Eye movements and attention in reading, scene perception, and visual search

Keith Rayner*

* University of California, San Diego, CA, USA

First Published on: 14 May 2009
Eye movements are now widely used to investigate cognitive processes during reading, scene perception, and visual search. In this article, research on the following topics is reviewed with respect to reading: (a) the perceptual span (or span of effective vision), (b) preview benefit, (c) eye movement control, and (d) models of eye movements. Related issues with respect to eye movements during scene perception and visual search are also reviewed. It is argued that research on eye movements during reading has been somewhat advanced over research on eye movements in scene perception and visual search and that some of the paradigms developed to study reading should be more widely adopted in the study of scene perception and visual search. Research dealing with “real-world” tasks and research utilizing the visual-world paradigm are also briefly discussed.

Keywords: Eye movements; Attention; Reading; Scene perception; Visual search.
It is well known that viewers can allocate attention independent of eye position in simple tasks in which they must hold their eyes on a fixation target while other stimuli are presented elsewhere in the visual field (Posner, 1980). But, how easy is it to have attention and eye position in different locations in complex tasks like reading, scene perception, and visual search? It is my contention that most of the time in such tasks, either (a) eye location (overt attention) and covert attention are overlapping and at the same location or (b) attention disengagement is a product of a saccade programme (wherein attention precedes the eyes to the next saccade target). Thus, while attention and eye position can be dissociated in these tasks, any such dissociation is generally a property of the processing system rather than some type of strategy used by readers or viewers in these tasks. Specifically, attention precedes a saccade to a given saccade target location (Deubel & Schneider, 1996; Henderson, 1993; Hoffman & Subramaniam, 1995; Irwin & Gordon, 1998; Irwin & Zelinsky, 2002; Kowler, Anderson, Dosher, & Blaser, 1995; Rayner, McConkie, & Ehrlich, 1978; Shepherd, Findlay, & Hockey, 1986), and this is a property of the processing system. There are many excellent studies of attention that do not involve eye movements, and they have been very informative with respect to understanding covert attention. However, that type of research is not discussed here. Rather, my goal is to review research on eye movements, as measures of overt attention, during complex cognitive processing tasks.

It is quite apparent that research utilizing eye movements to examine cognitive processing tasks is burgeoning as more and more researchers have started to use eye tracking techniques in the last few years. In large part, this is because systems to track the eyes have become more readily available and more user friendly. However, we have also entered an era of eye movement research, particularly in the domain of reading, in which a great deal of research is being driven by predictions that emerge from computational models of eye movement control. In this article, research on the following topics are reviewed with respect to reading: (a) the perceptual span (or span of effective vision), (b) preview benefit, (c) eye movement control, and (d) models of eye movements. I also discuss some related issues with respect to eye movements during scene perception and visual search. The review is somewhat selective in that I focus largely on research from my laboratory and related work. Furthermore, the review is not as systematic with respect to scenes and search as it is with reading; some issues that I find interesting with respect to scenes and search are the focus of this part of the review. In large part, it is argued that research on eye movements during reading has been somewhat advanced over research on eye movements in scene perception and visual search and that some of the paradigms developed to study reading should be more widely adopted in the study of scene perception and visual search. Research dealing with “real-world” tasks and research utilizing the visual-world paradigm are also briefly discussed.

Background information on eye movements

The two basic components of eye movements in the various tasks under consideration here are the movements themselves (called saccades) and the fixations (the period of time when the eyes remain fairly still and new information is acquired from the visual array). Since vision is suppressed during a saccade (Matin, 1974), new information is only acquired during fixations. There are special situations wherein information can be acquired during a saccade (see Campbell & Wurtz, 1979; Uttal & Smith, 1968), but under most normal circumstances we do not obtain new information during a saccade because the eyes are moving so quickly across the stable visual stimulus that only a blur would be perceived. Furthermore, masking, caused by the information available before and after the saccade, eliminates any perception of blurring. While new information is not encoded during saccades, cognitive processing does continue in most situations during the saccade (Irwin, 1998; Irwin & Carlson-Radavsk, 1996).

Eye movements are motor responses that take time to plan and execute. Saccade latency, the time needed to encode the location of a target in
the visual field and initiate an eye movement, is of
the order of 175–200 ms (Becker & Jürgens, 1979;
Rayner, Slowiaczek, Clifton, & Bertera, 1983).
However, it varies quite a bit as a function of the
exact nature of the situation. Saccade duration,
the amount of time that is takes to actually move
the eyes, is a function of the distance moved. A
2-deg saccade, typical of reading, takes about
30 ms, while a 5-deg saccade, typical of scene
perception, takes around 40 to 50 ms (Abrams,
Meyer, & Kornblum, 1989; Rayner, 1978a).

Eye movements are necessary because of the
anatomy of the retina and limitations due to
acuity outside of the fovea. In reading, for
example, the line of text that the reader is
looking at can be divided into three regions: the
foveal region (2 degrees in the centre of vision),
the parafoveal region (extending from the foveal
region to about 5 degrees on either side of
fixation), and the peripheral region (everything
beyond the parafoveal region). Although acuity is
very good in the fovea, it is not nearly so good
in the parafovea, and it is even poorer in the
periphery. Hence, viewers move their eyes so as
to place the fovea on that part of the stimulus
they want to see clearly. While the same kind of
constraints hold for scene perception and visual
search, unless the array is particularly dense,
viewers can typically process more information
around their fixation point (i.e., further into
eccentric vision) in these tasks than in reading.

It is tempting to think that eye movements in
each of these tasks would be controlled by the
same mechanisms, and that the same principles,
with respect to eye movements, should hold
across the three tasks. After all, the neural circuitry
for controlling eye movements is the same across
the tasks. However, it is actually somewhat
hazardous to generalize across these tasks in
terms of eye movement behaviour. For example,
Rayner, Li, Williams, Cave, and Well (2007c)
demonstrated that fixation durations and saccade
lengths in reading do not correlate with those
measures in scene perception and search (see also
T. J. Andrews & Coppola, 1999). Interestingly,
fixation duration and saccade length did not corre-
late significantly with each other\(^1\) within any of
these tasks (see also Castelhano & Henderson,
2008a; Rayner & McConkie, 1976). Why do eye
movement measures in reading not correlate well
with the same measures in scene perception and
visual search despite having the same underlying
neural circuitry? Presumably, the cognitive mech-
anism involved in the different tasks, and how
the cognitive system interacts with the oculomotor
system, differ as a function of the task.

In addition to the fact that there are not
particularly high correlations in eye movement
measures across tasks,\(^2\) it is also the case that the
basic descriptive characteristics of eye movements
differ as well. Table 1 shows the range of average
fixation durations typically associated with silent
reading, oral reading, scene perception, and
visual search. From Table 1, it is obvious that fix-
atations tend to be longer (a) in oral reading than in
silent reading and (b) in scene perception than in
reading, and that (c) the range of fixation dur-
ations is greater in visual search than in the other
tasks. Fixations are longer in oral reading than
silent reading because (a) the reader has to
produce each word as it is read, and (b) the eyes
(which move faster than the reader can produce
words) often stay in place longer so that they do
not get too far ahead of the voice. Fixations in
scene perception tend to be longer than those in
reading because information is taken in from a
wider area in scenes than in reading. And, the
large range associated with fixation durations in

\(^{1}\) While across a large passage of text, the correlation between fixation duration and saccade length tends to be close to zero
(Rayner & McConkie, 1976), it is the case that for certain segments of text one can find reasonable-sized correlations. Thus,
when the text is difficult, readers make long fixations and short saccades, leading to a significant correlation.

\(^{2}\) Within task, there are very high correlations for eye movement characteristics. Thus, Castelhano and Henderson (2008a) found
that fixation durations for a group of viewers tended to correlate highly independently of whether a photo or line drawing was used as
the stimulus. Likewise, saccade length was highly correlated. But, again, there was little correlation between fixation duration and
saccade length per se.
visual search is presumably related to how simple or complex the search array is.

With respect to saccade lengths, for reading the appropriate metric is letters rather than visual angle since the distance the eyes traverse from one saccade to the next is determined by letters rather than visual angle as long as the text is of normal size (Morrison & Rayner, 1981; see also McDonald, 2006a). However, for comparability to the other tasks, saccade size during reading is also given in degrees of visual angle in Table 1. It is also obvious that the distance the eyes move in scene perception and visual search can be highly variable depending on the complexity of the array; when the array is complex and crowded, saccades are shorter (the same would hold for a highly complex scene).

### Table 1. The range of mean fixation durations and the mean saccade length in silent reading, oral reading, scene perception, and visual search

<table>
<thead>
<tr>
<th>Task</th>
<th>FD (ms)</th>
<th>Deg</th>
<th>Letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silent reading</td>
<td>225–250</td>
<td>2</td>
<td>7–9</td>
</tr>
<tr>
<td>Oral reading</td>
<td>275–325</td>
<td>1.5</td>
<td>6–7</td>
</tr>
<tr>
<td>Scene perception</td>
<td>260–330</td>
<td>4–5</td>
<td></td>
</tr>
<tr>
<td>Visual search</td>
<td>180–275</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Note: FD = fixation duration; SL saccade length. The average fixation duration in scene perception and visual search can very much be influenced by the exact nature of the task that participants are given.

Eye movements in reading

As indicated in Table 1, the average fixation duration in reading is on the order of 225–250 ms, and the average saccade length is 7–9 letter spaces for readers of English and other alphabetic writing systems. However, it is important to keep in mind that these values are averages, and there is considerable variability in both. Thus, fixation durations can be as short as 50–75 ms and as long as 500–600 ms (or more), and saccade length can be as short as 1 letter space and as long as 15–20 letter spaces (or more). Regressions (saccades that move backwards in the text) are the third important component of eye movements in reading and occur about 10–15% of the time in skilled readers. The long saccades just mentioned tend to follow a regression since readers typically move forward in the text past the point from which they originally launched the regression. Most regressions are to the immediately preceding word, though when comprehension is not going well or the text is particularly difficult, more long-range regressions occur to earlier words in the text. Regressions are not particularly well understood because it is difficult to control them experimentally (though see Inhoff & Weger, 2005; Murray & Kennedy, 1988; Rayner, Juhasz, Ashby, & Clifton, 2003a; Weger & Inhoff, 2006, 2007; for an interesting discussion of regressions due to sentence parsing difficulties, see Mitchell, Shen, Green, & Hodgson, 2008). Finally, regressions need to be distinguished from return sweeps, which are right-to-left saccades from the end of one line to the beginning of the next. It is also instructive to note that the first and last fixations on a line are typically 5–7 letter spaces from the end of the line. Thus, about 80% of the text typically falls between the extreme fixations.

The values shown in Table 1 are for skilled readers of English. However, these values can be very much influenced by text difficulty, reading skill, and characteristics of the writing system. Thus, as text gets more difficult, fixations get longer, saccades get shorter, and more regressions are made (Rayner, 1998). Also, typographical variables like font difficulty can influence eye movements; more difficult to encode fonts yield longer fixations, shorter saccades, and more regressions (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006d; Slattery & Rayner, 2009). Beginning and dyslexic readers have longer fixations, shorter saccades, and more regressions (Rayner, 1998), as do less skilled readers (Ashby, Rayner, & Clifton, 2005). As far as writing system is concerned, the one that is most different from English is Chinese.
Chinese readers tend to have average fixation durations that are quite similar to readers of English, and their regression rate does not differ dramatically. Where they do differ is that their average saccade length is much shorter than that of readers of English as they typically move their eyes only 2–3 characters (which makes sense given that linguistic information in Chinese is more densely packed than in English). Likewise, readers of Hebrew (a language that is also more densely packed than English largely because vowels are deleted in the printed orthography) tend to yield shorter saccades (about 5.5 letter spaces) than readers of English (Pollatsek, Bolozky, Well, & Rayner, 1981), though their fixation durations are comparable.

One great virtue of eye movement data is that they give a good moment-to-moment indication of cognitive processes during reading. Thus, variables such as word frequency and predictability have strong influences on fixation times on a word (Rayner, 1998). However, the average fixation duration measure is not a particularly informative measure for inferring moment-to-moment processing; it is a valuable global measure, but there are also a number of local measures that provide more informative estimates of moment-to-moment processing time.

The problem with average fixation duration is related to two components of reading. First, readers skip words during reading;3 content words are fixated about 85% of the time, while function words are fixated about 35% of the time. Function words are skipped more because they tend to be short, and there is a clear relationship between the probability of fixating a word and its length. As word length increases, the probability of fixating a word increases (Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996) or conversely as the length of the word decreases, the probability of fixating it decreases, and the probability of skipping it increases (Brysbaert, Drieghe, & Vitu, 2005; Rayner, 1998). Words that are 2–3 letters long are only fixated around 25% of the time, whereas words that are 8 letters or more are almost always fixated. Second, longer words are often fixated more than once before leaving the word; that is, they are refixated (see McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; McDonald & Shillcock, 2004; Vergilino & Beauvillain, 2000, 2001; Vergilino-Perez, Collins, & Dore-Mazars, 2004). The joint problem of skipping and refixations has led to the development of alternative measures of fixation time. These measures are first-fixation duration (the duration of the first fixation on a word), single-fixation duration (those cases where only a single fixation is made on a word), and gaze duration (the sum of all fixations on a word prior to moving to another word). All of the measures are contingent on the word being fixated on a first-pass forward fixation.

If it were the case that readers fixated each word and only once on each word, then average fixation duration on a word would be a useful measure. But, as noted above, the reality is that many words are skipped, and some words are refixated. There is good reason (see below for discussion) to believe that the words that are skipped are processed on the fixation prior to the skip (and after the skip to some extent), and likewise there is good reason to think that words are refixated in order to fully process their meaning. The solution to this possible conundrum is to utilize all three measures just described, which provide a reasonable, though not perfect, estimate of how long it takes to process each word (Rayner, 1998). The reasons these measures are not a perfect estimate are, as is discussed below, that (a) preview information is obtained from a word prior to fixating it, and (b) there are spillover effects, wherein the processing of a given word spills over to the next fixation (Rayner & Duffy, 1986). When regions of interest are larger than a single word, a number of other measures like first-pass reading time, second-pass reading time, go-past time

---

3 The fact that words are skipped obviously means that readers do not invariably move forward in the text fixating on each successive word in its canonical order. However, some type of inner speech code presumably aids the reader to maintain the correct word order.
(the elapsed time from when a reader first enters a region until they move past it forward in the text, sometimes also referred to as regression path duration), and total reading time are generally computed.

When readers reread passages of text, the fixation pattern tends to remain fairly similar across readings (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Schnitzer & Kowler, 2006). Fixations become a bit shorter, and saccades become a bit larger; the main difference is that fewer fixations are made, and the frequency of regressions decreases. While average fixation durations and saccade sizes are only minimally affected when passages are reread, when fixation times associated with specific target words are examined via local measures (first fixation and gaze duration), there is a larger decrease in fixation time. This further demonstrates why average fixation time is not a particularly good measure of moment-to-moment processing.

Do the two eyes move in synchrony during reading (and other tasks)? It has generally been assumed that there is near-perfect binocular coordination during reading and that the two eyes start moving at the same time and land on the same letter. However, recent research (see Kirkby, Webster, Blythe, & Liversedge, 2008, for a comprehensive review) has demonstrated that up to 40–50% of the time the eyes are on different letters, and sometimes the two eyes are crossed (Heller & Radach, 1999; Liversedge, Rayner, White, Findlay, & McSorley, 2006a; Liversedge, White, Findlay, & Rayner, 2006b). Interestingly, the amount of disparity tends to be greater in beginning readers than in skilled readers (Blythe et al., 2006). More importantly, perhaps, word frequency and case alternation affect fixation duration in reading, but not fixation disparity (Juhasz, Liversedge, White, & Rayner, 2006). Thus, while researchers may need to worry about those rare situations in which the eyes would be on different words, it is clear that when both eyes are fixated within the same word robust effects like the frequency effect still emerge.

Four central issues with respect to eye movements during reading are now discussed. They are: (a) the perceptual span, (b) preview benefit, (c) the control of eye movements, and (d) models of eye movements.

**The perceptual span in reading**

Each time the eyes pause for roughly 225–250 ms, how much information is the reader able to process and use during that fixation? Readers often have the impression that they can clearly see the entire line of text, even the entire page of text. But, this is an illusion as experiments utilizing a gaze-contingent moving-window paradigm (McConkie & Rayner, 1975; Rayner & Bertera, 1979) have clearly demonstrated.

In experiments using the moving-window paradigm, the rationale is to vary how much information is available to a reader and then determine how large the window of normal text has to be before readers read normally (see Figure 1). Conversely, how small can the window be before there is disruption to reading? Thus, in the experiments, within the window area text is normally displayed, but outside of the window the letters are replaced (with other letters or with Xs or a homogenous masking pattern). Research using this paradigm has demonstrated that skilled readers of English and other alphabetic writing systems obtain useful information from an asymmetric region extending roughly 3–4 character spaces to the left of fixation (McConkie & Rayner, 1976a; Rayner, Well, & Pollatsek, 1980b; N. R. Underwood & McConkie, 1985) to about 14–15 character spaces to the right of fixation.

---

4 In the most extreme situation, the window contains only the fixated letter, thereby creating a situation in which the reader is literally forced to read letter by letter. In this situation, normal readers’ eye movement data are very much like the eye movement data of brain-damaged pure alexic or letter-by-letter readers (Johnson & Rayner, 2007; Rayner & Johnson, 2005).

5 In some situations, readers can obtain information further to the left of fixation (Binder, Pollatsek, & Rayner, 1999). For example, when a word is skipped, attention may often be directed to the left of fixation following the skip.
It is well known that viewers can allocate attention independently.

15-character window:

Xx xx xxxx xxxn that viewexxx xxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx

+  

Xx xx xxxx xxxx xhat viewers can xxxxxxxx xxxxxxxxxx xxxxxxxxxx

+  

29-character window:

Xx is well known that viewers can xxxxxxxx xxxxxxxxxx xxxxxxxxxx

+  

Xx xx xxxx xxxn known that viewers can allocaxx xxxxxxxx xxxxxxxxxx

+  

Asymmetric window:

xxxxxxxxxxxxxxxx that viewxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

+  

xxxxxxxxxxxxxxxxxxxxxxviewers canxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

+  

One-word window:

Xx xx xxxx xxxx that xxxxxxx xxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx

+  

Xx xx xxxx xxxx xxxx viewers xxx xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx

+  

Figure 1. Examples of the moving-window paradigm. The top line shows a normal line of text. The moving windows consist of the following examples (with two fixations, marked by the +, shown): a 15-character moving window (7 characters to the left and to the right of fixation), a 29-character window (14 characters to the left and right of fixation), an asymmetric window extending 3 characters to the left of fixation and 7 to the right (with the spaces between words filled in outside of the window), a one-word window, a three-word window, and a two-word window with the letters outside the window replaced with visually similar letters (first line) or random letters (second line). The data generally show that reading is easier when the spaces are not filled in. Also, when the window is small, reading performance is generally better when x's are outside the window than when letters are.

(Continued)
Two-word window:

\[ \text{Xx xx xxxx xxxx that viewers xxx xxxxxxxx xxxxxxxxxx xxxxxxxxxx} \]

+  

\[ \text{Xx xx xxxx xxxx xxxx viewers can xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx} \]

+  

Three-word window:

\[ \text{Xx xx xxxx xxxx that viewers can xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx} \]

+  

\[ \text{Xx xx xxxx xxxx xxxx viewers can allocate xxxxxxxxxx xxxxxxxxxx xxxxxxxxxx} \]

+  

Two-word window:

\[ \text{Lf rz voff hmevm that viewers asm cffeasfo sffomfrem rmbqombomf} \]

+  

\[ \text{Kw op hkml qeryp bvxs viewers can ityk1me sdhklrsty poytwmbsdjk} \]

+  

Figure 1. Continued.

(DenBuurman, Boersma, & Gerrissen, 1981; McConkie & Rayner, 1975; Rayner & Bertera, 1979; Rayner, Well, Pollatsek, & Bertera, 1982; N. R. Underwood & McConkie, 1985; N. R. Underwood & Zola, 1986). Indeed, if readers have the fixated word and the word to the right of fixation available on a fixation (and all other letters are replaced with visually similar letters), they are not aware that the words outside of the window are not normal, and their reading speed only decreases by about 10%. If two words to the right of fixation are available within the window, there is very little slowdown in reading (Rayner et al., 1982).

In some recent clever experiments, Miellet, O’Donnell, and Sereno (in press) introduced a variation of the moving-window paradigm, which they termed parafoveal magnification, in which letters around the fixation were normal sized, but letters increased in size in eccentric vision so as to compensate for the loss of acuity associated with parafoveal vision. They demonstrated that the size of the perceptual span to the right of fixation remained about 14 letter spaces (thus replicating prior research by McConkie & Rayner, 1975; Rayner & Bertera, 1979) with parafoveal magnification, and they also replicated
It is well known that XXXXXXXewers can allocate attention independent

It is well known that XXXXXXXs can allocate attention independent

It is well known that viewers XXXXXXXlocate attention independent

Figure 2. Examples of the moving-mask paradigm. Three consecutive fixations with a 7-character mask.

the finding (Morrison & Rayner, 1981) that the distance the eyes move is driven by letters. These findings are quite consistent with the view that attention and ongoing processing constraints, and not visual acuity, determine how much information can be obtained on each eye fixation in reading.

Readers do not utilize information from lines below the currently fixated line (Inhoff & Briihl, 1991; Inhoff & Topolski, 1992; Pollatsek, Raney, LaGasse, & Rayner, 1993). However, if the task is visual search rather than reading, then information can be obtained below the currently fixated line (Pollatsek et al., 1993). Finally, in moving-mask experiments (Fine & Rubin, 1999a, 1999b, 1999c; Rayner & Bertera, 1979; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981) in which a visual mask moves in synchrony with the eyes on each fixation, thus covering the letters in the centre of vision (see Figure 2), reading is very difficult, if not impossible, as central foveal vision is masked (and only letters in parafoveal vision are available for reading). In essence, the moving-mask paradigm creates an artificial foveal scotoma mimicking patients with brain damage that effectively eliminates their use of foveal vision.

Just as the characteristics of the writing system influence saccade size in reading, they also influence the size of the perceptual span. Thus, for readers of Chinese (which is now typically read from left to right in mainland China), the perceptual span extends 1 character to the left of fixation to 2–3 characters to the right (H. Chen & Tang, 1998; Inhoff & Liu, 1998). And the perceptual span for Hebrew readers is asymmetric and larger to the left of fixation (Pollatsek et al., 1981) since they read from right to left.

Reading skill also influences the size of the perceptual span since beginning readers (Häikö, Bertram, Hyöna, & Niemi, 2009; Rayner, 1986) and dyslexic readers (Rayner, Murphy, Henderson, & Pollatsek, 1989) have smaller spans than more skilled readers. Presumably, difficulty encoding the fixated word leads to smaller spans for both beginning and dyslexic readers. Older readers read more slowly than younger college age readers (Laubrock, Kliegl, & Engbert, 2006; Rayner et al., 2006d) and their perceptual span seems to be slightly smaller and less asymmetric than that of younger readers (Rayner, Castelhano, & Yang, in press).

Preview benefit in reading
Research using another type of gaze-contingent display change paradigm, the boundary paradigm (Rayner, 1975), has revealed important information about what kind of information is integrated across saccades. In the boundary paradigm, an invisible boundary is just to the left of a target word (see Figure 3), and before the reader crosses the boundary, there is typically a preview different from the target word. When the eyes cross the boundary, the preview is replaced by the target word. The
It is well known that readers can allocate attention independent +

It is well known that viewers can allocate attention independent +

It is well known that vievcnr can allocate attention independent +

It is well known that viewers can allocate attention independent +

It is well known that hbgkrsk can allocate attention independent +

It is well known that viewers can allocate attention independent +

Figure 3. Examples of the boundary paradigm. In all of the examples, the boundary is located after the last letter in the word that; when the reader’s eyes cross the boundary the preview word (shown in the first line of each example) changes to the target word (viewers, as in the second line of each example). In the first example, the preview is another word (readers); in the second example, the preview is a nonword (vievcnr) that is orthographically similar and shares the initial letters with the target word; and in the third example the preview is a random string of letters (hbgkrsk).

sentence always makes sense with the target word, and readers are unaware of the identity of the preview and of the display change.\(^6\) Research using this paradigm has revealed that when readers have a valid preview of the word to the right of fixation, they spend less time fixating that word (following a saccade to it) than when they do not have a valid preview (i.e., another word or nonword or random string of letters initially occupied the target location). The size of this preview benefit is typically of the order of 30–50 ms. Research using this technique has revealed that readers do not combine a literal representation of the visual information across saccades (McConkie & Zola, 1979; Rayner, McConkie, & Zola, 1980a). Rather, information about the beginning

\(^6\) In the boundary paradigm, the display change from the preview to the target word occurs during a saccade when vision is suppressed. Typically, readers are not aware of the change. But, can the results of boundary experiments be attributed to artefacts associated with the change? Inhoff, Starr, Liu, and Wang (1998) directly tested this by varying the speed of the display change and the refresh rate of the display monitor. They found no evidence to suggest that results of gaze-contingent change experiments were artefacts of the paradigm. Thus, when the timing of the display change is such that the change occurs during the saccade, researchers can be confident that the only thing that influences the data is the experimental manipulation. However, when the timing is such that the display change is too slow, a different pattern of data will appear. Thus, readers who are aware of the display change show a different pattern of data than do readers who are not aware (White et al., 2005a).
and ending letters of words (Briihl & Inhoff, 1995; Inhoff, 1989b; Johnson, Perea, & Rayner, 2007; Lima & Inhoff, 1985; Rayner et al., 1982) and orthographic codes (Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek, 2005b; Rayner, 1975; White, Rayner, & Liversedge, 2005b; Williams, Perea, Pollatsek, & Rayner, 2006) is integrated across saccades, as are abstract letter codes and phonological information (Ashby & Rayner, 2004; Ashby, Treiman, Kessler, & Rayner, 2006; Chace, Rayner, & Well, 2005; Miellet & Sparrow, 2004; Pollatsek, Lesch, Morris, & Rayner, 1992; Rayner et al., 1980a; Sparrow & Miellet, 2002).

While information about letter position, abstract letter codes, orthographic codes, and phonological codes is integrated across saccades (and serves as the basis of the preview benefit effect), it is quite interesting that semantic information is not (Altarriba, Kambe, Pollatsek, & Rayner, 2001; Hyöna & Häikö, 2005; Rayner, Balota, & Pollatsek, 1986; Rayner et al., 1980a). That is, words that typically produce priming in a standard naming or lexical decision task (e.g., the prime word tune primes the target word song) do not yield priming when the prime word is in parafoveal vision (with the target word presented as soon as the reader crosses the invisible boundary location). Although this result is sometimes considered puzzling, it is probably due to the fact that words in parafoveal vision are degraded sufficiently that readers cannot typically process their meaning.

This is not to say that words in parafoveal vision cannot be identified, because they can. When words are quite short or sufficiently constrained by context, as discussed below, readers skip over them, and it is generally agreed that these words are identified on the fixation prior to the skip.

Just as there is no strong evidence that semantic information is integrated across saccades, there is also no evidence that morphological information is integrated across saccades when reading English (Inhoff, 1989a; Juhasz, White, Liversedge, & Rayner, 2008; Kambe, 2004; Lima, 1987). On the other hand, readers of Hebrew do integrate morphological information across saccades (Deutsch, Frost, Peleg, Pollatsek, & Rayner, 2003; Deutsch, Frost, Pollatsek, & Rayner, 2000, 2005). Morphological information is more central to processing Hebrew than English, and the difference in findings presumably reflects this fact.

The amount of preview benefit readers obtain varies as a function of the difficulty of the fixated word. If the fixated word is difficult to process, readers get little or no preview benefit from the word to the right of fixation (Henderson & Ferreira, 1990; Kennison & Clifton, 1995; White, Rayner, & Liversedge, 2005a); if the fixated word is easy to process, readers get better preview benefit from the word to the right (Balota et al., 1985; Drieghe et al., 2005b). Interestingly, it has also been found that preview benefit is larger within words than across words (Hyöna, Bertram, & Pollatsek, 2004; Juhasz, Pollatsek, Hyöna, Drieghe, & Rayner, 2009; Pollatsek & Hyöna, 2005). This finding is discussed further below.

Finally, while eye fixations land nearer to the beginning of misspelled words, the effect holds whether the previous word is easy or difficult to process (White & Liversedge, 2006b).

Another interesting issue concerns the spatial extent of preview benefit. Specifically, do readers obtain preview benefit from word \( n + 2 \) (the word two to the right of the currently fixated word)? While it is clear that readers generally obtain preview benefit from word \( n + 1 \), readers typically do not get preview benefit from word \( n + 2 \) (Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Kliegl, Risse, & Laubrock, 2007; McDonald, 2005, 2006b; Rayner, Juhasz, & Brown, 2007a). It may be that when word \( n + 1 \) is a very short high-frequency word (2–3 letters) preview benefit is obtained from word \( n + 2 \) and that when readers target their next saccade to word \( n + 2 \) preview benefit is obtained.\(^7\) However, when readers fixate word \( n + 1 \) and

\(^7\) Chinese readers obtain preview benefit from characters and words to the right of their fixation (Inhoff & Liu, 1998; Liu, Inhoff, Ye, & Wu, 2002; Pollatsek, Tan, & Rayner, 2006b; Tsai, Lee, Tzeng, Hung, & Yen, 2004; J. Yang, Wang, Xu, & Rayner, in press; Yen, Tsai, Tzeng, & Hung, 2008), and they can sometimes obtain preview benefit from word \( n + 2 \).
word \( n + 2 \) in sequence, they obtain preview benefit from word \( n + 1 \) but not word \( n + 2 \).

**Parafoveal-on-foveal effects.** A topic that has been highly controversial recently is the extent to which the characteristics of the word to the right of fixation can influence the duration of the fixation on the currently fixated word. Such effects are referred to as parafoveal-on-foveal effects. Some studies have found that orthographic properties of the word to the right of fixation can influence the duration of the current fixation (Drieghe, Brysbaert, & Desmet, 2005a; Inhoff, Starr, & Shindler, 2000b; Pynte, Kennedy, & Ducrot, 2004; Rayner, 1975; Starr & Inhoff, 2004), while other studies have found no such effects (Rayner et al., 2007a; White & Liversedge, 2004, 2006a).

Furthermore, some recent studies have suggested that the meaning of the word to the right of fixation can produce parafoveal-on-foveal effects (Kennedy, Murray, & Boissiere, 2004; Kennedy & Pynte, 2005; Kennedy, Pynte, & Ducrot, 2002; Kliegl, 2007; Kliegl, Nuthmann, & Engbert, 2006; Pynte & Kennedy, 2006). Yet, other studies have shown inconsistent (Hyönä & Bertram, 2004) or no parafoveal-on-foveal effects due to word frequency (Calvo & Meseguer, 2002; Henderson & Ferreira, 1993; Rayner, Fischer, & Pollatsek, 1998a; Schroyens, Vitu, Brysbaert, & d’Ydewalle, 1999; White, 2008). Other studies have likewise shown no evidence of lexical parafoveal-on-foveal effects (Altarriba et al., 2001; Hyönä & Häikiö, 2005; Inhoff, Starr, & Shindler, 2000b; Rayner, 1975; Rayner et al., 2007a).

Are parafoveal-on-foveal effects real or are there other reasons why such effects sometimes appear in the eye movement record? There are two possible (and reasonable) explanations for parafoveal-on-foveal effects that do not assume that such effects are real. First, some fixations in reading are mislocated because saccades are not perfectly accurate and do not land on the intended target (McConkie, Kerr, Reddix, & Zola, 1988; Nuthmann, Engbert, & Kliegl, 2005); thus, parafoveal-on-foveal effects may arise due to inaccurately targeted saccades (Drieghe, Rayner, & Pollatsek, 2008b; Rayner, Warren, Juhasz, & Liversedge, 2004b). That is, some saccades that are meant to land on a given target word fall short of the target and land on the end of the previous word. However, in this scenario, attention is still allocated to the originally intended saccade target word such that processing of the target word influences the fixation on the previous word. Second, the studies that have typically reported evidence of parafoveal-on-foveal effects have largely been based on analyses of large data sets across a corpus of data (Kennedy & Pynte, 2005; Kliegl, 2007; Kliegl et al., 2006; Pynte & Kennedy, 2006), while those that have found no evidence for lexical parafoveal-on-foveal effects are based on experimental studies that provide greater control over other variables (see Rayner, Pollatsek, Drieghe, Slattery, & Reichle, 2007d).

At this point, there seems to be some converging agreement concerning the validity of orthographic parafoveal-on-foveal effects. However, there is controversy concerning the validity of lexical parafoveal-on-foveal effects (Rayner & Juhasz, 2004; Rayner, White, Kambe, Miller, & Liversedge, 2003d; Starr & Rayner, 2001; White, 2008). Given the possibility of mislocalized fixations and the fact that most of the positive evidence for these effects is based on corpus-based analyses, it seems reasonable to view such effects with caution (Rayner et al., 2007d; White, 2008).

**The control of eye movements in reading**

There are two components to the issue of eye movement control: What determines *when* to move the eyes, and what determines *where* to move? As noted above, across large segments of text, there is typically no correlation between how long the eyes remain fixated and how far they move (Rayner & McConkie, 1976). This has generally been taken to suggest that these

---

8 A number of studies using either multiple isolated word-processing tasks, in which participants must look at, for example, three words in succession or tasks that mimic reading, have also reported evidence for parafoveal-on-foveal effects (Kennedy, 2000; Kennedy et al., 2004; Kennedy et al., 2002).
two decisions are made somewhat independently. Furthermore, there is compelling evidence for this view. Specifically, Rayner and Pollatsek (1981) varied the physical aspects of the text and found that properties of an eye movement mirrored aspects of the available display. First, they used the moving-window paradigm and varied the size of the window from fixation to fixation. They found that saccade length varied as a function of the immediately preceding window size. Thus, when the window was small, saccade size was shorter than when the window was large. Second, the text was delayed at the onset of a fixation by a mask (with the time of the delay varying randomly from fixation to fixation). Rayner and Pollatsek found that a large percentage of fixation durations varied according to the delay. From this, they argued that most fixations in reading are under direct cognitive control, though there was also a subset of fixations that appeared to be preprogrammed (see also Morrison, 1984). Importantly, the manipulations affected saccade length and fixation duration independently, reinforcing the view that the decisions about where and when to move are made somewhat independently.

In general, it is clear that the decision of where to move next is largely driven by low-level properties of the text while the decision of when to move the eyes is largely driven by lexical properties of the fixated word (Rayner, 1998). However, as discussed below, lexical effects (specifically related to word predictability and word frequency) have some influence on where the eyes move (and considerable influence on how long they remain fixated).

Where to move the eyes. For English and other alphabetic languages, where to move the eyes next is strongly influenced by low-level cues provided by word length and space information. Thus, saccade length is influenced by the length of the fixated word and the word to the right of fixation (Inhoff, Radach, Eiter, & Juhasz, 2003; Juhasz et al., 2008; O'Regan, 1979, 1980; Rayner, 1979; White et al., 2005b). If the word to the right of fixation is either very long or very short, the next saccade will be longer than when a medium-size word is to the right of fixation. For example, if the 11 character spaces to the right of the fixated word consisted of two 5-letter words (with a space between) or a single 11-letter word, the saccade will be longer in the latter case (Juhasz et al., 2008; Rayner, 1979; White et al., 2005b). If there was a short word (2 to 4 letters) to the right of fixation, the next saccade would tend to be longer than when the next word is 5 to 7 letters, largely because the short word would be skipped (Juhasz et al., 2008).

It is also clear that the spaces between words (which demarcate how long words are) are used in targeting where the next saccade will land. When spaces are removed, reading slows down by as much as 30–50% (Morris, Rayner, & Pollatsek, 1990; Pollatsek & Rayner, 1982; Rayner et al., 1998a; Rayner & Pollatsek, 1996; Spragins, Lefton, & Fisher, 1976). Of course, spaces between words are not present in all writing systems. Interestingly, Kohsom and Gobet (1997) demonstrated that when space information was provided for readers of Thai (who are not used to reading with spaces between words), they read more effectively than normal. Also, work with three-lexeme compound words in German has shown that inserting spaces between the lexemes actually reduces overall reading time on the compounds (Inhoff, Radach, & Heller, 2000a). In the experiment by Inhoff et al., the interword spaces were more beneficial than other manipulations that were used to mark lexeme boundaries (e.g., capitalizing the first letter of each lexeme). More recently, Bai, Yan, Liversedge, Zang, and Rayner (2008) inserted spaces between Chinese words or

---

9 Although most research has indicated that word length is an important cue in deciding where to look next, Epelboim, Booth, and Steinman (Epelboim, Booth, Askkenazy, Taleghani, & Steinman, 1997; Epelboim, Booth, & Steinman, 1994, 1996) argued that word length per se is not a critical cue for eye guidance. Indeed, they claimed that reading unspaced text is "relatively easy". However, analyses by Rayner and Pollatsek (1996) demonstrated that even in their experiments, most readers slowed down when reading unspaced text.
between Chinese characters (for a similar study with Japanese, see Sainio, Hyöna, Bingushi, & Bertram, 2007). Whereas inserting spaces between characters interfered with reading, inserting spaces between words did not. Actually, it is quite surprising that the insertion of spaces between words was not interfering given that the Chinese readers have a lifetime of experience reading without spaces. All of these pieces of evidence suggest that even when interword spaces are orthographically illegal, they are beneficial to reading.

Although low-level visual information influences where readers move their eyes, semantic information does not. While word predictability and word frequency have influences (discussed below), semantic preprocessing of words does not influence where readers move their eyes. It was reported (Underwood, Bloomfield, & Clews, 1988; Underwood, Hyöna, & Niemi, 1987) that readers move their eyes further into a word when the informative information is at the end of the word than when it is at the beginning and that semantic preprocessing was responsible for the effect. However, neither Rayner and Morris (1992) nor Hyöna (1995) replicated the effect. If readers processed the meaning of the end of an upcoming word it would involve rather complex processing. Thus, the failure to find effects is not surprising.

Landing position effects. The spaces between words provide information about an upcoming word’s length in parafoveal vision leading to systematic tendencies with respect to where the eyes typically land. Rayner (1979) demonstrated that readers’ eyes tend to land halfway between the middle of a word and the beginning of that word, the preferred viewing location (PVL). It is generally argued that readers attempt to target the centre of words, but their saccades tend to fall short (McConkie et al., 1988; Rayner, 1979). When readers’ eyes land at a nonoptimal position in a word, they are more likely to refixate that word (O’Regan, 1990; Rayner et al., 1996). Experiments using the boundary change paradigm, which provided readers with an incorrect length preview of an upcoming word in the parafovea, have demonstrated that when readers send their eyes to what will turn out to be a nonoptimal position in the parafoveal word there will be an increase in reading time on the word once fixated (Inhoff et al., 2003; Juhasz et al., 2008; White et al., 2005b).

Where readers fixate in a word can be viewed not only as a landing site for that word but also as the launch site for the next saccade. Although the PVL in a word lies between the beginning and the middle of a word, this position varies as a function of the prior launch site. Thus, if the launch site for a saccade landing on a target word is far from that word (say 8 to 10 letter spaces), the landing position will be shifted to the left. Likewise, if the distance is small (2–3 letter spaces), the landing position is shifted to the right. Thus, the landing site distribution on a word depends on its launch site (McConkie et al., 1988; Rayner et al., 1996).

The inverted optimal viewing position effect. In contrast to the PVL, which represents where readers fixate in words, the optimal viewing position (OVP) represents the location in a word at which recognition time is minimized; the best place to be in a word to recognize it most quickly is at the centre of the word. The OVP effect has been examined in the context of isolated word recognition studies, in which eye movements were monitored (O’Regan & Jacobs, 1992; O’Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984), and two general effects have been reported. First, there is a refixation effect such that the further the eyes are from the optimal viewing position the more likely it is that a refixation will be made on the word. Second, there is a processing cost effect such that for every letter that the eyes deviate from the optimal viewing position, there is a cost of roughly 20 ms (O’Regan et al., 1984). Interestingly, however, although the refixation effect remains in
reading (as opposed to isolated word recognition), the processing cost is either greatly attenuated or absent (Rayner et al., 1996; Vitu, O’Regan, & Mittau, 1990). This finding is consistent with the view that either contextual information in reading overrides low-level visual processing or the information acquired about a word before it is directly fixated affects its later processing.

With respect to the OVP effect, another interesting finding has recently been documented. Specifically, and somewhat counterintuitively, when readers make only one fixation on a word, if that fixation is at the centre of the word (which is the optimal viewing position), then the fixation is longer than when a single fixation is at the end of the word. This effect, the inverted optimal viewing position (IOVP) effect, was first documented by Vitu, McConkie, Kerr, and O’Regan (2001), and possible reasons for the effect have been examined by Vitu, Lancelin, and Marrier d’Unienville (2007). Perhaps the best explanation for the effect has been put forward by Nuthmann, Engbert, and Kliegl (2005, 2007) who suggested that mislocated fixations are the primary source of the effect (see also Pollatsek et al., 2006d). The basic nature of this argument is that many single fixations falling on the beginnings or ends of words are not on the targeted word (though the targeted word is being processed), but are due to overshoots (for fixations falling on the beginning of words) and undershoots (for fixations falling at the end of a word) of the oculomotor system. Via clever modelling techniques, Nuthmann et al. showed how mislocalized fixations could account for this effect. Perhaps more interestingly, however, it is the case that there are also word frequency effects independent of where the eyes land in a word when single fixations are made (Rayner et al., 1996; Vitu et al., 2001). Thus, although single fixations are longer when they fall in the middle of a word than at the ends, in both cases low-frequency words receive longer fixations than high-frequency words.

**Skipping effects.** As noted earlier, words are sometimes skipped during reading. Obviously, skipped words must be processed in parfoveal vision (where stimuli are degraded by acuity limitations), which also reduces the speed of processing of these words (Rayner & Morrison, 1981). Two factors have a big impact on skipping: word length and contextual constraint. First, the most important variable in skipping is word length (Brysbaert et al., 2005; Drieghe, Brysbaert, Desmet, & De Baecke, 2004; Drieghe, Desmet, & Brysbaert, 2007; Rayner, 1998): Short words are much more likely to be skipped than long words. When two or three short words occur in succession, there is a good chance that two of them will be skipped. And short words (like the) preceding a content word are often skipped (Drieghe, Pollatsek, Staub, & Rayner, 2008a; Gautier, O’Regan, & LaGargasson, 2000; Radach, 1996).

Second, words that are highly constrained by the prior context are much more likely to be skipped than those that are not predictable (Balota et al., 1985; Binder et al., 1999; S. F. Ehrlich & Rayner, 1981; Rayner & Well, 1996; Schustack, Ehrlich, & Rayner, 1987; Vitu, 1991). Chinese words that are predictable from prior context are also skipped more than those that are not (Rayner, Li, Juhasz, & Yan, 2005). Word frequency also has an effect on word skipping, but the effect is smaller than that of predictability (Rayner et al., 1996). While predictability influences whether or not a word is skipped, it does not influence where in the word the fixation lands (Rayner, Binder, Ashby, & Pollatsek, 2001a; Vainio, Hyöna, & Pajunen, 2009), though (see below) predictability does influence how long readers look at a word.

It is a mistake to think that if a word is skipped it is not processed. Fisher and Shebilske (1985) demonstrated this by examining the eye movements of readers on a passage of text. They then deleted all words from the passage that these readers had skipped and asked a second group of readers to read the passage. This second group of readers had a difficult time understanding the text. So, skipped words do get processed. But, when are they processed? There is evidence to suggest that when a word is skipped, it is processed on the fixation prior to or after the skip. Thus, the
fixation prior to skipping has been shown to be inflated in comparison to when a given target word is not skipped (Kliegl & Engbert, 2005; Pollatsek, Rayner, & Balota, 1986; Rayner et al., 2003a), as has the fixation after the skip (Reichle, Rayner, & Pollatsek, 2003).

**When to move the eyes.** Over the past few years, it has become very clear that the ease or difficulty associated with processing the fixated word strongly influences when the eyes move (Liversedge & Findlay, 2000; Rayner, 1998; Starr & Rayner, 2001). Thus, fixation time on a word is influenced by a host of lexical and linguistic variables11 such as word frequency (Inhoff & Rayner, 1986; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner & Duffy, 1986; Rayner et al., 1996; Schilling, Rayner, & Chumbley, 1998; Slattery, Pollatsek, & Rayner, 2007; Vitu, 1991; White, 2008), word predictability (Balota et al., 1985; Drieghe et al., 2005a; S. F. Ehrlich & Rayner, 1981; Kliegl et al., 2004; Rayner & Well, 1996; Zola, 1984), number of meanings (Binder & Morris, 1995; Binder & Rayner, 1998; Dopkins, Morris, & Rayner, 1992; Duffy, Morris, & Rayner, 1988; Folk & Morris, 2003; Kambe, Rayner, & Duffy, 2001; Rayner, Cook, Juhasz, & Frazier, 2006b; Rayner & Frazier, 1989; Sereno, O’Donnell, & Rayner, 2006), age of acquisition (Juhasz & Rayner, 2003, 2006), phonological properties of words (Ashby, 2006; Ashby & Clifton, 2005; Folk, 1999; Jared, Levy, & Rayner, 1999; Rayner, Pollatsek, & Binder, 1998b; Sereno & Rayner, 2000; Slattery, Pollatsek, & Rayner, 2006), semantic relations between the fixated word and prior words (Carroll & Slowiaczek, 1986; Morris, 1994), and word familiarity (Chaffin, Morris, & Seely, 2001; R. S. Williams & Morris, 2004). The effect of word frequency (Yan, Tian, Bai, & Rayner, 2006) and word predictability (Rayner et al., 2005) on fixation times on words also holds for skilled readers of Chinese. It is also interesting that when viewers are presented passages of text and are asked to search for a target word in the text, the frequency effect disappears (Rayner & Fischer, 1996; Rayner & Raney, 1996). This is consistent with the view that what influences when to move the eyes during reading is different from visual search.

A number of other variables have also been examined with respect to processing time on words. McDonald and Shillcock (2003a, 2003b) reported that the transitional probability between two words (e.g., *defeat* following *accept* has a high transitional probability whereas *losses* following *accept* has a low transitional probability) influences how long readers look at a word. However, Frisson, Rayner, and Pickering (2005) subsequently found that the difference was due to the predictability of the target word. When predictability was controlled, there was no difference between words with high and low transitional probabilities. Miller, Juhasz, and Rayner (2006) examined the extent to which the location of the uniqueness point in a word (the point, moving from left to right, at which a given target word could only be that word) influenced fixation times and found no differences between early and late uniqueness points in terms of how long readers look at words. Thus, these two extralexical variables (transitional probability and uniqueness point) do not seem to influence the amount of time readers look at words. On the other hand, morphological properties of words do influence fixation times on target words. Thus, the frequency of the different lexemes in longer compound words (which overall tend to be low-frequency words) influences fixation time on each lexeme. Specifically, a high-frequency beginning or ending lexeme in a compound word is fixated for a shorter duration than are low-frequency lexemes (S. Andrews, Miller, & Rayner, 2004; Bertram & Hyöna, 2003; Bertram, Pollatsek, & Hyöna, 2004; Hyöna et al., 2004; Hyöna & Pollatsek, 1998; Juhasz, 2007, 2008; Juhasz, 2008).

---

11 It is not possible to cite all of the many studies demonstrating the effects that are described here. Thus, in general the articles originally documenting the demonstrations and some recent demonstrations of the effects are listed (with an emphasis on studies from my lab).
Inhoff, & Rayner, 2005; Juhasz, Starr, Inhoff, & Placke, 2003; Pollatsek & Hyöna, 2005; Pollatsek, Hyöna, & Bertram, 2000a). And, as noted earlier, the amount of preview benefit for lexemes comprising the second half of a compound word is larger, around 100 ms, than the preview benefit obtained across words (Hyöna et al., 2004; Juhasz et al., 2009). Thus, Juhasz et al. found that there was more preview benefit when a correct preview (ball) was provided (in the boundary paradigm) for the lexeme at the end of a compound word (basketball) than when a partial preview (badk) was provided. More critically, there was more preview benefit for the lexeme ball in basketball than in tennis ball.

Experiments examining the influence of neighborhood frequency on eye movements (Perea & Pollatsek, 1998; Pollatsek, Perea, & Binder, 1999) have generally found that the effects typically do not occur in first-pass reading-time measures but in later measures (like regressions and spillover times). Rayner, White, Johnson, and Liversedge (2006e) and White, Johnson, Liversedge, and Rayner (2008) examined eye movements when reading words with jumbled letters and found that while it was fairly easy to read such text, there was always a cost associated with transposing the letters.

In short, it is the case that some variables have strong influences immediately when a word is fixated (such as frequency, age of acquisition, and predictability), while other variables seem to yield later occurring effects. However, there is no doubt that cognitive processing activities have a strong influence on when the eyes move. Perhaps the most compelling evidence that cognitive processing of the fixated word is driving the eyes through the text comes from experiments in which the fixated word either disappears or is masked after 50–60 ms (Ishida & Ikeda, 1989; Liversedge et al., 2004; Rayner et al., 1981; Rayner, Liversedge, & White, 2006c; Rayner, Liversedge, White, & Vergilino-Perez, 2003b). Basically, these studies show that if readers are allowed to see the fixated word for 50–60 ms before it disappears, they read quite normally. This does not mean that words are completely processed in 50–60 ms, but rather that this amount of time is sufficient for the processing system to encode the word. Interestingly, if the word to the right of fixation also disappears or is masked, then reading is disrupted (Rayner et al., 2006c); this quite strongly demonstrates that the word to the right of fixation is very important in reading. More critically, when the fixated word disappears after 50–60 ms, how long the eyes remain in place is determined by the frequency of the word that disappeared: If it is a low-frequency word, the eyes remain in place longer (Rayner et al., 2003b, 2006c). Thus, even though the word is no longer there, how long the eyes remain in place is determined by that word’s frequency. This is very compelling evidence that the cognitive processing associated with a fixated word is the engine driving the eyes through the text.

It is thus quite clear that lexical variables have strong and immediate effects on how long readers look at a word. While other linguistic variables can have an influence on how soon readers move on in the text, it is generally the case that higher level linguistic variables have somewhat later effects, unless the variable more or less “smacks you in the eye”. So, for example, when readers fixate on the disambiguating word in a syntactic garden path sentence there is increased fixation time on the word (Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983; Rayner & Frazier, 1987) and/or a regression from the disambiguating word back to earlier parts of the sentence (Frazier & Rayner, 1982; Meseguer, Carreiras, & Clifton, 2002; Mitchell et al., 2008). Readers also have longer fixations at the end of clauses and sentences (Hirotani, Frazier, & Rayner, 2006; Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). And, when readers encounter an anomalous word, they fixate on it longer, and the effect is quite immediate (Rayner et al., 2004b; Staub, Rayner, Pollatsek, Hyöna, & Majewski, 2007; Warren & McConnell, 2007); when a word indicates an implausible, but not truly anomalous, event, there will be an effect registered in the eye movements and attention.
movement record, but it is typically delayed a bit, showing up in later processing measures (Joseph et al., 2008; Rayner et al., 2004b).

Interestingly, when sentences with an anomalous word (such as *carrot* in “Jane used a pump to inflate the large carrot”) are embedded in cartoon or fantasy-like contexts and are compared with real-world contexts where inflating a carrot with a pump is not anomalous (Warren, McConnell, & Rayner, 2008), the earliest measures of fixation time (first-fixation duration and gaze duration) still yielded longer fixations on the anomalous word than the control word (*carrot* in a sentence like “Jane used a knife to chop the large carrot”). However, the go-past measure revealed disruption only in the real-world context. These results suggest that contextual information did not eliminate the initial disruption, but moderated it quickly thereafter.\(^{12}\)

**Using eye movements to study sentence and discourse processing.** In much of the foregoing discussion, the premise has largely been that lexical processing is the engine driving the eyes through the text. However, there is good reason to believe that higher order comprehension processes influence eye movements primarily when something does not compute (Clifton, Staub, & Rayner, 2007; Staub & Rayner, 2007). When readers are garden-pathed by syntactic ambiguity (Binder, Duffy, & Rayner, 2001; Boland & Blodgett, 2001; Clifton et al., 2003; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner et al., 1983; Rayner & Frazier, 1987; Rayner, Garrod, & Perfetti, 1992) their fixations get longer, and they often make shorter saccades and more regressions (Altmann, Garnham, & Dennis, 1992; Rayner & Sereno, 1994). In cases such as this, higher order comprehension processes can override the normal default situation in which lexical processing is driving the eyes and result in longer fixations or regressions back to earlier parts of the text.

It is interesting that eye movement data have more or less become the gold standard in experiments, which are too numerous to list here, dealing with sentence processing and syntactic ambiguity resolution. Because of its precise temporal properties, eye tracking is generally deemed to be the preferred way to study online sentence processing. In contrast, it is striking that there have not been nearly as many studies utilizing eye movement data to examine online comprehension and discourse-processing effects. While there are a few studies (see, for example, Birch & Rayner, 1997; Cook & Myers, 2004; Duffy & Keir, 2004; Duffy & Rayner, 1990; K. Ehrlich & Rayner, 1983; Garrod, O’Brien, Morris, & Rayner, 1990; Garrod & Terras, 2000; O’Brien, Raney, Albrecht, & Rayner, 1997; O’Brien, Shank, Myers, & Rayner, 1988; Rayner, Chace, Slattery, & Ashby, 2006a; Sturt, 2003) in which eye movements were monitored to assess immediate comprehension in discourse processing, the number of such studies pales in comparison to those that used more gross reading-time measures. The time may be ripe for more studies to use eye movement data to understand moment-to-moment discourse processing.

**Models of eye movement control in reading**

Given the vast amount of information about eye movements during reading that has accumulated in the past 25–30 years, it is not surprising that a number of models of eye movements in reading have recently appeared. The E-Z Reader model (Pollatsek, Reiche, & Rayner, 2006d; Rayner, Ashby, Pollatsek, & Reiche, 2004a; Rayner, Pollatsek, & Reiche, 2003c; Rayner, Reiche, & Pollatsek, 1998c; Reiche, Pollatsek, Fisher, & Rayner, 1998; Reiche, Pollatsek, & Rayner, 2006; Reiche, Rayner, & Pollatsek, 1999; Reiche et al., 2003) is typically regarded as the most influential of these models. Due to space limitations, other models are not discussed in length here.\(^{13}\) Other

---

\(^{12}\) See also Filik (2008) for slightly different data on this issue, and Ferguson and Sanford (2008) for data consistent with Warren et al. (2008).

\(^{13}\) For a comprehensive overview of these models, see the special issue of *Cognitive Systems Research* (Reichle, 2006), and see Reichle et al. (2003) for a comparison of the models. See also Reichle and Laurent (2006) for a model describing how learning influences eye movements.
models include SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005), SHARE (Feng, 2006), Glenmore (Reilly & Radach, 2006), EMMMA (Salvucci, 2001), SERIF (McDonald, Carpenter, & Shillcock, 2005), Mr. Chips (Legge, Hooven, Klitz, Mansfield, & Tjan, 2002; Legge, Klitz, & Tjan, 1997), and the competition/activation model (S. Yang, 2006; Yang & McConkie, 2001). These models are all fully implemented, but they differ on a number of dimensions. Mr. Chips is an ideal observer type of model, whereas all of the others strive to account for the actual performance of readers. In some of the models the eyes are driven by lexical processing whereas in others eye movements are largely viewed as being primarily influenced by oculomotor constraints. Some models allow for (lexical) parallel processing of words, whereas in others lexical processing is serial so that the meaning of word $n + 1$ is not accessed until the lexical processing is complete (or nearly complete) for word $n$.

These models share many similarities, but they differ on some precise processing details and on how certain effects are explained. E-Z Reader has the most lexical involvement, and the competition/activation model has the least amount. E-Z Reader involves serial lexical processing of words, whereas SWIFT and Glenmore allow for parallel lexical processing of words. It is generally agreed that lexical processing has to have a strong influence on the decision of when to move the eyes in order to account for much of the data described above.

E-Z Reader does a good job of predicting how long readers look at words, which words they skip, and which words will most likely be refixedated. Importantly, it can account for global aspects of eye movements in reading, as well as more local processing characteristics (competitor models also account for much of the data). As a computational model, E-Z Reader has the virtue of being highly transparent, so it makes very clear predictions, and when it cannot account for certain data it is very clear why it cannot (thus enabling one to modify the model). The model has also generated interesting research questions (Drieghe, 2008; Inhoff, Eiter, & Radach, 2005; Inhoff, Radach, & Eiter, 2006; Miellet, Sparrow, & Sereno, 2007; Pollatsek, Reichle, & Rayner, 2006b, 2006c; Rayner et al., 2004a; Reingold & Rayner, 2006) because it makes clear predictions. Also, changes have been made to the model when resulting data have necessitated it (Rayner et al., 2004a). Furthermore, the model has also made it possible to account for data patterns that in the past may have been difficult to explain (see Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008; Rayner et al., 2004a, Rayner et al., 2006d, for examples). However, the model is not perfect, and it has some limitations. For example, in the model as initially developed, higher order processes due to sentence parsing and discourse variables had no influence, and, as noted above, higher order processes can at times influence when the eyes move and how long they remain in place. As a first step in accounting for such nonlexical variables, a new version of the model (E-Z Reader 10) strives to account in part for these effects (Reichle, Warren, & McConnell, 2009). E-Z Reader has also been extended to account for eye movement data of elderly readers (Rayner et al., 2006d), as has SWIFT (Laubrock et al., 2006), and for eye movement data of Chinese readers (Rayner, Li, & Pollatsek, 2007b).

With careful experimentation and with the implementation of computational models that simulate eye movements during reading, great advances have been made in understanding eye movements in reading (and inferring the mental processes associated with reading). In the next two sections, eye movements during scene perception and visual search are discussed. Although there has not been as much eye movement research on these areas as on reading, it is still the case that some

---

14 A recent neurophysiologically inspired model (Heinzle, Martin, & Hepp, 2009) has many of the same properties as those inherent in E-Z Reader, including serial lexical processing.
clear conclusions emerge from the work that has been done.

**Eye movements in scene perception**

Fixation durations during scene perception tend to be longer than those in reading, and saccade size tends to be larger. The average fixation duration tends to be closer to 300 ms (but varies as a function of the task and the characteristics of the scene). Average saccade size tends to be 4–5 deg (though it too can vary as a function of task and the exact nature of the scene¹⁵). Whereas there is a well-defined task for readers, exactly what participants should do in a scene perception task is more variable. Sometimes participants are asked to look at the scene in anticipation of a memory test while other times they may be asked to indicate if a certain object is present in the scene. Under the latter instructions, scene perception becomes very much a visual search task.

Examination of the eye movement pattern (or the scan path) of a viewer on a scene demonstrates that viewers do not fixate every part of the scene. Most fixations tend to fall on the informative parts of the scene. Thus, viewers tend to not fixate on the sky or the road in front of a house. Furthermore, viewers are able to obtain the gist of the scene in a single glance. That is, the gist is understood so quickly that it is obtained even before the eyes begin to move (De Graef, 2005). Castelhano and Henderson (2008b) found that when viewers are shown a scene for as little as 40 ms, they are able to extract enough information to understand the scene gist. Thus, the gist is thought to be acquired during the first fixation in order to orient subsequent fixations to appropriate/interesting regions in the scene. Actually, it is not well understood how the gist is acquired so rapidly, and this issue requires further research.

While the gist of the scene can pretty much be obtained on the first fixation, how much information do viewers obtain on each fixation as they look around a scene? It is clear that information is acquired over a wider range of the visual field in scene perception than is the case for reading. As with reading, the best way to address this issue is via gaze-contingent paradigms. Yet surprisingly few such studies have been reported. Henderson, McClure, Pierce, and Schrock (1997) used a moving-mask procedure (to cover the part of the scene around the fixation point) and found that although the presence of a foveal mask influenced looking time, it did not have nearly as serious effects for object identification as a foveal mask has for reading (Rayner & Bertera, 1979). Saida and Ikeda (1979) used a moving-window paradigm and found that the functional field of view is quite large and can consist of about half of the total scene regardless of the absolute size of the scene (at least for scenes that were up to 14.4 by 18.8 degrees). They found a serious deterioration in recognition of a scene when the window was limited to a small area (about 3.3 × 3.3 degrees) on each fixation. Performance gradually improved as the window size became larger up to about 50% of the entire scene. Saida and Ikeda noted that there was considerable overlap of information across fixations. In this study and other studies that have used the moving-window paradigm (Castelhano & Henderson, 2007; van Diepen & d’Ydewalle, 2003; van Diepen, Ruelens, & d’Ydewalle, 1999; van Diepen & Wampers, 1998) scene information was presented normally within the window area around a fixation point, but the information outside of the window was degraded in a systematic way.¹⁶

Other studies have attempted to address the question via other techniques. For example,

---

¹⁵ More densely packed scenes lead to longer fixations and shorter saccades.

¹⁶ A very promising variation on gaze-contingent moving-windows and moving-masks paradigms, discussed in the context of visual search, that has not yet been fully exploited is to use multiresolution displays (Loschky & McConkie, 2002; Reingold & Loschky, 2002; Reingold, Loschky, McConkie, & Stampe, 2003). With these types of display, a clear view of the scene can be provided around the fixation point with increasing degradation of the scene outside of the window.
Nelson and Loftus (1980) examined how close to fixation an object had to be for it to be recognized as having been in the scene. They found that objects located within about 2.6 degrees from fixation were generally recognized, but recognition depended to some extent on the characteristics of the object. They also suggested that qualitatively different information is acquired from the region within 1.5 degrees around fixation than from regions further away (see also Nodine, Carmody, & Herman, 1979). While a study by Parker (1978) suggested that the functional field of view for specific objects in a scene is quite large (with a radius of at least 10 degrees around fixation resulting in a perceptual span of up to 20 degrees), other more recent studies using better controlled stimuli and more natural images (Henderson & Hollingworth, 1999; Henderson, Williams, Castelhano, & Falk, 2003) suggest that the functional field of view only extends about 4 degrees away from fixation.

It appears that the answer to the question of how large the perceptual span in scene perception is has not been answered as conclusively as it has in reading. Viewers typically gain useful information from a fairly wide region of the scene, which also probably varies as a function of the scene and the task. For instance, the ease with which an object is identified has been linked to its orientation (De Graef, 2005), frequency within a scene context (Hollingworth & Henderson, 1998), and how well camouflaged it is (De Graef, Christiaens, & d’Ydewalle, 1990). As has been shown in reading (Henderson & Ferreira, 1990), it is likely that the ease of identifying a fixated object has an effect on the extent of processing in eccentric vision.

**Preview benefit in scene perception**

Just as in reading, viewers obtain preview benefit from objects that they have not yet fixated (Henderson, 1992; Henderson, Pollatsek, & Rayner, 1987, 1989; Henderson & Siefert, 1999, 2001; Pollatsek, Rayner, & Collins, 1984; Pollatsek, Rayner, & Henderson, 1990) with the amount of the preview benefit of the order of 100 ms; thus, preview benefit is larger in scene perception than in reading. Interestingly, as is now rather well known, viewers are rather insensitive to changes in scenes. Grimes and McConkie (1995) and McConkie and Currie (1996) asked observers to view scenes with the task of memorizing what they saw. They were also told that changes could be made to the image while they were examining it, and they were instructed to press a button if they detected the changes. While observers viewed the scenes, changes were made during a saccade (when vision is suppressed). Remarkably, observers were unaware of most changes, which included the appearance and disappearance of large objects and the changing of colours.

Henderson and Hollingworth (2003) likewise demonstrated that low-level sensory information is not preserved from one fixation to the next (see also Henderson, Brockmole, & Gajewski, 2008). Other studies found that viewers were unable to detect changes when there was a corresponding disruption to processing, such as with the simultaneous onset of patches covering portions of the scene (O’Regan, Rensink, & Clark, 1999). However, this lack of awareness of changes does not mean that there is no recollection of any visual details, but rather that the likelihood of remembering visual information is highly dependent on the processing of that information (Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). Thus, knowing about the processes that go on during a fixation is very important if one wants to predict how well information from a scene is stored.

Early theories of transsaccadic memory (Jonides, Irwin, & Yantis, 1982; McConkie & Rayner, 1976b) proposed that information is integrated across saccades in an integrative visual buffer (with properties like iconic memory). However, experiments described above in the context of reading (McConkie & Zola, 1979; Rayner et al., 1980a) as well as nonreading experiments using relatively simply arrays (Irwin, Yantis, & Jonides, 1983; O’Regan & Lévy-Schoen, 1983; Rayner & Pollatsek, 1983) demonstrated that this view is incorrect and that viewers do not integrate sensory information presented on
separate fixations in a visual buffer. More recent work with more naturalistic scenes has arrived at the same conclusion, and there is evidence that visual short-term memory (VSTM), which is thought to be at a higher level than a visual buffer, serves a primary role in integrating information across saccades (Hollingworth, Richard, & Luck, 2008). Thus, memory across saccades during scene perception appears to be due to higher level visual codes (Carlson-Radvansky, 1999; Carlson-Radvansky & Irwin, 1995; Hollingworth et al., 2008; Irwin, 1991, 1992; Irwin & Gordon, 1998), which are abstracted away from precise sensory representations, with VSTM as the basis for integration.

Eye movement control in scene perception

Where do viewers look in scenes? Since the pioneering work of Buswell (1935) and Yarbus (1967), it has been widely recognized that viewers’ eyes are drawn to important aspects of the visual scene and that their goals in looking at the scene very much influence their eye movements. Quite a bit of early research demonstrated that the eyes are quickly drawn to informative areas in a scene (Antes, 1974; Mackworth & Morandi, 1967). It is also clear that the saliency of different parts of the scene influence what part of the scene is fixated (Mannan, Ruddock, & Wooding, 1995, 1996; Parkhurst & Niebur, 2003). A large amount of empirical and computational research has recently been devoted to understanding the factors that govern fixation position in scenes (Foulsham, Kingstone, & Underwood, 2008; Henderson, 2003; Itti & Koch, 2000; Melcher & Kowler, 2001; Parkhurst, Law, & Niebur, 2002; Rutishauser & Koch, 2007; Tatler, Baddeley, & Vincent, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006), and much of this work revolves around how saliency (which is typically defined in terms of low-level components of the scene, such as contrast, colour, intensity, brightness, spatial frequency, etc.) influences where viewers look.

Saliency is not the only factor involved in determining where to look, and there are questions about how important it is (Foulsham & Underwood, 2008; Henderson, Brockmole, Castelhano, & Mack, 2007). It has also become increasingly clear that there are strong cognitive influences on where viewers look as well (see Henderson, 2007; Torralba, Oliva, Castelhano, & Henderson, 2006). Neider and Zelinsky (2006a) examined the influence that context has on the placement of eye movements in search of certain objects within pseudo-realistic scenes. Viewers were asked to look for target objects that are typically constrained to certain parts of the scene (i.e., jeep on the ground, hot air balloon in the sky). When a target was present, fixations were largely limited to the area where one would expect to find the target object (i.e., ground or sky), while, when the target was absent, search was less restricted to these areas. They also found that when the target was in the expected area, search times were on average 19% faster. From these results, it seems that not only do viewers focus their fixations in areas of a scene that most likely contain the target to improve search, but also they are flexible in the application of these restrictions, and they very quickly adopt a “look everywhere” strategy when necessary. Thus, it seems that search strategies in scenes are guided by the scene context, but not with strict adherence.

Henderson (1993) found that viewers tended to fixate near the centre of an object and that there was a greater tendency to undershoot the centre than to overshoot. Furthermore, landing position influenced fixation time as the duration of the first fixation on an object decreased, and the probability of refixating the object increased as the deviation of the initial landing position from the centre of the object increased. Also, Mannan et al. (1995, 1996) found that viewers tend to look in pretty much the same locations across three viewings of a scene even though the scenes had been either high-pass or low-pass filtered. Thus, low-level visual information must be critical in deciding where to look next.

What role does memory play in directing fixations in scenes? Many of the models using saliency as the primary driving force of eye movements do
not consider how information gathered initially may influence the placing of subsequent fixations. In a recent study, Castelhano and Henderson (2007) investigated whether this initial representation of a scene can be used to guide subsequent eye movements on a real-world scene. Viewers were shown a very short preview of the search scene and then were asked to find the target object. The moving-window technique was used in this phase of the study, thus eliminating any immediately available visual information. A preview of the search scene itself elicited the most efficient searches when compared to a meaningless control (the preview yielded fewer fixations and the shortest scan path to the target). When a preview of another scene within the same semantic category was shown (thereby providing general semantic information without the same visual details) there was no improvement in search. These results suggest that the initial representation used to improve search efficiency was not based on general semantic properties, but on something more specific. When a reduced scale of the search scene was shown as the preview, search efficiency measures were as high as when the full-scale preview was shown. Taken together, these results suggest that the initial scene representation is based on abstract, visual information that is useful across changes in spatial scales. Thus, the information used to guide eye movements has two sources: the saliency of the scene and the information in memory about that scene and scene type.

Are the eyes drawn to unusual parts of a scene? A somewhat contentious issue concerns the extent to which the eyes are drawn to highly informative, unusual, or emotional aspects of a scene; the evidence is somewhat uneven as some experiments indicate that they are, while others suggest they are not. Early experiments found that the eyes move quickly to an object that is out of place in a scene (Friedman, 1979; Loftus & Mackworth, 1978). Unfortunately, these studies did not control physical distinctiveness very well (Rayner & Pollatsek, 1992), and, when it was controlled, studies (Brocket & Henderson, 2008; De Graef, 1998; De Graef et al., 1990; Henderson, Weeks, & Hollingworth, 1999) failed to replicate the finding that semantically inconsistent objects were fixated earlier than consistent objects (but see Underwood & Foulsham, 2006; Underwood, Humphreys, & Cross, 2007).

More recently, Becker, Pashler, and Lubin (2007) and Harris, Kaplan, and Pashler (2008) renewed interest in the extent to which semantically incongruent objects in scenes attract the eyes. Unlike many earlier experiments that used line-drawings of scenes, Becker et al. used colour photographs, which viewers examined for 8 seconds. Some aspect of the scene was anomalous; for example, a stop sign was green instead of red. Unlike the earlier experiments, a large number of viewers looked at a rather limited number of scenes. Becker et al. found that viewers fixated anomalous items (the green stop sign) earlier (both in time and in order of eye fixations) than the control objects (the red stop sign) and argued that the results indicate that violations of canonical form can be detected from extrafoveal vision and can affect the likelihood of fixating them.

Harris et al. (2008) introduced emotional, yet somewhat unusual, aspects into scenes; for example, in a control scene people are tossing a beach ball at a beach, while in the emotional scene the beach ball is replaced by a baby. They found that viewers looked earlier at the emotional aspect of the scene. A number of other recent studies (Calvo & Lang, 2005; Calvo & Nummenmaa, 2007; Kirchner & Thorpe, 2006; Nummenmaa, Hyöna, & Calvo, 2006) have likewise reported that the eyes quickly move to emotional objects or scenes (though in these studies, the object/scene is usually presented in parafoveal vision, and the latency of a saccade from a central fixation point is measured for an emotional scene/object versus a neutral scene/object).

Rayner, Castelhano, and Yang (2009a) used the scenes from Harris et al. (2008) as well as a large number of other scenes with weird or unusual aspects. Like Becker et al. (2007) and Harris et al., they found that the eyes were drawn to the weird parts of the scene earlier than when the weird aspect was missing. Yet, the eyes being drawn to the weird part of the scene was not instantaneous;
it was fixated within 3 fixations (in comparison to the same part of the scene being fixated within 3.5 fixations in the control condition). Underwood, Templeman, Lamming, and Foulsham (2008) likewise reported that the incongruous objects attract attention earlier than congruous objects, but the effect was not apparent until the picture has been displayed for several seconds.

Are there cultural influences on where to look in scenes? It has recently been suggested that Asian (Chinese) participants look at scenes differently than English-speaking participants. Specifically, Chua, Boland, and Nisbett (2005) reported that Chinese viewers spent less time looking at the focal objects in a scene and more time looking at the scene background than their English-speaking counterparts. These results were discussed in the wider context of a general theory of cultural differences in cognition (Nisbett, 2003) whereby Asian cultures lead people to not place as much value on the individual as on the group, while the American culture places more emphasis on the individual. According to Chua et al., this underlying cultural difference in thinking led the Chinese viewers to look more at the background and spend relatively less time (in comparison to the Americans) looking at the focal objects.

However, three recent reports have raised some questions about the validity of Chua et al.’s (2005) findings. Rayner, Li et al. (2007b) reported no differences in the looking patterns of Chinese and American participants, with both groups looking more at focal objects than the background information. Boland, Chua, and Nisbett (2008) noted that the study was not a direct replication of Chua et al. given that the same materials were not used (and the focal objects were more apparent in their study than in the Rayner et al. study), and the task varied between the two studies (expectation of a memory test in Rayner et al. and a rating task for scene likeability in Chua et al.). However, Evans, Rotello, Li, and Rayner (2009) used the original scenes from Chua et al. (as well as additional scenes for increased power) and Chua et al.’s task, and also found no differences between the two groups (either with the entire set of stimuli or with the subset that had been previously used by Chua et al.). Furthermore, Rayner et al. (2009a) argued that if Chinese viewers paid more attention to the background information in a scene it might take them longer to notice the weird object in the scene. But, there was no difference between the Chinese and American viewers in their study.

When do viewers move their eyes when looking at scenes? Given that attention precedes an eye movement to a new location within a scene (Henderson, 1992; van Diepen & d’Ydewalle, 2003), it follows that the eyes will move once information at the centre of vision has been processed, and a new fixation location has been chosen. When this shift in attention (from the centre of fixation to the periphery) takes place in the course of a fixation was investigated by van Diepen and d’Ydewalle (2003). They had observers view scenes whose information at the centre of fixation was masked during the initial part of fixations (with the mask coming on 20–90 ms from fixation onset). In another case, the periphery was masked at the beginning of each fixation (for 10–85 ms). They found that when the centre of fixation was masked initially, fixation durations increased with longer mask durations. When the periphery was masked, they found a slight increase in fixation durations, but not as much as with a central mask. Interestingly, they found that the average saccades size decreased, and the number of fixations increased, with longer mask durations in the periphery. They argued that with the longer peripheral masking durations the visual system does not wait for the unmasking of peripheral information, but instead chooses information that is immediately available. These results suggest that the extraction of information at the fovea occurs fairly rapidly, and attention is then directed to the periphery to choose a viable saccade target almost immediately following the extraction of foveal information. The general timing of the switch between central and peripheral information processing needs further investigation, but it is likely that the variability of information across scenes will make it
more difficult to delineate a specific time frame as has been done in reading.

Recent studies by Henderson and Pierce (2008) and Rayner, Smith, Malcolm, and Henderson (2009b) are quite informative with respect to issues related to the timing of eye movements in scene perception. Henderson and Pierce (2008) adapted the paradigm used by Rayner and Pollatsek (1981; Morrison, 1984) to create a scene onset delay paradigm. In these experiments, a visual mask was presented at the beginning of eye fixations as viewers examined a scene. The duration of the mask was varied (with scene onset delays as short as 40 ms and as long as 1200 ms), and the scene did not appear until the designated mask duration was exceeded. Then the scene appeared and remained visible until the viewer made a saccade. Scene onset delays took place on every tenth fixation. As in reading, Henderson and Pierce found that one population of fixations was under direct control of the current scene, increasing in duration as the delay increased. However, a second population of fixations was relatively constant across delay. Also like reading, the data pattern did not change whether the scene delay duration was random or blocked, suggesting that the effects are not under the strategic control of viewers. The results support a mixed model of eye movement control and indicate that the durations of some fixations are determined regardless of scene presence while others are under direct moment-to-moment control of scene analysis. Interestingly, the percentage of fixations under direct control was much greater in reading than in scene perception.

Rayner et al. (2009b) adapted the disappearing text/masked text paradigm (Rayner et al., 1981, 2003b, 2006c) discussed above to create a situation in which a scene was masked at certain points after the onset of each new fixation. Interestingly, they found that viewers needed 150 ms to view the scene before the mask was not disruptive. Obviously, this is much longer than the 50–60 ms that was needed in reading for the mask to not cause disruption and also longer than one might predict given that viewers can obtain the gist of a scene on the first fixation.

Models of eye movement control in scene perception. A number of models of eye movement control in scene perception have recently appeared. For the most part, these models focus on where to move the eyes next, and little effort has been made to specify when the eyes move (or what influences the decision to move the eyes). A fair number of computational models (Baddeley & Tatler, 2006; Itti & Koch, 2000, 2001; Parkhurst et al., 2002) use the concept of a saliency map (following from Findlay & Walker, 1999) to model eye fixation locations in scenes. In this approach, bottom-up properties in a scene make explicit the locations of the most visually prominent regions of the scene. The models are basically used to derive predictions about the distribution of fixations on a given scene.

While these models can account for some of the variability in where viewers fixate in a scene, they are limited in that the assumption is that fixation locations are driven primarily by bottom-up factors, and it is clear that higher level factors also come into play in determining where to look next in a scene (Castelhano & Henderson, 2007). A model that includes more in the way of top-down and cognitive strategies was recently presented by Torralba et al. (2006). Indeed, as noted previously, while there has been a lot of research to localize where viewers move their eyes while looking at scenes, there has been precious little in the way of research attempting to determine what controls when the eyes move. This is in contrast with reading where the issues of where to move the eyes and when to move the eyes have both received considerable attention. Models of eye movement control in scene perception need to better explain the factors influencing when to move the eyes.

Eye movements and visual search

Visual search has received considerable attention over the past 40 years. However, the majority of this research has been done without measuring eye movements (Findlay & Gilchrist, 1998), and it has often been assumed that they are not particularly important in understanding search. However,
This attitude seems to be largely changing as many recent experiments have utilized eye movements to understand the process. Many of these studies deal with very low-level aspects of search and often focus on using the search task to uncover properties of the saccadic eye movement system (see Findlay & Gilchrist, 2003). But, it is becoming very clear that eye movement studies of visual search, like reading and scene perception, can provide important information on moment-to-moment processing in search (Trukenbrod & Engbert, 2007; Williams & Pollatsek, 2007; Zelinsky, Rao, Hayhoe, & Ballard, 1997).

Here the focus is primarily on research using eye movements to examine how viewers search through arrays to find specific targets. As noted at the outset, fixation durations in search tend to be highly variable. Some studies report average fixation times as short as 180 ms while others report averages on the order of 275 ms. This wide variability is undoubtedly due to the fact that how difficult the search array is (or how dense or cluttered it is) and the exact nature of the search task strongly influence how long viewers pause on average. Typically, saccade size is a bit larger than that in reading (though saccades can be quite short with dense arrays) and a bit shorter than that in scene perception. Two important points regarding eye movements during search are first discussed. Then a brief review of (a) the perceptual span, (b) preview benefit, (c) eye movement control, and (d) models of eye movements is presented.

The search array matters. Perhaps the most obvious thing about visual search is that the search array makes a big difference in how easy it is to find a target. When the array is cluttered and/or dense (with many objects and/or distractors) search is more costly than when the array is simple (or less dense), and eye movements typically reflect this fact (Bertera & Rayner, 2000; Greene & Rayner, 2001a, 2001b). The number of fixations and fixation duration both increase as the array becomes more complicated, and the average saccade size decreases (Vlaskamp & Hooge, 2006). Additionally, the configuration of the search array has an effect on the pattern of eye movements. In an array of objects arranged in an arc, fixations tend to fall in between objects, progressively getting closer to the area where viewers think the target is located (Zelinsky & Loschky, 2005; Zelinsky et al., 1997). On the other hand, in randomly arranged arrays, other factors such as colour of the items and shape similarity to the target object influence the placement of fixations (Williams, Henderson, & Zacks, 2005).

Does visual search have a memory? Horowitz and Wolfe (1998) initially proposed that during visual search viewers do not have good memory and that the same item will be resampled during the search process. This question has provoked a considerable amount of research. However, Horowitz and Wolfe made this assertion based on reaction time functions, and eye movement data are ideal for addressing the issue (since one can examine how frequently the eyes return to a previously sampled part of the array). Indeed, a number of eye movement experiments (Beck, Peterson, Boot, Vomela, & Kramer, 2006a; Beck, Peterson, & Vomela, 2006b; Geyer, von Mühlener, & Müller, 2007; Peterson, Kramer, Wang, Irwin, & McCarley, 2001) make it quite clear that viewers generally do not return to previously searched items.

The perceptual span in visual search Rayner and Fisher (1987a, 1987b) used the moving-window paradigm as viewers searched through horizontally arranged letter strings for a specified target letter. They found that the size of the perceptual span varied as a function of the difficulty of the distractor letters; when the distractor letters were visually similar to the target letter, the size of the perceptual span was smaller than when they were distinctly different from the target letter. They suggested that there were two qualitatively different regions within the span: a decision region (where information about the presence or absence of a target is available), and a preview region where some letter information is available but where information on the absence of a target is not available.
Bertera and Rayner (2000) had viewers search though a randomly arranged array of letters and digits for the presence of a target letter. They used both the moving-window and moving-mask techniques. They varied the size of the array (so that it was 13 by 10 deg, 6 by 6 deg, or 5 by 3.5 deg), but the number of items was held constant (so in the smaller arrays, the information was more densely packed). Mask size was 0.3, 1, 1.7, 2.3, or 3 deg; window size was 1, 2.3, 3.7, 5, and 5.7 deg. There were also control conditions in which no mask or window was present. Not surprisingly, the moving mask had a deleterious effect on search time and accuracy, and the larger the mask, the longer the search time, the more fixations were made, and the longer were the fixations; saccade size was affected by array size, but mask size had little effect. In the moving-window condition, search performance reached asymptote when the window was 5 deg (all letters/digits falling within 2.5 deg from the fixation point were visible with such a window size while all other letters were masked).

In a study that is conceptually very similar to that of Bertera and Rayner (2000), Cornelissen, Bruin, and Kooijman (2005) asked viewers to search for the letter O among distractors (Cs, with the orientation of the opening randomly varied). The search arrays consisted of a 7 × 5 hexagonal matrix (three rows of seven and two rows of six objects) containing 32 Cs (distractors) and a single O (the target). The overall array size was 38 × 28 deg, and the objects were 5 deg apart. Mask size and window size was 5, 10, or 15 deg. Like Bertera and Rayner and other studies that have used gaze-contingent displays (see also Greene, 2006; Greene & Rayner, 2001b; Pomplun, Reingold, & Shen, 2001), Cornelissen et al. found that search times, fixation durations, and number of fixations all became larger as mask size increased and as window size became smaller. However, saccade length appeared to be more strongly affected by window and mask size manipulations in the Cornelissen et al. study than in the Bertera and Rayner study. Given that the arrays differed between Bertera and Rayner’s study and Cornelissen et al. (they were more structured and larger in the latter study), it is difficult to make generalizations regarding any differences in results. What is certain from the two studies is that, as noted earlier, array size matters in search, and foveal masks (which mimic central scotomas) provide greater disruption to search than moving windows (which create tunnel vision).

Finally, other recent studies have investigated the perceptual span via gaze-contingent multi-resolution moving windows (Geisler, Perry, & Najemnik, 2006; Loschky, McConkie, Yang, & Miller, 2005). Within this paradigm, information outside the window is degraded in a manner that simulates resolution degradation (i.e., blurring) at various eccentricities from an observer’s area of fixation. The eccentricity in degrees at which display resolution drops to one half of its value at the fixation point is termed $e_2$. In a multiresolution display, the value of $e_2$ controls the extent of blurring into the parafovea, such that the smaller the value of $e_2$, the steeper the drop-off in resolution. Findings from these studies suggest that during any single fixation, when $e_2$ is about 6 deg, the parafoveal blur imposed on a scene is not detectable. Thus, viewers do not notice that the scene has been artificially blurred. Consistent with other studies discussed here dealing with the use of information beyond the point of fixation, even when artificial blurring went undetected in eccentric vision, search performance was affected.

**Preview benefit**

It is undoubtedly the case that viewers obtain preview benefit during search (and the issue is probably related to whether or not there is memory for items in search discussed earlier). Typically, studies of preview benefit in visual search provide a viewer with a preview of the search array (or part of the array) for a set period of time (such as 500 ms), or no preview in a control condition. Then the array is presented in its entirety. Generally, it is found (Watson & Inglis, 2007) that there are fewer fixations on previewed stimuli (and if they are fixated, for shorter durations) in the preview condition than in the control condition in which no preview of the array is provided. In a variation of the preview
presentation paradigm, van Zoest, Lleras, Kingstone, and Enns (2007) showed that when a search display was blanked for 900 ms and was re-presented during search, viewers were quick to respond to targets that were near the point of fixation before the interruption. In contrast, under similar search conditions when a gaze-contingent paradigm was used to present the target at the current point of fixation after the blank interval, viewers were not quick to respond. In effect, viewers were better at responding to the target if they autonomously found it, presumably with some preview from the previous fixation.

While all of the studies reviewed here are interesting, and suggestive of preview benefit, it is striking that there is no research directly using the types of gaze-contingent boundary paradigms that have been used in reading to study preview benefit in visual search. An approximation towards this is found in a study by Pomplun et al. (2001). They showed that in visual search with gaze-contingent windows, when particular features are visible outside the window, saccades are selectively made towards these particular features within the window. Perhaps the time is ripe to develop boundary paradigms (as used in reading research) to study preview benefit in visual search.

Eye movement control in visual search
Where and when to move the eyes. While there have been considerable efforts undertaken to determine the factors involved in deciding where and when to move the eyes in visual search (Greene, 2006; Greene & Rayner, 2001a, 2001b; Hooge & Erkelens, 1996, 1998; Hooge, Vlaskamp, & Over, 2007; Jacobs, 1986; Pomplun, Reingold, & Shen, 2003; Vaughan, 1982), a clear answer to the issue has not emerged. Some have concluded that fixation durations in search are the result of a combination of preprogrammed saccades and fixations that are influenced by the fixated information (Hooge et al., 2007; Vaughan, 1982). Others have suggested that the completion of foveal analysis is not necessarily the trigger for an eye movement (Hooge & Erkelens, 1996, 1998) while others have suggested that it is (Greene & Rayner, 2001b). Still others (Trukenbrod & Engbert, 2007) have demonstrated that fixation position is an important predictor of the next saccade and influences both the fixation duration and selection of the next saccade target. Rayner (1995) suggested that the trigger to move the eyes in a search task is something like: Is the target present in the decision area of the perceptual span? If it is not, a new saccade is programmed to move the eyes to a location that has not been examined (see also Motter & Belky, 1998a, 1998b; Najemnik & Geisler, 2005, for similar arguments). As with reading and scene perception, attention would move to the region targeted for the next saccade.

The decisions about where and when to move the eyes is undoubtedly strongly influenced by the characteristics of the specific search task and the density of the visual array, as well as viewer strategies (van Zoest, Donk, & Theeuwes, 2004). It seems that parallels between visual search and scene perception are greater than with reading, in that visual saliency plays a greater role in directing fixations. Also, search for targets in visual search displays and scenes have different dimensions that are more variable than reading. For instance, with respect to search tasks, there are many different types of targets that people may be asked to search for. Searching for a certain product in a grocery store shelf or searching for a particular person in a large group picture or for a word in a dictionary may well yield very different strategies than skimming text for a word (and hence influence eye movements in different ways). Although the task is generally much better defined in visual search than in scene perception, it typically is not as well specified as in reading.

Models of eye movement control in visual search
Perhaps the most well-known model of eye movement control is that of Findlay and Walker (1999). This model focuses on saccade generation based on parallel processing and competitive inhibition and, like many of the models of scene perception, relies heavily on the notion of a saliency map. While the model is unquestionably very interesting and very much tied to
neurophysiological properties of the oculomotor system, it is not a fully implemented model.

One fully implemented model is the target acquisition model (TAM) of Zelinsky (2008). TAM accounts for eye movements in search contexts ranging from fully realistic scenes to objects arranged in circular arrays to search for Os embedded in Qs. It also accounts for manipulations such as set size, target eccentricity, and target–distractor similarity, and it handles a number of important findings on eye movements and visual search (X. Chen & Zelinsky, 2006; Dickinson & Zelinsky, 2005, 2007; Neider & Zelinsky, 2006a, 2006b; Zelinsky, 1996, 2001; Zelinsky & Loschky, 2005; Zelinsky & Sheinberg, 1997). Comparisons of scan paths of the model to human viewers reveal that the model nicely mimics viewers’ behaviour, and it is difficult when presented with scan paths of the model and viewers to determine which is which. As impressive as TAM is with respect to simulating eye movements and scan paths during search, it does not provide an account of the determinants of when to move the eyes, and, hence, it does not predict fixation durations in search. Hopefully, future instantiations of TAM will lead to a better understanding of the mechanisms involved in accounting for how long the eyes pause in search.

Eye movements and visual cognition

Although there are obviously many differences between reading, scene perception, and visual search, there are some important generalizations that can be made. First, how much information is processed on any fixation (the perceptual span or functional field of view) varies as a function of the task. The perceptual span is obviously smaller in reading than in scene perception and visual search. Thus, for example, fixations in scene perception tend to be longer, and saccades are longer because more information is being processed on a fixation. Second, the difficulty of the stimulus influences eye movements: In reading, when the text becomes more difficult, eye fixations get longer, and saccades get shorter; likewise in scene perception and visual search, when the array is more difficult (crowded, cluttered, dense), fixations get longer, and saccades get shorter. Third, the difficulty of the task (reading for comprehension vs. reading for gist, searching for a person in a scene vs. looking at the scene for a memory test, and so on) clearly influences eye movements. Finally, in all three tasks it seems that viewers integrate visual information somewhat poorly across saccades (Najemnik & Geisler, 2005; Rayner, 1998), and that what is most critical is that there is efficient processing of information on each fixation.

To this point, research on reading, scene perception, and visual search has been the entire focus. In the next section, two lines of recent research utilizing eye movements are discussed: “real-world” tasks and the visual-world paradigm. Space limitations preclude discussion of eye movements during music perception (Gilman & Underwood, 2003; Land & Furneaux, 1997; Rayner & Pollatsek, 1997; Truitt, Clifton,Pollatsek, & Rayner, 1997), face perception (Henderson, Williams, & Falk, 2005; Williams & Henderson, 2007), typing (Inhoff, 1991; Inhoff & Gordon, 1997), driving (Chapman & Underwood, 1998; Land & Tatler, 2001; Pollatsek, Narayanaan, Pradhan, & Fisher, 2006a; Recarte & Nunes, 2000, 2003), problem solving and concept learning (Knoblich, Ohlsson, & Raney, 2001; Rehder & Hoffman, 2005), sports (Huber & Krist, 2004; Land & McLeod, 2001), mental rotation (Nakatani & Pollatsek, 2004), chess (Reingold, Charness, Pomplun, & Stampe, 2001), and advertising (Pieters, Wedel, & Liechty, 2008; Rayner, Miller, & Rotello, 2008; Rayner, Rotello, Stewart, Keir, & Duffy, 2001b). These, and other areas not mentioned,
are now active areas of research that utilize eye movement data to elucidate underlying cognitive processes in the various tasks.

**Eye movements in “real-world” tasks and the visual-world paradigm**

Recently, there has been considerable interest in using eye movements in tasks that presumably share many components with scene perception and visual search. Specifically, there has been considerable interest in “real-world” or “natural” tasks (Hayhoe & Ballard, 2005; Hayhoe, Shrivastava, Mruczek, & Pelz, 2007; Land & Hayhoe, 2001) and in a paradigm called the visual-world paradigm (Altmann, 2004; Altmann & Kamide, 1999; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus & Spivey-Knowlton, 1996; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In the so-called real-world or natural tasks, people’s eye movements are monitored as they engage in tasks such as making a sandwich or a cup of tea. While these studies are very interesting and while they demonstrate the much greater flexibility in recording eye movements that the new generations of eye tracking devices afford, there is something puzzling about the term. Are reading, scene perception, and visual search not real-world or natural tasks? Perhaps they are called real-world or natural tasks simply for the lack of a better descriptor. Certainly, there are many experiments on eye movements and driving (which is obviously also a real-world task), but these studies are typically not designated as real-world or natural tasks. A more appropriate label for these “real-world” studies would probably be something like eye movements and “action tasks” where some overt action is required on the part of the participant. Nonetheless, these types of studies do reveal important information about the relationship between eye movements and action plans (and the execution of such plans).

In the visual-world paradigm, participants hear some type of auditory input with a visual array in front of them. Eye movements show a systematic relationship between what is being listened to and where the eyes tend to go (and how quickly they go) in the visual array. The visual-world paradigm as such combines aspects of scene perception, visual search, and language processing. While the work is quite interesting, it is somewhat curious that there have been no real critics of the paradigm to date. For example, what do participants in the visual-world paradigm think they are supposed to do? What else are they supposed to do besides look at the visual array? There are also issues related to the nature of data analysis in the paradigm (see Altmann & Kamide, 2004, for a good discussion). Virtually all researchers who have adopted the paradigm have been advocates of its virtues (and there are obviously many). However, it is generally the case that paradigms that prove to be the most useful over the long run are also those where the basic assumptions underlying the paradigm are challenged, and empirical evidence is brought to bear on the issues at hand. This has certainly been the case over the past 40 years with respect to eye movements during reading.

**CONCLUDING COMMENTS**

Hopefully it is apparent that a great deal of knowledge has been gleaned from studies using eye movements to examine reading, language processing, scene perception, visual search, and other cognitive-processing tasks. As noted earlier, the present review is not intended as a comprehensive or exhaustive review as were earlier articles.

---

18 The task originated with Cooper (1974) but has been effectively utilized by Tanenhaus, Altmann, and others to study a number of topics ranging from auditory word recognition to syntactic parsing. Some recent work by Altmann (2004), Richardson and Spivey (2000), and Ferreira, Apel, and Henderson (2008) is very interesting in that they show, perhaps surprisingly, that listeners fixate on now–empty regions that had previously been occupied by relevant objects. Ferreira et al. suggested that the “looking at nothing” finding perhaps provides some clues about how the visual system creates and stores internal memory representations and that looking at nothing aids retrieval of these representations.
(Rayner, 1978b, 1998). Rather, I have largely reviewed work with some ties to my laboratory or work that I otherwise find interesting and appealing.

My contention is that research on eye movements during reading has advanced more rapidly and systematically than research on scene perception and visual search. This is probably due to the fact that stimulus characteristics (in reading, there is a limited set of letters that make up words, whereas in scene perception the scene is not as constrained by stimulus properties) and the task (the task in reading and visual search is quite straightforward, but exactly what viewers do in scene perception is not as obvious) are more amenable to experimental manipulations in reading than in scenes (especially) and search. Research on reading has significantly benefited from the use of the gaze-contingent paradigm, and while researchers in the domains of scene perception and visual search have been utilizing such paradigms recently, it is the case that there are many issues in both domains where the paradigms could be effectively used. It would also be appropriate for research of the latter type to acknowledge the prior work done on reading (which, unfortunately, does not always happen).

Another area where research on reading has been advanced over that on scene perception and visual search is with respect to the development of computational models to account for eye movement data. Models of eye movement control in reading tend to do a good job of accounting both for where readers look and how long they look at words. Models of eye movement control in scene perception and visual search have largely focused on where viewers look to the exclusion of when they move their eyes. Hopefully, this will be remedied in the next few years.

All in all, it is clear that major advances have been made with respect to understanding eye movements in reading, scene perception, and visual search. Although it has become increasingly clear that eye movements provide a very good (and precise) index of mental processing in various tasks, it is the case that eye movement research perhaps does not have quite the status many concede to various brain imaging techniques (even though eye movement data typically have better temporal resolution). On the other hand, it is certainly the case that more and more researchers are turning to eye movement recording and data as a means to examine important issues about how the brain/mind handles information in various tasks. Many brain imaging techniques now enable researchers to also record eye movements (though rather crudely), and attempts to simultaneously record eye movements and event-related potentials in reading and other tasks look very promising (Baccino & Manunta, 2005; Dambacher & Kliegl, 2007; Sereno & Rayner, 2003). Thus, the future looks very bright with respect to the possibility of learning more about cognitive processing and how information is processed in the tasks described above via the use of eye movements.

Original manuscript received 7 May 2008
Accepted revision received 4 September 2008
First published online 14 May 2009

REFERENCES


Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don’t: Mediating the mapping between language and the visual world. In J. M. Henderson...


