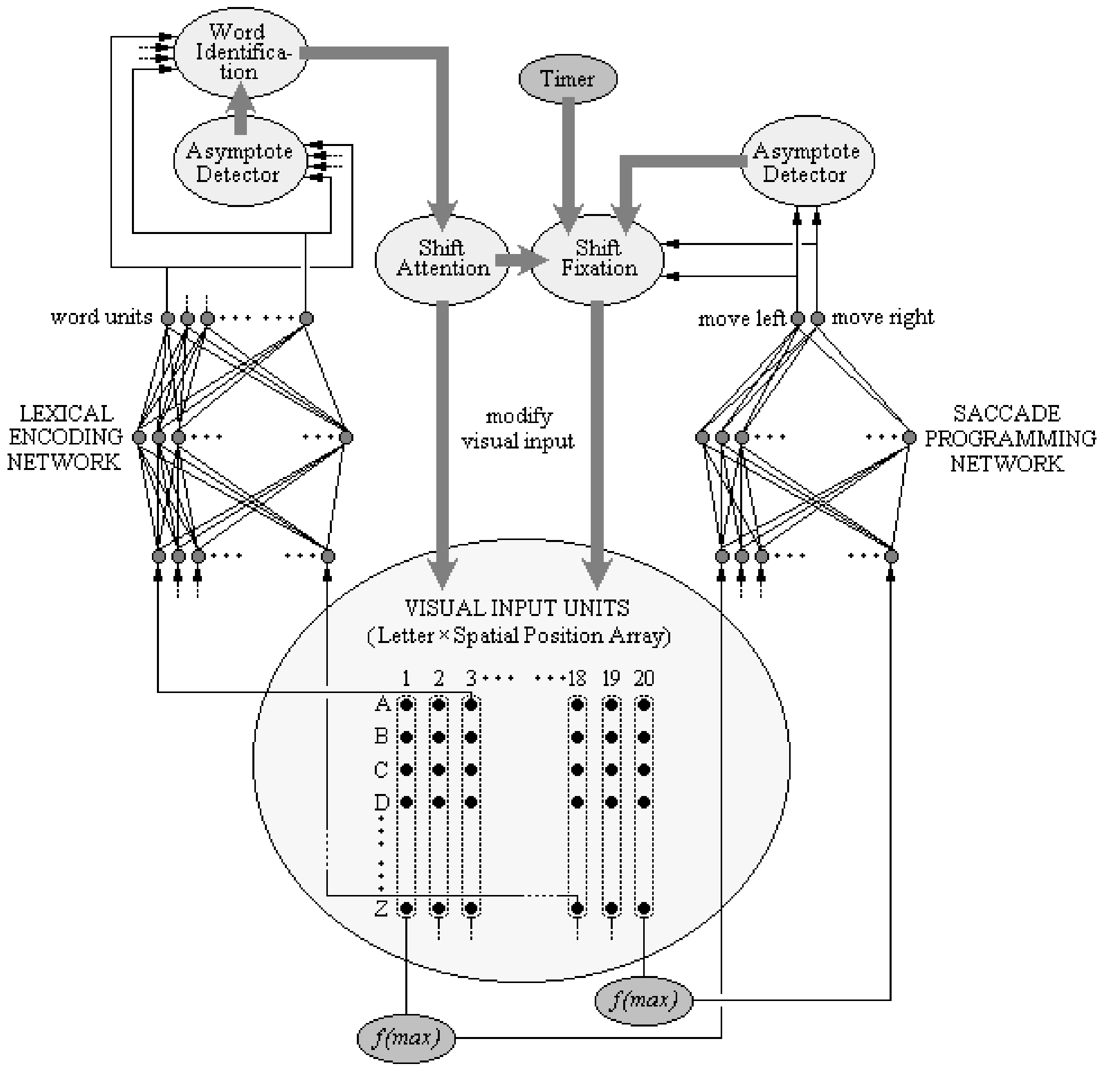

Figure 12. A schematic diagram of the *Attention-Shift* model (Reilly 1993). In the model, visual input is represented by an array of 26 letters that can be in any of 20 different spatial locations (position 8 is the center of the fovea). The core of the model consists of two connectionists networks that work in tandem to identify words and move the eyes. The first network, labeled "Lexical Encoding" in the Figure, has as its input the activation values of each letter from the 16 central spatial positions. This information is used to identify individual words, which are represented by the word units as unique 8-bit patterns. The input to the second network, labeled "Saccadic Programming" in the Figure, are the maximal values from each spatial position, which is used to compute the direction and amplitude of the saccades. The "Asymptote Detectors" determine when the networks have settled into stable activation patterns, and thus provide an index of processing time. Word identification causes attention to shift, which modifies the visual input by reducing the activation values of unattended spatial input units (this is represented by the thick arrows in the Figure). Attention shift also enable saccades, which are executed after the "Saccadic Programming" network has settled into a stable pattern or after a certain time interval (which is determined by the "Timer"). Saccades also modify the visual input by boosting the activation values of the letters in the next word.

sometimes encoded in parallel, the model is able to account for the skipping of short words, as well as parafoveal preview benefit. It is of interest, though, that the ASM does not account for either of these phenomena in the same way that E-Z Reader does. In our model, skipping occurs whenever the word being fixated is identified, attention shifts to the next word, and it too is identified (in the parafovea). Thus, the models provide quite different accounts of the same phenomena: Whereas the ASM (a guidance-by-attentional-gradient model) allows some degree of parallel processing of upcoming words, E-Z Reader (a sequential-attention-shift model) allows for parafoveal processing via covert shifts of attention. There is one noteworthy difference between the two models with respect to parafoveal processing, however: In contrast to our model, the ASM does not explain why predictable words are skipped more often than less predictable words.

Finally, like E-Z Reader, saccadic programming in ASM is prone to noise, so that individual words can be refixated and/or skipped due to simple oculomotor error. Reilly (1993) has not, however, demonstrated that the model can reproduce the complex dependencies between the locations and durations of launch sites and the landing site distributions. We therefore contend that, unlike E-Z Reader, the ASM has not – at present – provided a complete account of the visual, oculomotor, and language-processing determinants of eye-movement control in reading.

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

4.9. EMMA

Salvucci (2000; 2001) has recently extended many of the core principles in E-Z Reader to provide a general theory of the interrelationships among cognition, attention, and eye movements. This model, EMMA (Eye Movements and Movements of Attention), has been implemented within the *ACT-R/PM* production-system architecture (Anderson & Lebiere 1998; Byrne & Anderson 1998). *Productions* are procedural condition-action pairs (i.e., "if-then" statements) that perform operations on units of declarative knowledge. For example, the production:

If ($letter_1$ = "c" & $letter_2$ = "a" & $letter_3$ = "t"),
then (word = "cat")

encodes the percept "cat," so that the meaning of the word can then be retrieved from semantic memory.

In EMMA, the encoding time for both words and objects, T_{enc}, is given by Equation 24. In Equation 24, the frequency of occurrence is scaled within the range (0, 1), ε is the eccentricity of the word or object (as measured by the angular distance between it and the fovea), and K and k are free parameters which scale the encoding time and eccentricity parameter, respectively. Like E-Z Reader, EMMA is a sequential-attention-shift model. EMMA also shares the following assumptions with E-Z Reader 7. First, encoding times are a function of both normative frequency and foveal eccentricity. Second, the actual amount of time that is required to encode a given object or word is determined stochastically by sampling random values from gamma distributions having fixed means (cf. Equations 1, 2, and 3, in E-Z Reader 7, and Equation 24, in EMMA) and standard deviations. Third, saccadic programming is completed in two sequential stages (the first being subject to cancellation by subsequent programs, the second not), the durations of which are also sampled from gamma distributions having fixed means and standard deviations. Finally, although saccades are directed towards the centers of their intended targets, they often deviate from their targets because of Gaussian motor error.

$$T_{enc} = K\,[-\log(\text{frequency})]\,e^{k\varepsilon} \tag{24}$$

Although EMMA and E-Z Reader share many common assumptions, there are a few notable differences. First, in contrast to our model, encoding time in EMMA is not modulated by predictability, so that the model cannot account for predictability effects (Balota et al. 1985; Ehrlich & Rayner 1981; Rayner & Well 1986). Second, the distinction between the first and second stages of lexical processing in E-Z Reader corresponds to the encoding and cognitive-processing stages in EMMA, respectively. As cognitive processing completes, it directs the visual system to encode additional information. However, because only the rate of encoding (and not cognitive processing) is modulated by normative frequency, EMMA cannot account for the interaction between parafoveal preview benefit and foveal processing difficulty (Henderson & Ferreira 1990; Kennison & Clifton 1995; Schroyens et al. 1999). Finally, in EMMA, foveal eccentricity is measured in terms of angular disparity rather than character spaces. Although this last difference between the two models is largely cosmetic, it allows EMMA to simulate tasks other than reading.

So far, EMMA has successfully predicted the patterns of fixation durations and locations in equation solving (i.e., mental arithmetic) and visual search tasks (i.e., subjects scan visual arrays of alphanumeric characters and indicate the presence of pre-defined targets). EMMA has also been fitted to the same six word-based measures used to evaluate E-Z Reader (i.e., the mean fixation duration and fixation probability values observed in the Schilling et al. 1998 sentence corpus). In each of these tasks, the core principles governing attention and eye movements were the same in the model, and only the productions mediating the central, or cognitive, components of the tasks were changed. We view the successes of EMMA as being very encouraging because they suggest that the core principles of the model (which are shared by E-Z Reader) are general enough to describe the link between cognitive processing and eye movements in a variety of task domains. These successes also provide converging evidence supporting the validity of the basic principles shared by E-Z Reader and EMMA. However, the link between cognitive processes and eye movements might not be as tight in tasks where there are no externally composed task demands (such as scene perception).

4.10. Reader

In contrast to all of the models discussed thus far (including E-Z Reader), this model attempts to explain reading in its entirety, including the encoding of visual features, lexical processing, semantic and syntactic analysis, and the schema-guided comprehension and abstraction of key ideas that normally occur during reading (Carpenter & Just 1983; Just & Carpenter 1980; 1987; Thibadeau et al. 1982). In this model, eye movements are tightly linked to cognitive processing. This coupling is based on two assumptions. The first is the *immediacy hypothesis,* which stipulates that each word is processed to the farthest extent possible when it is fixated. The second is the *eye-mind hypothesis,* which stipulates that the eyes remain fixated on a word until the processing on that word has been completed. Both the durations and locations of individual fixations are thus determined by the immediate processing of the word that is being fixated. Thus, Reader (like our model) is clearly a sequential-attention-shift model in that attention (and in the case of Reader, all cognitive processing) is sequentially shifted from one word to the next.

Reader was implemented as a computer simulation with a production-system cognitive architecture (Anderson 1983; Anderson & Libiere 1998; Newell 1990). In Reader, the productions are activation-based; that is, they direct activation towards units of declarative knowledge. These units of declarative knowledge, in turn, have thresholds that must be exceeded if the information is to be "active" in working memory (and thereby satisfy the conditions of other productions). The values of these thresholds are adjusted to modulate the cost associated with using each production. For example, the thresholds of those productions that mediate lexical access are adjusted to reflect each word's normative frequency of occurrence, so that low-frequency words take longer to identify (and are consequently fixated longer) than high-frequency words. Also, in the most recent version of the model (Just & Carpenter 1992), the amount of activation that is available to support processing is limited (and is a free parameter) so that individual differences in working memory capacity can be used to simulate individual differences in reading ability.

The major strength of the Reader model is its compre-

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

hensiveness. As mentioned above, the model attempts to explain the entire reading process and therefore does reasonably well simulating a number of language-related reading phenomena, such as word-frequency effects, increased reading times on lexically ambiguous words, and the processing difficulties which are found with syntactically ambiguous sentences. Unfortunately, the model is extremely complex (it consists of 225 productions; Just & Carpenter 1987), and thus lacks the conciseness and controllability of other computational models (e.g., the inner workings of the model are not transparent, and can only be described verbally). It is also difficult to evaluate the model's performance because it depends upon the complex interplay of the productions, many free parameters, and the regression weights on several independent variables (e.g., whether or not a word is the first in a sentence) that are necessary to convert production cycles (arbitrary units of time) into processing time. Furthermore, the model only makes predictions about the locations of fixations at the level of individual words, using a composite measure (*gazes*) that counts skipping as 0-msec fixation durations in the average. This means that the model does not really make precise predictions about which word is fixated. In addition, apart from word-length effects, the model fails to account for any of the phenomena that are explained by the oculomotor models (e.g., landing site distributions).

In addition to the above shortcomings, Reader has been criticized because of the immediacy and eye-mind assumptions. With respect to the former, there is considerable evidence that the lexical processing of a word is often initiated before the word has been directly fixated (i.e., parafoveal preview: Balota et al. 1985; McConkie & Rayner 1975; Pollatsek et al. 1992; Rayner 1975). Furthermore, the depth of linguistic processing assumed before the eyes are allowed to move seems somewhat implausible. With respect to the eye-mind hypothesis, as we have noted a couple of times, there is evidence that the normative frequency of $word_n$ can affect how long the eyes remain on $word_{n+1}$ (Rayner & Duffy 1986; Rayner et al. 1989). These spillover effects indicate that the eyes often leave a word before the processing of that word is complete, contrary to the eye-mind assumption. Moreover, it seems quite implausible that each word can be encoded to the linguistic depth assumed in the model before an eye movement is programmed. This would produce fixation durations (and gaze durations) much longer than those usually encountered in normal reading. Thus, even if eye movements during reading are partially guided by language processing, the Reader model greatly over-simplifies how this occurs.

4.11. Comparison of the models

The processing models extend the theoretical coverage of the oculomotor models by attempting to specify how *the* key component of reading – word identification – affects (and is affected by) both the visual and oculomotor systems. This is important because a large number of linguistic variables have well-documented effects on eye movements during reading (for reviews, see Rayner 1998; Rayner & Duffy 1988; Rayner & Sereno 1994). Indeed, much of the interest surrounding the use of the eye-tracking methodology is that it affords a relatively non-intrusive, on-line way to study language processing. Of course, the processing models are not equally successful in handling the phenomena addressed by the oculomotor models. Table 1 lists the various eye-movement phenomena that have been observed during reading (as we discussed earlier in this article), and which E-Z Reader can explain. In Table 1, we have also presented for comparison a summary of the performance of the other eye-movement control models with respect to each of these phenomena. Thus, we have indicated whether or not (or the extent to which) each of the models can account for particular phenomena. A "Yes" indicates that the model can explain a result; a "No" indicates that (as the model is currently instantiated) it does not; finally, in some cases, we have indicated that the model provides a limited (labelled "Ltd") account in that the account is incomplete.

Table 1. *A comparison of the Reading Models[a,b] with respect to reading-related phenomena[c] that are explained by the E-Z Reader Model (See Table 1 Notes located at end of main Notes section.)*

	Minimal-Control	Strategy-Tactics	Word-Targeting	Push-Pull	SWIFT	Glenmore	Mr. Chips	Attention-Shift	E-Z Reader	EMMA	Reader
	Oculomotor ←			Oculomotor-Cognitive Dimension				→	Cognitive		
Reading Phenomena	POC	POC	POC	POC	GAG	GAG	GAG	GAG	SAS	SAS	SAS
Landing Site Distributions	No	Yes	Yes	Ltd	No	Yes	No	Yes	Yes	No	No
Systematic Range Error	No	Yes	Yes	No	No	Ltd	No	No	Yes	No	No
Word-Based Measures	Ltd	No	No	Ltd	Ltd	Yes	No	Ltd	Yes	Yes	Ltd
Frequency Effects	No	No	No	No	Yes	Yes	No	Ltd	Yes	Yes	Ltd
Parafoveal Preview	Ltd	No	No	No	Ltd	Yes	Ltd	Ltd	Yes	Ltd	No
Spillover Effects	No	No	No	No	No	Yes	No	No	Yes	No	No
Costs for Skipping	No	No	No	No	Yes	Yes	No	Ltd	Yes	No	No
Predictability Effects	No	No	No	No	Yes	No	No	No	Yes	No	Yes

Table 1 indicates that E-Z Reader handles the phenomena discussed in this article. Of course, one might argue that the inventory of phenomena in Table 1 is incomplete, and that there are also other ways by which to evaluate a computational model. Let us examine each of these objections in turn. First, we acknowledge that Table 1 is incomplete. For example, it does not include neighborhood effects (Perea & Pollatsek 1998; Pollatsek et al. 1999a) or lack of case change effects across fixations (McConkie & Zola 1979; Rayner et al. 1980). For E-Z Reader to be able to account for these effects, it would be necessary to extend the model to account for how letter processing maps onto word identification (which is something that we intend to do in future research). Nevertheless, the phenomena contained in Table 1 represent a substantial body of research and are not trivial to explain (as indicated by the fact that many of models have difficulty explaining a majority of them). Moreover, there is obviously some consensus that these phenomena are important "benchmarks" in that so much effort has been spent developing models to explain these phenomena. Thus, although we agree that Table 1 is not exhaustive, it does represent the basic results that any viable model of eye-movement control in reading must be able to explain.

A second criticism – that there are other ways to evaluate computational models – is more difficult to address because what constitutes a "good" model is somewhat subjective (see Hintzman 1991, for a discussion of some of the issues related to the evaluation of computational models). Rather than arguing that our model is better than another, we believe that it may be more productive simply to discuss why we think our model is a "good" model. To begin with, E-Z Reader describes and summarizes a large body of data (those in Table 1). Moreover, it does so in a relatively simple fashion. Although successive versions of the model have included additional free parameters, we have always maintained our "minimalist" approach to modelling; that is, we have added new parameters only when it was absolutely necessary (e.g., to explain some aspect of the data that could not otherwise be explained) or when it made the model more psychological or physiologically plausible.[17] Our reason for doing this is that we wanted the model to be transparent. That is, we wanted the model to be simple enough for us to understand why it worked and why – in some cases – it failed. (We believe that one of the major shortcomings of other modelling approaches, e.g., connectionism and production systems, is that the models are often too complicated to be summarized in a concise and precise manner.)

One final criterion that we use for evaluating our model is its utility as a heuristic device. That is, one measure of a model's usefulness is the degree to which it makes clear predictions that don't depend on specific settings of parameter values, but instead flow from the basic assumptions of the model. For example, prior to any attempts to fit the model, it was clear than an earlier version of our model (E-Z Reader 5; Reichle et al. 1998) predicted inflated fixation durations on word$_n$ in cases where word$_{n+1}$ is skipped. This prediction was subsequently confirmed (Pollatsek et al. 1986; Reichle et al. 1998; but see Note 4). Similarly, the model is currently being used as an analytical tool to evaluate the basic assumptions of other theories of language processing, as will be discussed in the next section of this paper. Finally, we believe that – with everything else being equal – it is better to have a model that at least has the potential to map the behavioral phenomena that are being explained onto their underlying neural processes. As the last section of this paper will indicate, we are currently striving to link the cognitive processes of E-Z Reader onto known brain structures.

On the basis of the preceding analysis, therefore, we conclude that the E-Z Reader model provides the most comprehensive and complete theory of eye-movement control in reading, while still being transparent enough that many of its qualitative properties flow from basic assumptions rather than from specific parameter values. In the final section of this article, we will briefly discuss the possible roles that E-Z Reader may play in future reading research.

5. Future research

In this section, we will focus on a few of the ways in which the E-Z Reader model may be used to guide future reading research, and, conversely, how this research may guide the development of future reading models. This discussion will focus on two main issues. First, we will briefly discuss how the model has been used as an analytical tool to examine some key assumptions about eye movements and language processing. More specifically, our discussion will focus on the ways in which the model might be used to better understand higher-level linguistic processing in the context of natural reading. Second, we will consider how recent advances in cognitive neuroscience have influenced our understanding of eye-movement control in reading, and then speculate on how our model might be viewed in light of this new information.

5.1. Language processing

The core principles of E-Z Reader have been adapted to several different task domains, which suggests that it is capturing the basic "engine" that drives eye movements in tasks like reading. However, as we have indicated above, it is incomplete, as it only takes into account certain relatively "low-level" aspects of the reading process (i.e., up to the level of lexical access). However, we are optimistic that as better quantitative descriptions of higher-order language processing are developed, additional processing modules could be interfaced with our model to expand the domain of the model. This would undoubtedly be beneficial for two reasons. First, our model could be used to help guide what to look for in the eye movement record to test theories of language processing. Second, because a large number of higher-level language processing phenomena are known to affect eye movements during reading (see Rayner 1998, Table 2), the capacity to simulate these results using language models could provide additional benchmarks for evaluating future models of eye-movement control. Two examples of this "bootstrapping" approach to understanding reading and language are discussed below.

5.1.1. Lexical ambiguity. There are now a large number of eye-movement studies (Binder & Rayner 1998; Dopkins et al. 1992; Duffy et al. 1988; Kambe et al. 2001; Rayner & Duffy 1986; Rayner & Frazier 1989; Sereno 1995; Sereno et al. 1992; Wiley & Rayner 2000) that have examined how lexically ambiguous words are processed during reading. The basic findings from this research suggest that both

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

meaning dominance (i.e., the relative frequency of the various meanings of the ambiguous word) and contextual information influence the processing of such words. For ambiguous words with two equally likely meanings (e.g., "straw"), readers' gaze durations are longer on such words in neutral, contexts than on a control word matched in length and word frequency. However, when the prior context disambiguates the meaning that should be instantiated, gaze durations are no longer on the ambiguous word than on the control word. Thus, the contextual information helps guide the reader's choice of the appropriate meaning. For ambiguous words where one meaning is much more dominant than the other (e.g., "bank"), when the prior context was neutral, readers look no longer at the ambiguous word than the control word. However, when the subsequent text in the sentence makes it clear that the subordinate meaning should be instantiated, fixation times on the disambiguating information are quite long and regressions back to the target word are frequent (suggesting that the reader incorrectly selected the dominant meaning and now has to recompute the subordinate meaning). Conversely, when the disambiguating information that precedes the biased ambiguous word indicates that the subordinate meaning is instantiated, readers' gaze durations on the ambiguous word are lengthened. Apparently, the contextual information increases the level of activation for the subordinate meaning so that the two meanings are in competition (just as the two meanings of a balanced ambiguous word like "straw" are in competition in a neutral context). This general pattern of results has been interpreted in the context of the *Reordered Access Model* (Duffy et al. 1988) and the data have been simulated using a constraint-satisfaction framework (Duffy et al. 2001).

Using the basic principles of E-Z Reader, we were able to simulate the pattern of data present in these eye-movement studies. This was done by: (1) treating the subordinate meaning of ambiguous words as if readers were dealing with a low-frequency word; and (2) allowing disambiguating context to decrease the time required to complete lexical processing of ambiguous words. Although our early efforts indicated that the model can predict the gaze duration on the ambiguous target words, we were unable to simulate an important finding; namely, that spillover fixations are much longer for ambiguous words than for words matched to the frequency of the subordinate meaning (Sereno et al. 1992). However, the important point for this discussion is that we suspect that, by implementing aspects of the Reordered Access Model into the architecture of our model, progress can be made in understanding lexical ambiguity resolution in reading.

5.1.2. Morphology. A recent survey of prominent reading researchers indicated that one of the major areas of residual ignorance in the domain of reading research concerns the role of morphology in visual word identification (Kennedy et al. 2000). In the last few years, researchers have had some success investigating the role of morphology in word identification by examining how eye movements are affected by the morphemic variables during natural reading (Andrews et al., in press; Hyönä & Pollatsek 1998; 2000; Juhasz et al. 2003; Pollatsek et al. 2000). For example, Hyönä and Pollatsek (1998) examined the eye movements of Finnish readers while reading long compound words embedded in single sentences. The data indicated, among other things, that although the whole-word frequency influenced fixation durations on the word, the frequency of the constituent words of the compounds influenced fixation durations as well. Interestingly, the effect of the frequency of the second constituent was first seen a bit later in processing than the effects of either the frequency of the first constituent or the frequency of the whole word (i.e., on the duration of the second fixation on the word instead of the duration of the first fixation of the word). These findings suggest that access of the compounds is a "race" between a direct lexical look-up process and a compositional process in which the components are assembled (a similar conclusion comes from a study of English suffixed words; Niswander et al. 2000). E-Z Reader 7, which already includes races between various components, is a natural framework to be expanded upon to explain such phenomena. However, expanding the model in this direction is not trivial, as it entails positing that units smaller than "the set of letters between the spaces" can influence the decision of when to move the eyes. Thus, among other things, one has to think carefully about which letter subsets of a word can play an active role in this decision. We are currently working on an expanded version of the model that simulates the major trends that were observed in these data (Pollatsek et al. 2003).[18]

5.1.3. Conclusion. Our discussions of lexical ambiguity and the role of morphology in word identification were meant to illustrate how our model of eye-movement control might be used to advance our understanding of language-related phenomena. These two examples were selected because researchers in both of these areas have made extensive use of data from eye-movement experiments and because explaining these phenomena clearly involved relatively small increments in the development of our model. Of course, this is not to say that eye movements have not already been used in a productive manner to address other language-related questions; on the contrary, eye movements have been used to study a wide array of linguistic phenomena, including (but not limited to) other types of ambiguity resolution (e.g., syntactic and phonological ambiguity), semantic and repetition priming, anaphor and co-reference, and discourse processing (for reviews, see Rayner 1998; Rayner & Pollatsek 1989; Rayner & Sereno 1994). We think that E-Z Reader will also prove to be a useful platform from which to model these other psycholinguistic phenomena.[19]

5.2. Cognitive neuroscience

As mentioned at the beginning of this section, the last decade has witnessed unprecedented advances in our general understanding of the mind-brain relationship. New methodologies, such as brain-imaging (e.g., PET, fMRI), electrophysiological recording (e.g., EEG), and single- and multiple-cellular recording techniques, have provided invaluable tools for examining the relationship between cognitive processes and their neural substrates. Likewise, new theoretical advances, such as those offered by biologically plausible connectionist models (Churchland & Sejnowski 1992; McClelland & Rumelhart 1986; Rumelhart et al. 1986b), promise to bridge the chasm that has until recently separated cognitive psychology from neuroscience (Churchland 1986). It therefore seems appropriate to consider how these recent advances will further our understanding of eye-movement control in reading, and, conversely, how cognitive models of reading might be used to guide neuroscience research.

5.2.1. The neural basis of reading. There is a growing consensus that most high-level and/or complex cognitive processes (e.g., language processing) are supported by large-scale networks that are themselves composed of several cortical and subcortical regions (Mesulam 1990; 1998; Posner & Raichle 1997). Consequently, it is not surprising that reading (which subsumes a large number of complex cognitive operations) is mediated by several of these large-scale networks. In the specific case of reading, these include (minimally) the networks that support vision, attention, eye-movement control, and language. In this section, we will provide a brief overview of these systems, and then speculate about how the language-processing system might interface with the systems that are responsible for programming and executing saccades.

The most natural place to begin an analysis of the neural systems underlying reading is the printed page. Visual processing of the text begins in the retina and progresses by way of the optic nerve to the optic chiasm and then the optic tract. From there, the visual "stream" splits into two pathways: The first projects to the lateral geniculate nucleus, and then the occipital cortex; the second innervates several subcortical structures, including one that is known to play a key role in eye movements – the superior colliculus (Leigh & Zee 1999; Sparks & Mays 1990). On the basis of results from numerous electrophysiological recording experiments with non-human primates, it has been estimated that there are 30 or more distinct cortical areas that are involved in vision (Felleman & Van Essen 1991; Maunsell & Newsome 1987; Van Essen & DeYoe 1995), although many of these areas perform functions that are less central to reading (e.g., motion perception). However, the low-level visual features (which comprise graphemes) are extracted and represented within the primary visual and extrastriate cortices (Grill-Spector et al. 1998).

The visual-processing stream continues on past this first analysis via two anatomically and functionally distinct pathways (Maunsell & Newsome 1987; Sagi & Julesz 1985; Ungerleider & Mishkin 1982; Van Essen & DeYoe 1995). The ventral, or "what," pathway extends along the inferior temporal cortices, and is thought to play an important role in feature integration and object recognition (Ishai et al. 1999; Tanaka 1996). Because words can be considered to be visual objects, the ventral system has also been implicated in the integration of those visual features which are necessary to represent visual word forms (Cohen et al. 2000; Poldrack et al. 1998). However, the location of the word-form area(s) remains controversial (see Posner et al. 1999a; 1999b; and Price 1997), and there is some evidence suggesting that the left medial extrastriate cortex is also intrinsically involved in the recognition of word forms (Peterson et al. 1989; 1990; Pugh et al. 2000).

The dorsal, or "where," pathway is thought to represent spatial information, such as the relative positions and orientations of objects (Ungerleider & Haxby 1994; Ungerleider et al. 1998). (For this reason, the dorsal system may also provide an interface between perception and action; Goodale & Milner 1992.) The dorsal pathway has also been implicated in visuospatial attention. In particular, the regions around the intraparietal sulci (i.e., the *parietal eye fields*) are thought to be central components of the visuospatial attention network. The other components include the superior colliculus (part of the mid-brain), the pulvinar nucleus of the thalamus, and a region that includes the precentral sulci/gyri and the posterior tips of the superior frontal sulci (i.e., the *frontal eye fields*) (Corbetta et al. 1993; Goldberg 1994; Kim et al. 1999; Leigh & Zee 1999; Luna et al. 1998; Rafal & Robertson 1995; Sweeney et al. 1996). Recent neuroimaging and electrophysiological recording research suggests that this network is involved in both covert and overt shifts of visuospatial attention, and that covert attention is probably represented in motor (more specifically, eye movement) coordinates (Corbetta 1998; Kim et al. 1999). This attention network also modulates both the analysis of objects in the ventral visual-processing pathway (Corbetta 1998) and perceptual processing in the striate and extrastriate cortices (Somers et al. 1999).

Although much less is known about language than the other components of reading, a long history of neuropsychological evidence (Caplan 1992) and a large number of more recent neuroimaging experiments indicate that the left inferior frontal gyrus (*Broca's area*) and the posterior part of the left superior and middle temporal gyri (*Wernicke's area*) are the two major language-processing areas. Both areas are engaged by a variety of receptive and expressive language tasks, including: (1) reading (Bavelier et al. 1997; Binder et al. 1997); (2) speech comprehension (Binder et al. 1997; Caplan et al. 1999; Schlosser et al. 1998; Stromswold et al. 1996); and (3) speech production (Bookheimer et al. 1997; Müller et al. 1997). The exact functional roles of these two language-processing areas are not known, but it has been suggested that Broca's area is involved in articulatory and syntactic processing, and that Wernicke's area supports lexical and semantic processing (Mesulam 1990). This hypothesis is (in part) based on the close proximity between Broca's area and the primary motor cortex. Wernicke's area, which receives input from the primary auditory cortex, may play a large role in lexical processing, such as binding the phonological word forms to their semantic representations (which are distributed elsewhere in the associative cortex; Mesulam 1998).

Because a single language network is presumably used to understand both written and spoken language, one of the central questions in reading research has been: How are the graphemes on a printed page converted into linguistic-based codes? The results of several recent neuroimaging experiments suggest that the left angular gyrus (which is located in the posterior part of the inferior parietal lobule) plays a critical role in computing grapheme-to-phoneme correspondences (Horwitz et al. 1998; Pugh et al. 2000). Because the left angular gyrus lies at the juncture of the extrastriate cortex and Wernicke's area, it is ideally situated to convert the orthographic word forms into their phonological counterparts. From the angular gyrus, the phonological word forms could then be used to gain access to semantic representations via Wernicke's area.

With respect to the time course of orthographic, phonological, and semantic processing, a recent meta-analysis (Posner et al. 1999a; 1999b) provides compelling evidence that key components of word-form processing can be completed within the time window that is necessary for it to function as a signal to initiate saccadic programming. The results of a recent ERP experiment, for example, indicate that certain aspects of lexical processing (e.g., word fre-

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

quency) can be discerned within 120–150 msec of word onset (Sereno et al. 1998). This would leave plenty of time (up to 130–180 msec) for the oculomotor system to program a saccade if one assumes a 250–300 msec fixation. This is an ample amount of time to initiate and complete the labile stage of saccadic program. (In E-Z Reader 7, the time needed to do this, $t(M_1)$, is equal to 187 msec.) Of course, additional programming time is available to the extent that pre-attentive visual processing (which, in our model, subsumes the first 90 msec of processing) allows early processing of parafoveal words. Nonetheless, the Sereno et al. results only show that it is plausible that word identification drives eye movements; they do not demonstrate that word identification drives eye movements, nor do their data suggest how the linkage is made. One possibility is discussed in the next section of this paper.

5.2.2. Specifying a neural implementation. E-Z Reader provides a functionalist account of eye-movement control in reading. As we have stated on previous occasions (Reichle et al. 1998; 1999), the model is neither a deep model of linguistic processing, nor a deep model of oculomotor control; instead, the model is simply our attempt to specify the functional relationships among a few key parameters (i.e., word frequency, predictability, retinal acuity, saccadic accuracy) to explain the time course of word identification and eye-movement control during reading. Consequently, up to now, we have remained completely agnostic about how the cognitive operations in our model might be implemented in the brain. Given the current state of cognitive neuroscience, however, it seems appropriate that this question should at least be considered.

Our answer – which at this time is obviously very speculative – is depicted schematically in Figures 13 and 14. Figure 13 depicts the eye movements that might occur as word_n and word_{n+1} are in turn fixated, the cognitive processes (as specified in our model) which give rise to this pattern of eye movements, and the cortical and subcortical systems in which these cognitive processes occur. Figure 14 shows both where in the brain these neural systems are localized (indicated by the numbers in the text below), and how processing is coordinated among these systems.

The sequence of events depicted in Figures 13 and 14 begins when the visual image of word_n hits the retina. After approximately 90 msec, the features that make up the word's orthographic form are being processed within the primary visual cortex (1). The individual letter features are then integrated at successively higher levels of the visual

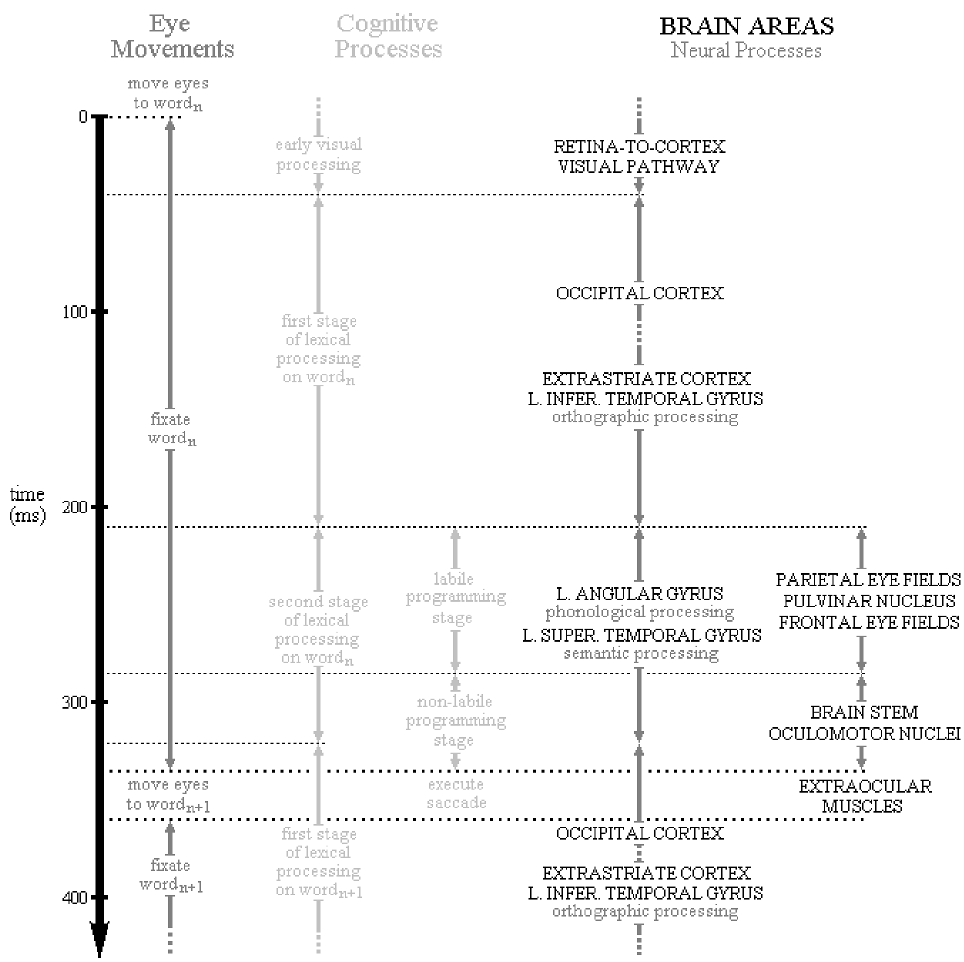

Figure 13. The time course of cognitive and neural processing during reading. The left side of the figure shows the pattern of fixations and saccades as the eyes move from word_n to word_{n+1}. The center of the figure shows the cognitive processes specified by the E-Z Reader model. The right side of the figure shows the neural processes (and their locations within the brain) that may mediate these cognitive processes.

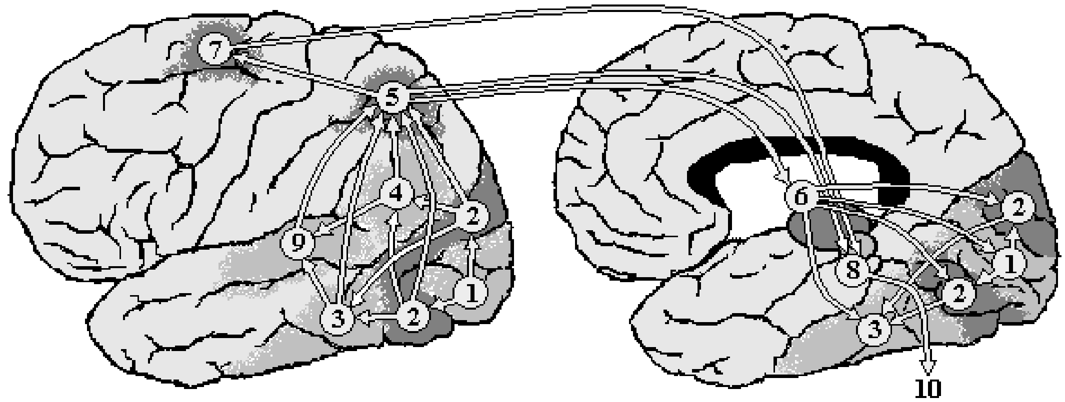

Figure 14. Sagittal views of the left lateral (left side of figure) and medial (right side of figure) cortical, thalamic (i.e., pulvinar nucleus), and mid-brain (i.e., superior colliculus) structures that may mediate the control of eye movements during reading. The number in the figure correspond to the following brain structures: (1) primary visual cortex (Brodmann's Area [*BA*] 17); (2) extrastriate cortex (*BA*s 18 & 19); (3) inferior temporal gyrus (*BA*s 20 & 37); (4) posterior inferior parietal lobule (i.e., angular gyrus; *BA* 39); (5) intraparietal sulci (i.e., parietal eye fields; *BA*s 7 & 40); (6) pulvinar nucleus of the thalamus; (7) superior prefrontal and posterior superior frontal gyri (i.e., frontal eye fields; *BA*s 6 & 8); (8) superior colliculus; (9) posterior middle and superior temporal gyri (i.e., Wernicke's area; *BA*s 21 & 22); and (10) the motor circuits of the brainstem which control the extraocular muscles and actually move the eyes. Although the figure only shows the left hemisphere, the right-hemisphere homologues of structures 1, 2, 5, 6, and 7 are also components of the visuospatial, attention, and oculomotor networks. Finally, the processing pathways among the areas depicted in the figure are not the only pathways that are known to exist; rather, the figure shows a few of the major pathways that have been shown to exist and which have a pattern of connectivity that is sufficient to support those cognitive processes that are important components of reading.

system as processing cascades from the striate to the extrastriate cortex (2). After approximately 150–250 msec, $word_n$'s orthographic form has been assembled in the left extrastriate cortex (2) and/or left inferior temporal gyrus (3), and this orthographic word form has been used to either access or assemble its phonological representation within the left angular gyrus (4).

Up to this point in time, both the eyes and attention have been focused on $word_n$. With the partial (i.e., orthographic and/or phonological) identification of $word_n$, however, the parietal eye fields (5) disengage visuospatial attention. The pulvinar nucleus of the thalamus (6) then moves the attentional "spotlight" forward, so that the frontal eye fields (7) and superior colliculus (8) can start using the low-spatial frequency information (e.g., word length) from the primary visual cortex to begin programming a saccade to $word_{n+1}$. This saccadic program takes (on average) approximately 240 msec to complete. During this time, the processing of $word_n$ continues; its orthographic (2 & 3) and/or phonological form(s) (4) are used to access the word's meaning by way of connections through Wernicke's area (9) to various parts of the associative cortex. If the meaning is accessed before the saccadic program has been completed, then the pulvinar (6) enhances the processing of $word_{n+1}$ (by shifting the internal attentional "spotlight" to the next word) and a preview benefit ensues. Otherwise, a saccade is executed by neural circuitry in the brainstem (10; see Leigh & Zee 1999) and the extraocular muscles, thereby moving the eyes move forward to $word_{n+1}$.

Again, it is important to note that saccadic programming in our model is initiated after the first stage of lexical processing on an attended word has been completed, whereas attention shifts only occur after an attended word has been identified. Attention is thus allocated serially, from one word to the next as each new word is identified. The serial allocation of attention is necessary because it preserves the temporal order of the words, along with any syntactic information that may be dependent upon word order (Pollatsek & Rayner 1999). This is, of course, not to say that some properties of an upcoming word might not occasionally be encoded in parallel to those of the word that is currently the focus of attention; as reviewed earlier, there is some evidence that (under certain conditions) properties of two words can indeed be encoded in parallel (Inhoff et al. 2000a; Kennedy 1998; 2000; Kennedy et al. 2002; Starr & Inhoff, in press). However, we believe that the default process during normal reading is one in which attention is allocated serially, so that the meaning of each new word that is identified can be integrated into a larger sentence representation, which is at least partially dependent upon word-order information. Furthermore, the version of our model presented in this paper (E-Z Reader 7) includes an early, pre-attentive visual processing stage that surveys the "terrain" of the upcoming text. Orthographic irregularities in the parafoveal might therefore register through this pre-attentive visual processing. This would allow the model to account for parafoveal-on-foveal effects stemming from unusual word beginnings in a manner that does not depend upon the serial shifts of attention that are normally associated with lexical processing.

6. Conclusion

Our contention throughout this paper has been that, although E-Z Reader does not provide a deep explanation of language processing, vision, attention, or oculomotor control, it does provide a viable framework for thinking about how these different cognitive processes interact during the course of normal reading. Like the oculomotor models that were discussed earlier in this paper, E-Z Reader can account for the effects of several basic visual and oculomotor variables on eye movements. In contrast to these models, however, E-Z Reader also accounts for many of the impor-

https://doi.org/10.1017/S0140525X03000104 Published online by Cambridge University Press

tant linguistic variables that are known to affect eye movements during reading. The model thus reflects our belief that, in order to account for the complex relationship between language processing and eye movements during reading, any adequate model of eye-movement control during reading will (almost by definition) have to include an account of language processing. Although our sketch of how the cognitive processes in E-Z Reader might map onto the neural systems responsible for guiding the eyes during reading is undoubtedly a gross over-simplification of what will undoubtedly turn out to be a much more complicated story, we would still argue that the mapping is precise enough to guide future cognitive neuroscience research.

Finally, it is worth emphasizing that E-Z Reader, like all of the other models reviewed in this paper, was developed primarily to explain the results of eye-tracking experiments. This should not be surprising because eye-tracking technology has proven to be an invaluable tool for studying reading. It is only natural that, as our understanding of eye movements and their determinants improve, this knowledge should be used to make inferences about the cognitive processes that occur during reading, and that these inferences should in turn be used to guide our modeling efforts. Because the last decade has witnessed unprecedented theoretical and methodological advances in the study of cognitive neuroscience, however, it is almost certain that these advances, too, will guide the development of the next generation of reading models. Like eye-movement data in the past, the discoveries of tomorrow will provide important guideposts for developing and evaluating future models.

ACKNOWLEDGMENTS
We thank our colleagues Charles Clifton, Susan Duffy, Julie Fiez, and Charles Perfetti for their many helpful discussions about various issues discussed in this article. Preparation of the article was supported by Grants HD17246 and HD27650 from the National Institutes of Health. The second author was also supported by a Research Scientist Award (MH02550) from the National Institute of Mental Health and by the Leverhulme Trust as a Leverhulme Professor at the University of Durham.

Requests for reprints should be sent to: Erik Reichle, Department of Psychology, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260; or via e-mail to: *reichle@pitt.edu.*

NOTES

1. Many models of word-identification have been proposed (Brown 1991; Bullinaria 1997; McClelland & Rumelhart 1981; Paap et al. 1982; Plaut et al. 1996; Seidenberg 1989; Seidenberg & McClelland 1989) to explain how orthography maps onto phonology and/or meaning, and how this process is affected by lexical variables (e.g., normative frequency, grapheme-phoneme regularity, etc.). Unfortunately, these models are generally limited in two ways: First, the entry point into these models is usually some highly abstract orthographic representation that bears little resemblance to the features that one might expect to be encoded by the visual system (e.g., homogenous retina acuity). Second, the models are generally fit to data from paradigms other than natural reading (e.g., lexical decision latencies). The models therefore say very little about the relationships among vision, eye movements, and word identification. Two interesting exceptions to this are McClelland's (1986) *programmable blackboard* model of reading and Shillcock et al.'s (2000) *split processing* model. The former model was designed to examine how fixation locations and visual acuity restrictions affect the model's word recognition performance; similarly, the split processing model was designed to examine how bisection of the visual field (and hence words) by the two cerebral hemispheres might explain why words are identified most rapidly when they are fixated near their centers.

2. We did not have a deep reason for choosing the name of our model. "E-Z Reader" was the name of a fictional character in a children's educational program *The Electric Company* in the U.S. and was clearly a spoof on the title of the movie *Easy Rider.*

3. Our discussion of parafoveal preview effects pertains to the processing of English. Indeed, there is some recent evidence (Deutsch et al. 2000; 2003) that indicates that, in Hebrew, morphological previews (in the form of the root morpheme, which is distributed throughout the word) provide preview benefit effects.

4. There is currently some disagreement regarding the extent to which the duration of a fixation prior to a skip is inflated. While there are reports of such an effect (Pollatsek et al. 1986; Reichle et al. 1998), others have reported null effects (Engbert et al. 2002; Radach & Heller 2000). In a very recent examination, we found effects on the order of 23 msec prior to a skip.

5. There is some dispute concerning the influence of "higher order" variables on where readers fixate. For example, Lavigne et al. (2000) reported that the eyes moved further into a word when that word was both high-frequency and predictable from the prior context. However, Rayner et al. (2001) and Vonk et al. (2000) found no such effect. In addition, Underwood et al. (1990; see also Hyönä et al. 1989) reported that the eyes moved further into words when the informative part of the word was at the end of the word. But Rayner and Morris (1992) and Hyönä (1995b) were unable to replicate this finding. On the other hand, there appears to be general agreement that an orthographically irregular letter cluster at the beginning of a word results in the eyes' initial landing position deviating toward the beginning of the word (Beauvillain & Doré 1998; Beauvillain et al. 1996; Hyönä 1995b).

6. A single set of parameter values were used in all of the simulations reported in this paper. These values were estimated by completing multiple grid-searches of the parameter's space so as to find the set that yielded the best overall fit to the Schilling et al. (1998) sentence corpus. For a complete description of our grid-search procedure, see the Appendix of Reichle et al. (1998).

7. Strictly speaking, Equation 1 produces word length effects (holding the eccentricity of the center of the word constant) only if the word straddles the fixation point. We used the arithmetic mean of the absolute distances in these formulas because of computational simplicity. However, if this were changed to some other combination rule (e.g., the geometric mean), then the equation would predict word length effects in all cases.

8. Frequency and predictability are not the only (nor necessarily the best) predictors of the time needed to identify a word in text. One problem with using frequency is that, even if the number of times a reader sees a given word in print was a perfect predictor of the time to identify the word, the Francis and Kučera (1982) norms (and other norms) are derived from corpuses that are unlikely to be representative of the texts that most readers encounter. (Another limitation of the Francis & Kučera norms is that they are derived from a fairly small corpus – only one million words.) Likewise, the predictability norms are also very crude estimates of how sentence context affects "on-line" lexical processing; in contrast to what actually happens during natural reading, the readers in these close-task studies have no visual information about the target words, but unlimited time to use all of the words in the sentence prior to the targets to guess their identities. Finally, the time needed to identify a word is likely to be a function of many other variables, including its part of speech, its concreteness, and the frequency with which it is encountered in spoken language. In summary, then, our decision to use frequency and predictability was not based on any a priori belief that these variables provide a complete explanation of lexical processing during reading. Instead, we are using them because they are known to produce significant effects in reading, and because they are clearly important determinants of word identification speed (i.e., how often a reader has seen the word before and how much top-down influence there is on the word).

9. In the current version of the model, for simplicity, attentional processing of word$_{n+1}$ (or words in general) is assumed to begin only when early visual processing of the entire word is completed. We are currently exploring versions of the model in which this assumption is relaxed, and attentional processing can begin when the early visual processing of parts of words is complete.

10. In our model, both the early pre-attentive visual processing and the non-labile stage of saccadic programming were halted during actual saccades. The former assumption was made because there is evidence that virtually no visual information is extracted during eye movements (Ishida & Ikeda 1989; Wolverton & Zola 1983). The latter assumption was necessary to ensure that a saccade could not be initiated while the eyes were already in motion. It should be noted that lexical processing does continue during saccades (Irwin 1998).

11. Figure 6 indicates that the model is underestimating the durations of single fixations. This problem stems from our increased estimate of the time needed to complete the labile stage of saccadic programming (i.e., $t(M_1) = 187$ msec). Because this "competitor" takes longer completing the "race" that determines whether or not a word will be refixated (i.e., the race between L_1 and M_1), the predicted durations of the first of two or more fixations is slightly too long, as indicated by the fact that the first-fixation durations are similar in length to the single-fixation durations. This also causes the single fixation durations for lower frequency words to be a bit too short. We don't think this is a major conceptual problem, as the primary goal in our simulations was to fit first-fixation durations and gaze durations rather than single-fixation durations. The problem seems fixable, however, by reducing $t(M_1)$ a bit and increasing the effect of frequency on the first stage of lexical access a bit. These changes shouldn't produce any catastrophic effects on other aspects of the fit, although perhaps the gaze durations may not fit quite well as in the current simulation.

12. We did not actually examine the landing site distributions in the Schilling et al. (1998) data because there were too few observations and because the properties of the distributions that we wanted to simulate are quite robust and have been reported in several places (e.g., McConkie et al. 1988; 1991; Rayner et al. 1996).

13. Interestingly, Vitu et al. (2001) recently reported an inverted optimal viewing position effect in reading in which readers' fixations were longer when they fixated near the center of a word than when they fixated away from the center of the word (when only one fixation was made on the word). Like Rayner et al. (1996), Vitu et al. also found frequency effects such that low-frequency words were fixated longer than high-frequency words.

14. In its current version, the model predicts that people will read about as effectively in a moving window condition in which the word to the left of fixation (word$_{n-1}$) and the fixated word (word$_n$) are visible as when the word to the right of fixation (word$_{n+1}$) and the fixated word (word$_n$) are visible (assuming word-boundary information is preserved to guide eye movements). This conflicts markedly with the findings in moving window studies (McConkie & Rayner 1975) where information to the right of the fixated word facilitates reading far more than information to the left of the fixated word. Perhaps the model does not depend critically on this attentional assumption and good predictions can be obtained with better attentional assumptions.

15. The model derives its name from Glenmore, Ireland – the place where much of the model was first developed (cf. Reilly & Radach 2003).

16. These results are open to alternative interpretations because the task was not natural reading, and thus did not actually require eye movements. Instead, the subject was required to read text on a computer monitor that was displayed through a stationary nine-character "window." The text was manually advanced via pressing keys that moved the text forward (1–9 character spaces) or backwards (1–3 character spaces), and a mask (covering 1, 3, or 5 character spaces) was placed over the center of the viewing window to occlude letters in the scotoma conditions.

17. For example, we previously argued that the last version of the model discussed in Reichle et al. (1998), E-Z Reader 5, is superior to an earlier version, E-Z Reader 3, even though the latter model provided a slightly better aggregate fit to the Schilling et al. (1998) data. This claim was based primarily on a qualitative argument: In E-Z Reader 5 (but not E-Z Reader 3), the rate of lexical processing decreases as the disparity between the word being processed and the fovea increases. Although this feature of E-Z Reader 5 makes the model more psychologically plausible, the counter-argument could be made that the lack of an improvement of the model's overall performance does not warrant the additional of two parameters. However, Salvucci and Anderson (1998; 2001) recently found additional evidence supporting our claim. Briefly, Salvucci and Anderson first replicated the Schilling et al. experiment with a different subject population, and then used several different eye-movement protocol algorithms to determine how well E-Z Readers 3 and 5 could account for the eye-movement data of individual subjects. They also examined how well the models could account for two sequential measures: (1) the proportions of saccades of each given length; and (2) the proportions of saccades of each given length following saccades of various lengths. The results of these analyses indicated that E-Z Reader 5 fit all three measures better than did E-Z Reader 3, and that E-Z Reader 5 in fact provided a better account of the finer-grained, sequential aspects of the observed eye-movement data. Moreover, these results suggest that E-Z Reader 7 (which also includes the visual acuity assumption) may also provide better quantitative fits than earlier, simpler, versions of the model.

18. Furthermore, our simulations to date (Pollatsek et al. 2003) indicate that a simple race model (i.e., a race between two independent processes, a direct look-up process and a constructive process) is unlikely to account for the observed pattern of data in Hyönä and Pollatsek (1998) and Pollatsek et al. (2000). This is an illustration of how modeling can help sharpen one's thinking about such issues.

19. Because the effects of higher-order language processing are often delayed and/or apparent over a wider temporal window than are the effects of lower-order language processing, the former may actually be less difficult to simulate than the latter. Paradoxically, it may be more difficult to evaluate a model's capacity to simulate higher-order linguistic effects for these same reasons.

TABLE 1 NOTES:

a. The primary references for the reading models are: (1) *Minimal-Control* (Suppes 1990; 1994); (2) *Strategy-Tactics* (O'Regan 1990; 1992b); (3) *Word-Targeting* (McConkie et al. 1988; Reilly & O'Regan 1998); (4) *Push-Pull* (Yang & McConkie 2001); (5) SWIFT (Engbert et al. 2002); (6) Glenmore (Reilly & Radach 2003); (7) *Mr. Chips* (Klitz et al. 2000; Legge et al. 1997); (8) *Attention-Shift* (Reilly 1993); (9) *E-Z Reader* (Reichle et al. 1998; 1999); (10) *EMMA* (Salvucci 2000a; 2000b); and (11) *Reader* (Just & Carpenter 1980; 1987; 1992; Thibadeau et al. 1982).

b. *GAG* indicates that the model assumes that attention is distributed as a gradient during reading (i.e., "guidance by attentional gradient"); *SAS* indicates that the model assumes the serial allocation of attention from one word to the next during reading (i.e., "sequential attention shift"); *POC* indicates that the model is primarily an oculomotor model and thus makes no specific assumptions about how attention is allocated during reading.

c. *Yes* indicates that a model can explain a result; *No* indicates that the model (as it is currently instantiated) does not explain a result; *Ltd* indicates that the model's account of a phenomenon is incomplete or limited (e.g., the model predicts parafoveal preview benefit, but the benefit is not modulated by foveal processing difficulty).