Binocular coordination and return-sweep saccades among skilled adult readers

Adam J. Parker Department of Psychology, Faculty of Science & Technology, Bournemouth University, Talbot Campus, Poole, UK

Mirela Nikolova School of Psychology, University of Southampton, Highfield Campus, Southampton, UK

Timothy J. Slattery Department of Psychology, Faculty of Science & Technology, Bournemouth University, Talbot Campus, Poole, UK

Simon P. Liversedge School of Psychology, University of Central Lancashire, Preston, UK

Julie A. Kirkby Department of Psychology, Faculty of Science & Technology, Bournemouth University, Talbot Campus, Poole, UK

During reading, binocular coordination ensures that a unified perceptual representation of the text is maintained across eye movements. However, slight vergence errors exist. The magnitude of disparity at fixation onset is related to the length of the preceding saccade. Return-sweeps are saccadic eye movements that span a line of text and direct gaze from the end of one line to the start of the next. As these eye movements travel much farther than intraline saccades, increased binocular disparity following a return-sweep is likely. Indeed, increased disparity has been a proposed explanation for longer line-initial fixations. Thus, we sought to address the following questions: Is binocular disparity larger following a return-sweep saccade than it is following an intraline saccade, and is the duration of a line-initial fixation related to binocular disparity and coordination processes? To examine these questions, we recorded binocular eye movements as participants read multiline texts. We report that, following return-sweeps, the magnitude of disparity at fixation onset is increased. However, this increased magnitude of disparity is unrelated to the duration of line-initial fixations. We argue that longer line-initial fixations result from a lack of parafoveal preview for words at the start of the line.

Introduction

Reading requires the coordination of the eyes so that high-acuity foveal vision can operate to allow for encoding of the words on a page. Saccadic eye movements direct gaze from one word to another. Saccades are followed by momentary periods of stillness, called fixations, during which specific regions of the text are projected to a relatively constant location on the retina and information is encoded (Hessels, Niehorster, Nyström, Andersson, & Hooge, 2018). Saccades typically move the eyes' gaze seven to nine character spaces for English readers. Binocular coordination ensures that a stable, unified perceptual representation of the text is maintained across saccades and fixations in order for visual and linguistic processing to proceed without disruption. However, it would appear that during reading, the two eyes' visual axes are regularly unaligned by more than one character space (Blythe, Liversedge, & Findlay, 2010; Blythe et al., 2006; Jainta, Blythe, Nikolova, Jones, & Liversedge, 2015; Jainta, Hoormann, Kloke, & Jaschinski, 2010; Liversedge, Rayner, White, Findlay, & McSorley, 2006; Liversedge, White, Findlay, & Rayner, 2006, Nikolova, Jainta, Blythe, Jones, & Liversedge, 2015; Nuthmann & Kliegl, 2009; Vernet & Kapoula,

Citation: Parker, A. J., Nikolova, M., Slattery, T. J., Liversedge, S. P., & Kirkby, J. A. (2019). Binocular coordination and return-sweep saccades among skilled adult readers. *Journal of Vision*, *19*(6):10, 1–19, https://doi.org/10.1167/19.6.10.

https://doi.org/10.1167/19.6.10 Received November 20, 2018; published June 11, 2019 ISSN 1534-7362

2009; for reviews, see Kirkby, Webster, Blythe, & Liversedge, 2008; Kirkby, White, & Blythe, 2011). This results from transient divergence during saccades where the abducting eye (the temporally moving eye) typically makes a larger, faster movement than the adducting eye (the nasally moving eye; Collewijn, Erkelens, & Steinman, 1988). This divergence results in observed binocular disparity at fixation onset. The magnitude of binocular disparity has been reported to vary as a function of saccade amplitude during both reading (Kirkby, Blythe, Drieghe, & Liversedge, 2011) and nonreading tasks (Collewijn et al., 1988; Kirkby, Blythe, Benson, & Liversedge, 2010).

In the context of single-line reading studies, the magnitude of fixation disparity is typically small, likely because the range of saccade amplitudes entering analysis is largely based on forward moving saccades of seven to nine characters. However, when readers process multiline text and arrive at the end of a line, they are required to execute a specific saccade, known as a return-sweep, to direct their gaze toward the start of a new line. The distance traversed by return-sweeps is larger than what is typically observed for intraline reading saccades, and this distance is directly related to the horizontal extent of the text (i.e., line length; Hofmeister, Heller, & Radach, 1999). It is not yet clear whether these distinct qualities of return-sweeps influence binocular disparity and compensatory vergence movements at the point where readers begin to process a new line of text. Line-initial fixations (i.e., those just following a return-sweep) are longer in duration than typical, intraline reading fixations (Parker, Slattery, & Kirkby, 2019; Rayner, 1977).[1] This has led to the assertion that a longer period for vergence movements is required to resolve increased disparity following a return-sweep (Stern, 1978). This prolonged period of vergence may subsequently influence the rate at which information is encoded following a return-sweep. Hence, the focus of the current study was to examine the basic characteristics of binocular coordination across multiline texts and provide an empirical examination of Stern's predictions concerning binocular coordination processing and line-initial fixation durations.

Return-sweeps in reading

Return-sweep saccades move the eyes from one line to the next. The majority of return-sweeps are launched from line-final words (Parker et al., 2019). Line-final and line-initial fixations generally occur five to seven characters from the extremes of the line (Rayner, 1998). Previous examinations of return-sweeps have grouped fixations into four populations (Parker et al., 2019): intraline, line-final, accurate line-initial, and undersweep. Intraline fixations are nonadjacent to return-sweeps. Line-final fixations are those that precede a return-sweep. Line-initial fixations are those that follow return-sweeps and they can be further categorized into two specific groups: Accurate line-initial fixations are those that land close enough to the target of the return-sweep and are immediately followed by a rightward saccade so that a reader can begin their left-to-right pass. Undersweep fixations are those that land on a new line and are followed by an immediate leftward corrective saccade to position the fovea closer to the start of the line prior to the left-to-right pass. While accurate line-initial fixations tend to be longer than intraline fixations (Heller, 1982; Hawley, Stern, & Chen, 1974; Rayner, 1977), line-final (Abrams & Zuber, 1972; Rayner, 1977) and undersweep fixations (Heller, 1982) tend to be shorter. Undersweep fixations are considered uninvolved in ongoing linguistic processing (Hawley et al., 1974; Shebilske, 1975).

As readers progress toward the end of a line of text, fixation times decrease (Kuperman, Dambacher, Nuthmann, & Kliegl, 2010; Pynte & Kennedy, 2006). Consistent with this trend, the last fixation on a line has been reported to be shorter than typical intraline reading fixations (Rayner, 1977). Rayner argued that reduced parafoveal processing toward the end of the line results in these shorter line-final fixations. However, Kuperman et al. and Mitchell, Shen, Green, and Hodgson (2008) have suggested that this reduction in fixation duration results from oculomotor programming in response to line boundaries. Consistent with this assumption, Hofmeister (1997) found that line-final fixations were unaffected by text degradation of 50%, while all other reading fixations were increased by 20 ms.

Eye movements are the result of muscular contractions and are subject to both saccadic range and random error (McConkie, Kerr, Reddix, & Zola, 1988). As return-sweeps move the eyes farther than typical reading saccades, they are more likely to be influenced by these error components. However, return-sweeps very infrequently overshoot their target (Hofmeister et al., 1999). Instead, they have a systematic tendency to undershoot the start of the line (Heller, 1982; Hofmeister et al., 1999; Parker, Kirkby, & Slattery, 2017; Parker et al., 2019).

Early investigations of return-sweep accuracy considered the influence of typographical factors (Tinker, 1963). Longer lines have typically been associated with increased return-sweep undershoot error (RUE). For instance, D. G. Paterson and Tinker (1940) observed that reading times were longer for passages of 30 words formatted as longer lines. They attributed this increase in reading time to increased RUE. It is important to note that the passages viewed by participants were not counterbalanced across conditions. This makes it

difficult to determine whether this effect varied directly as a function of line length or whether it was modulated by the difficulty of the text in each condition. Similarly, differential characteristics of words presented at return-sweep launch sites and landing positions could have influenced targeting and accuracy between conditions. Despite these limitations, subsequent studies have reported similar findings. In general, the frequency of RUEs increases with the length of the intended return-sweep, with shorter lines yielding fewer RUEs (Beymer, Russell, & Orton, 2005; Schneps et al., 2013), while the landing position of the return-sweep is shifted to the right (Hofmeister et al., 1999).

More recent work, however, has elucidated these line-length findings. Parker et al. (2019) provided evidence to suggest that it is the distance from the left margin at which return-sweeps land that influences the likelihood of initiating a corrective saccade. This evidence is consistent with Becker's (1976) explanation of corrective saccades. That is, corrective saccades are not a product of saccade length but rather a response to a deviation of return-sweep landing position from the saccade target, which, once it exceeds a certain threshold, triggers a corrective saccade.

Accurate line-initial fixations tend to be longer than intraline fixations. While several explanations for this increased duration exist (Kuperman et al., 2010; Pynte & Kennedy, 2006; Rayner, 1977), one of particular relevance to our research was proposed by Stern (1978), who argued that longer line-initial fixations result from vergence movements and a period of orientation to compensate for increased divergence occurring during the return-sweep. However, in the absence of evidence to support Stern's assertion, an alternative possibility—a lack of parafoveal preview—may also explain this effect. Numerous studies have shown that when readers are able to preprocess a word in parafoveal vision, subsequent fixation times on that word are shorter (for a review, see Schotter, Angele, & Rayner, 2012). Words receiving line-initial fixations are unavailable for parafoveal preprocessing, as they lie outside of the perceptual span, and can only be processed foveally following a return-sweep. Consistent with this, Parker et al. (2017) reported longer gaze durations on line-initial words which could not be preprocessed during a line-final fixation than on words presented midline which were available for parafoveal preprocessing.

Binocular coordination during reading

Humans typically make use of both of their eyes when they read. Because of disconjugacy between the two eyes originating during a saccade, binocular coordination is required during a fixation to ensure a unified perceptual representation of written words. This unified percept results from motor and sensory fusion (Pratt-Johnson & Tillson, 2001). Motor fusion comprises the physiological mechanism of vergence. Due to transient divergence during a saccade, fixation disparity is typically observed at fixation onset (Blythe et al., 2010; Blythe et al., 2006; Jainta et al., 2010; Liversedge, Rayner, et al., 2006; Nuthmann & Kliegl, 2009; Vernet & Kapoula, 2009). Liversedge, White, et al. (2006) provided a comprehensive description of binocular coordination during reading. They reported that the two visual axes were, on average, 1.9 character spaces disparate when the eyes were unaligned, which accounted for 47% of all fixations during single-line reading. The unaligned fixations were further categorized as crossed (8%) and uncrossed (39%). Crossed fixations were those where the left eye fixated farther to the right than the right eye, and uncrossed fixations were those where the left eye fixated farther to the left than the right eye. Note that the distribution of crossed and uncrossed fixations varies between studies (for a discussion, see Kirkby, Blythe, Drieghe, Benson, & Liversedge, 2013; for a theoretical explanation of crossed and uncrossed fixations, see Shillcock, Roberts, Kreiner, & Obregón, 2010). Fine-grained oculomotor (vergence) movements are then made during fixations to resolve these disparities and maximize retinal correspondence between the two eyes (e.g., Liversedge, White, et al., 2006). Sensory fusion serves to combine the two retinal representations into a unified perceptual representation for higher level processing (Howard & Rogers, 1995, 2012; Worth, 1921). Sensory fusion occurs when disparities between the two retinal representations fall within Panum's fusional area (Blythe et al., 2010).

Studies of binocular coordination during reading have mainly examined how the visual or lexical characteristics of the text influence fixation disparity (for reviews, see Kirkby et al., 2008; Kirkby, Blythe et al., 2011). Hendriks (1996) reported that vergence velocities were higher when reading prose compared to a list of unrelated words, and argued that during prose processing, readers tolerate increased disparity at fixation onset as they use contextual information to aid lexical identification, whereas unrelated words can be processed only visually. Most importantly, Hendriks considered binocular coordination to be influenced by the properties of the text.

Heller and Radach (1999) had six participants read 200 lines of text presented in normal case or MiXeD cAsE. Mixed-case fonts result in longer gaze durations (Reingold, Yang, & Rayner, 2010) and are considered to be more effortful to read than single-case fonts (Rayner, McConkie, & Zola, 1980), as they may disrupt word-identification processes (Coltheart & Freeman, 1974). Heller and Radach observed a reduced magnitude of fixation disparity for passages in mixed

case relative to normal case. Similar to Hendriks (1996), they concluded that readers tolerate an increased magnitude of fixation disparity when reading under easy reading conditions (i.e., normally presented text). In contrast, Juhasz, Liversedge, White, and Rayner (2006) failed to replicate the effect of mixed-case fonts on fixation disparity. Additionally, they found no effect of lexical frequency (a measure of processing difficulty) on the magnitude of fixation disparity. That is, the magnitude of fixation disparity was similar for high- and low-frequency words despite fixation-duration results demonstrating clear disruptive effects for both mixed-case fonts and words of reduced frequency. Blythe et al. (2006) and Nikolova et al. (2015) have subsequently reported no modulatory influence of frequency on the magnitude of fixation disparity. Another word-level variable that appears to have no influence on the magnitude of fixation disparity is word length. When participants read sentences in which the length of a target word was varied to be four or 10 characters long, K. B. Paterson, McGowan, and Jordan (2013) observed similar magnitudes of disparity across the different-length target words. Similar findings have been reported in the nonreading literature, where the horizontal extent of the target is unrelated to the magnitude of fixation disparity (Kirkby et al., 2010). These mixed results warrant no firm conclusion concerning whether fixation disparity is influenced by the processing difficulty of the text.

The magnitude of fixation disparity has been shown to vary as a function of fixation position along the line of text. Heller and Radach (1999) reported that the magnitude of fixation disparity accumulated over the first line of text, with an average magnitude of approximately two character spaces. This trend then slowed and reversed for subsequent lines. Note, however, that Heller and Radach conducted no statistical analysis on this portion of the data. Jainta et al. (2010) reported that when participants read single sentences in German, selected from the Potsdam-Sentence-Corpus (Kliegl, Nuthmann, & Engbert, 2006), there was a significant linear relationship between fixation position on the screen and fixation disparity. That is, the magnitude of disparity at fixation onset increased as readers moved from left to right. Kirkby, Blythe et al. (2011) also reported an effect of fixation position on the magnitude of fixation disparity in children, but not in adults. If the magnitude of fixation disparity does increase across the line, it is possible that the magnitude of disparity at fixation onset may be greatest for line-final fixations.

The magnitude of fixation disparity is influenced by the length of the preceding saccade. Hendriks (1996) noted faster vergence velocities following long saccades than after short saccades (see also Collewijn et al., 1988; Zee, Fitzgibbon, & Optican, 1992). Similarly, Kirkby et al. (2010) reported that when participants were required to saccade from a central dot to a dot positioned at either 5.5° or 2.1° away from the center, increased disparity was noted for dots positioned 5.5° away. In a reading task, Kirkby, Blythe et al. (2011) found that disparity increased with incoming saccade length for both adults and children. One exception to this is reported by Schotter, Blythe et al. (2012), who presented participants with stimuli as normal or looming text. Looming text was additionally presented in two conditions where stimulus size and stimulus disparity varied, to manipulate the congruency of monocular and binocular depth cues. In their initial analyses, the researchers reported no effect of saccade length on the magnitude of disparity at fixation onset or offset. They argued that this was likely the result of text properties being correlated with saccade length and accounting for more variance in the data. Subsequent analysis revealed that saccade length was related to the magnitude of fixation disparity at onset when manipulations of increasing retinal size and increasing retinal disparity were excluded.

Together these results have implications for return-sweep saccades. Given the relationship between saccade length and the magnitude of disparity at fixation onset, it is expected that an increased magnitude of disparity would be observed for line-initial fixations. In turn, an increased period for vergence movements may be required to maximize the correspondence between the two visual axes prior to sensory fusion (Stern, 1978). Assuming that reduced disparity is facilitative to lexical processing, such speculation would predict a clear relationship between the magnitude of disparity at fixation onset and fixation duration. While Kliegl et al. (2006) reported no influence of the magnitude of fixation disparity on fixation durations, this may be due to the relatively small disparities associated with normal intraline fixations. If disparity is significantly greater at the onset of fixations which follow a return-sweep, then these would be excellent test cases for Stern's assertion of a relationship between fixation disparity and fixation duration during natural reading.

The present study

The present study sought to answer the questions: Is binocular disparity larger following a return-sweep saccade than it is following an intraline saccade, and is the duration of a line-initial fixation related to binocular coordination processes? Thus, our aim was twofold: first, to empirically evaluate the claim that binocular disparity will be greater following a return-sweep saccade than it is following intraline saccades; and second, to empirically evaluate Stern's (1978)

assertion that longer accurate line-initial fixations are the result of binocular coordination processes which compensate for increased divergence during a return-sweep. We addressed these aims by recording binocular eye movements of participants reading multiline texts. Given the relationship between saccade length and the magnitude of disparity at fixation onset, we included a line-length manipulation to vary the length of the return-sweep. To foreshadow, not only did this enable us to replicate prior line-length and return-sweep findings, it provided an opportunity to contrast the durations of two groups of line-initial fixations with different magnitudes of disparity.

To provide both a measure of fixation disparity and an index of the vergence movements that occurred during fixations, we examined the magnitude of fixation disparity at both fixation onset and offset. We also considered the proportion of different alignments across multiline texts. For intraline reading fixations, we expected disparities similar to those reported in prior work—that is, of approximately one character space—on a significant proportion of fixations (e.g., Liversedge, White, et al., 2006). Previous research has reported that an accumulation of fixation disparity occurs across a line of text in skilled adult readers (Heller & Radach, 1999; Jainta et al., 2010), yet other research has not (Kirkby, Blythe et al., 2011; note that Schotter, Blythe et al., 2012, also reported null effects when depth cues were not varied). Based on these mixed results, we did not make clear predictions for the magnitude of disparity for line-final fixations. However, our predictions for post-return-sweep (i.e., line-initial) fixations were clear. Return-sweeps typically move much farther than intraline saccades, and fixation disparity is related to this distance for both reading (Kirkby, Blythe et al., 2011) and nonreading (Collewijn et al., 1988; Kirkby et al., 2010) tasks. Therefore, we predicted that the magnitude of binocular disparity at fixation onset would be greater for line-initial fixations than for intraline reading fixations. Given the distinction between accurate line-initial and undersweep fixations, we predicted that greater disparities would be observed following an accurate return-sweep than an undersweep, as accurate return-sweep saccades tend to travel farther than undersweep saccades.

To evaluate Stern's (1978) assertion that longer line-initial fixations result from the need to converge following a return-sweep, we examined two metrics: the durations of fixation populations and the relationship between line-initial fixation duration and the magnitude of disparity at fixation onset. For the durations of each fixation population, following Parker et al. (2019) we expected to observe line-final fixations that were shorter than intraline reading fixations. We expected that undersweep fixations would also be shorter than intraline fixations, while accurate line-initial fixations would be longer. If a greater magnitude of disparity were observed for accurate line-initial fixations in the long-line condition, following Stern, it would be predicted that the durations of these fixations would be greater than in the short-line condition. A similar account would hold for undersweep fixations in the long-line condition. However, no effect of line length would be expected for intraline and line-final fixations. Our second analysis which aimed to address Stern's assertion would predict that an increased magnitude of disparity at fixation onset would result in longer line-initial fixations. However, given previous null effects of the magnitude of disparity on fixation duration (e.g., Kliegl et al., 2006), it is possible that a series of null findings would strengthen alternate claims that longer line-initial fixations result from a lack of preview (Parker et al., 2017) or from planning a series of saccades across the subsequent line (Kuperman et al., 2010).

Method

Participants

Fourteen native English speakers from the University of Southampton (10 women, four men; mean age: 21.4 years; range: 18–34 years) participated in exchange for course credit. All participants had normal or corrected-to-normal vision (with soft contact lenses) and no history of reading impairment. Tests of visual acuity and a Titmus Stereotest indicated that participants had 20/20 or better acuity in each eye at 4 m and functional stereopsis (minimal stereoacuity of 40 arcsec). Each participant gave informed consent before the experiment. The experimental procedure was approved by Bournemouth University's Research Ethics Code of Practice and the University of Southampton Ethics and Research Governance Office and followed the conventions of the Declaration of Helsinki.

Apparatus

The apparatus, viewing conditions, and calibration procedures were identical to those reported in prior work (e.g., Jainta, Blythe, & Liversedge, 2014; Nikolova, Jainta, Blythe, & Liversedge, 2017, 2018). Binocular eye movements were recorded using two Fourward Technologies dual-Purkinje-image eye trackers. The analogue signals were sampled at a rate of 1000 Hz (i.e., once every millisecond) using custom-designed software and an analogue-to-digital board. Red text was displayed on a black background on a

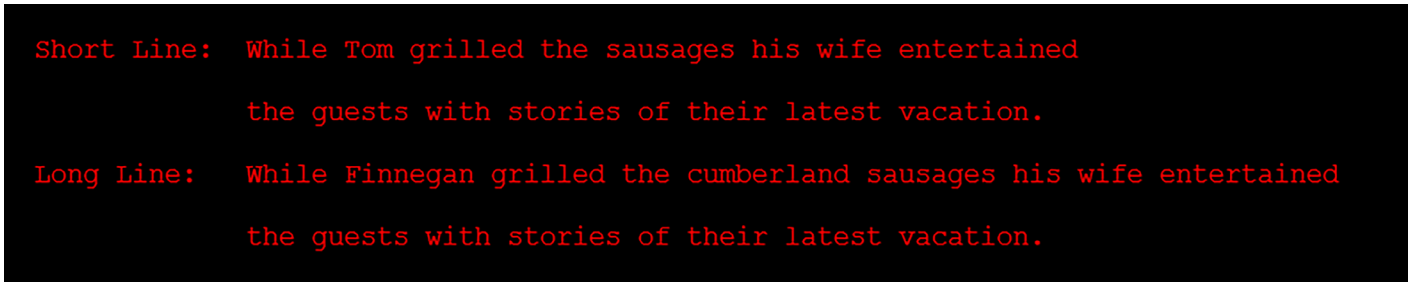

Figure 1. Examples of stimuli in the two experimental conditions: short and long line lengths.

Philips 21B582BH 21-in. monitor with a pixel resolution of 1,024 × 768.[2] Monocular calibration was achieved through the use of Cambridge Research Systems FEI shutter goggles. Viewing distance was 100 cm, so that one character of monospaced font (Courier New) subtended 0.25° of visual angle. Participants minimized head movements by leaning against two cushioned forehead rests and biting on an individually prepared bite bar.

Materials

Experimental stimuli consisted of 60 neutral passages of text. Each passage contained one to two sentences displayed across two lines, which were formatted to one of two line lengths: 56 characters (14° of visual angle) or 73 characters (18° of visual angle; see Figure 1). Words in the text varied in length from one to 13 letters (mean = 4.67) and had an average Zipf frequency (van Heuven, Mandera, Keuleers, & Brysbaert, 2014), based on the SUBTLEX database (Brysbaert & New, 2009), of 5.70 (range: 1.47 to 7.67). While the first line of text varied, line-final words and the second line of text were identical across conditions. This ensured that the lexical characteristics of words occurring at return-sweep launch sites and landing positions were identical across conditions, to avoid content differences in these regions. Passages were counterbalanced so that each participant read an equal number in each condition.

Procedure

After providing informed consent, participants completed the visual acuity and Titmus Stereotest. To calibrate the dual-Purkinje-image eye trackers, a monocular procedure was used (the left eye was occluded during calibration of the right eye, and vice versa). Participants fixated each of nine points in a 3 × 3 grid from the top left to the bottom right. Calibration was repeated if the distance between the recorded eye position and the actual validation point on the screen exceeded 0.25° of visual angle. The mean error observed was 0.04° on the horizontal plane and 0.01° on the vertical plane. Following successful calibration, participants completed two practice trials to familiarize themselves with the experimental setup. The practice trials were followed by further calibration and validation, after which the experiment began. After every five trials or track loss, the trackers were recalibrated. On average, 19.6% of trials were followed by a recalibration.

At the start of each trial, participants fixated a circle on the left-hand side of the screen. After 1,000 ms, the stimulus replaced the circle. After reading the passage, participants pressed a button on a button box to end the trial. After a third of the passages, participants used the button box to respond to a yes/no comprehension question. After the first 30 trials, participants were given a break. Additional breaks were given whenever required. The whole experiment lasted approximately 45–60 min.

Data analysis

Data analysis was conducted using custom-designed software. Following previous work (e.g., Kirkby et al., 2013), fixations and saccades were manually identified in order to avoid contamination by dynamic overshoots (Deubel & Bridgeman, 1995) or artifacts due to blinks. We excluded trials with track loss or blinks that were adjacent to the return-sweep (10.3%), fixations shorter than 80 ms or longer than 800 ms (6.3%), and the first and last fixation on each trial (10.0% of fixations). Disparity at fixation onset and offset was calculated by deducting the horizontal start of fixation position (degrees of visual angle from the center of the screen) for the right eye from that of the left eye. A value of 0° represents perfect alignment of the two eyes; positive values represent crossed fixations; negative values represent uncrossed fixations. We analyzed fixations only when disparity fell within 2.5 standard deviations of the mean for each participant. This enabled the exclusion of atypically large fixation disparities (>2°), which

may have resulted from tracker error. Trimming procedures based on fixation disparities led to 1.3% of fixations being removed. Analysis was conducted on the remaining 10,986 fixations.[3]

For statistical analysis, we used linear mixed-effects models (LMMs) constructed using the lme4 package (Version 1.1-18; Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.5.1.; R Core Team, 2018). For each predictor, we report regression coefficients (b), standard errors (SE), and t values. We used the two-tailed criterion $|t| > 1.96$ for significance, corresponding to $\alpha = 0.05$. The z values for generalized LMMs (GLMMs) are interpreted similarly. To conserve power lost to unnecessary complexity, we used a parsimonious approach to model the random-effects structure (Bates, Kliegl, Vasishth, & Baayen, 2019). By this approach, a backward selection procedure was used. That is, each random slope was removed one at a time, with the highest order interactions explored first. At each stage, the model was compared to each previous model. Where the removal of a random slope did not affect the model (i.e., $p > 0.2$; Barr, 2013), that removal was deemed justifiable. All numerical variables were centered prior to analysis.

Results

Across participants, 93% of comprehension questions were answered correctly. In the following sections, we report analyses aimed at addressing our specific research questions. The first group of analyses examined the basic characteristics of binocular coordination over multiline texts. The second examined Stern's (1978) assertion that longer line-initial fixations can be attributed to binocular coordination processes. In Supplementary File S1, we replicated previous examinations of return-sweep and corrective-saccade parameters and lexical and nonlexical influences on the magnitude of fixation disparity. That is, longer lines were associated with return-sweep landing positions which were farther from the left margin and more frequent undersweep fixations. Additionally, the length of the preceding saccade (but not word length or lexical frequency) modulated the magnitude of disparity at fixation onset. These replications within the data lend credibility to the novel aspects. To contrast return-sweep fixations with intraline fixations, we coded intraline fixations as −1. Line-final, accurate line-initial, and undersweep fixations were coded as 0.5, −0.5, and 1, respectively. Intraline reading fixations represent the intercept to which return-sweep fixations were compared.

What are the basic characteristics of binocular coordination during the reading of multiline text?

Fixation disparity prior to and following a return-sweep

Our primary goal was to investigate the basic characteristics of binocular coordination during the reading of multiline texts. This enabled us to address the question: is binocular disparity larger following a return-sweep saccade than it is following an intraline saccade? Disparity at fixation onset and offset were calculated to provide a static measure of fixation disparity and an index of the vergence movements that occurred during the fixation. The magnitude of disparity was measured as the absolute difference between the fixation positions of the two eyes, as this measure enabled calculation of disparity without consideration of direction (i.e., crossed or uncrossed). The mean absolute difference in saccade amplitude for the two eyes was also computed. The distribution of each variable is shown in Figure 2.

The model fit to absolute disparity at fixation onset—lmer(dv~ fixation population * line length + (1 + line length | participants) + (1 + line length | items))—revealed that compared to the intercept (intraline reading fixations), the magnitude of disparity was significantly increased at fixation onset for accurate line-initial fixations and undersweep fixations (see Table 1). The magnitude of disparity at fixation onset did not differ significantly between intraline reading fixations and line-final fixations. Additionally, the accuracy of return-sweeps and line length interactively influenced the magnitude of absolute fixation disparity, whereby the magnitude of disparity was greater following an accurate return-sweep in the long-line condition compared with the short-line condition. An identical model fit to absolute disparity at fixation offset indicated that disparity did not differ as a function of line length or fixation population at offset. Thus, vergence movements resolve increased disparities across different fixation populations during the fixation.

The LMM fit to the absolute difference in preceding-saccade amplitude for each population of reading fixation—lmer(dv~ fixation population * line length + (1+ fixation population * line length | participants) + (1 | items))—revealed that in comparison to intraline reading fixations, the difference in saccade amplitude was increased for line-final, accurate line-initial, and undersweep fixations. The magnitude of difference did not differ between line-length conditions for intraline or line-final fixations. The interaction between line length and return-sweep accuracy indicated that the difference in preceding-saccade amplitude between the two eyes was greater in the long-line condition for accurate line-initial and undersweep fixations. Given

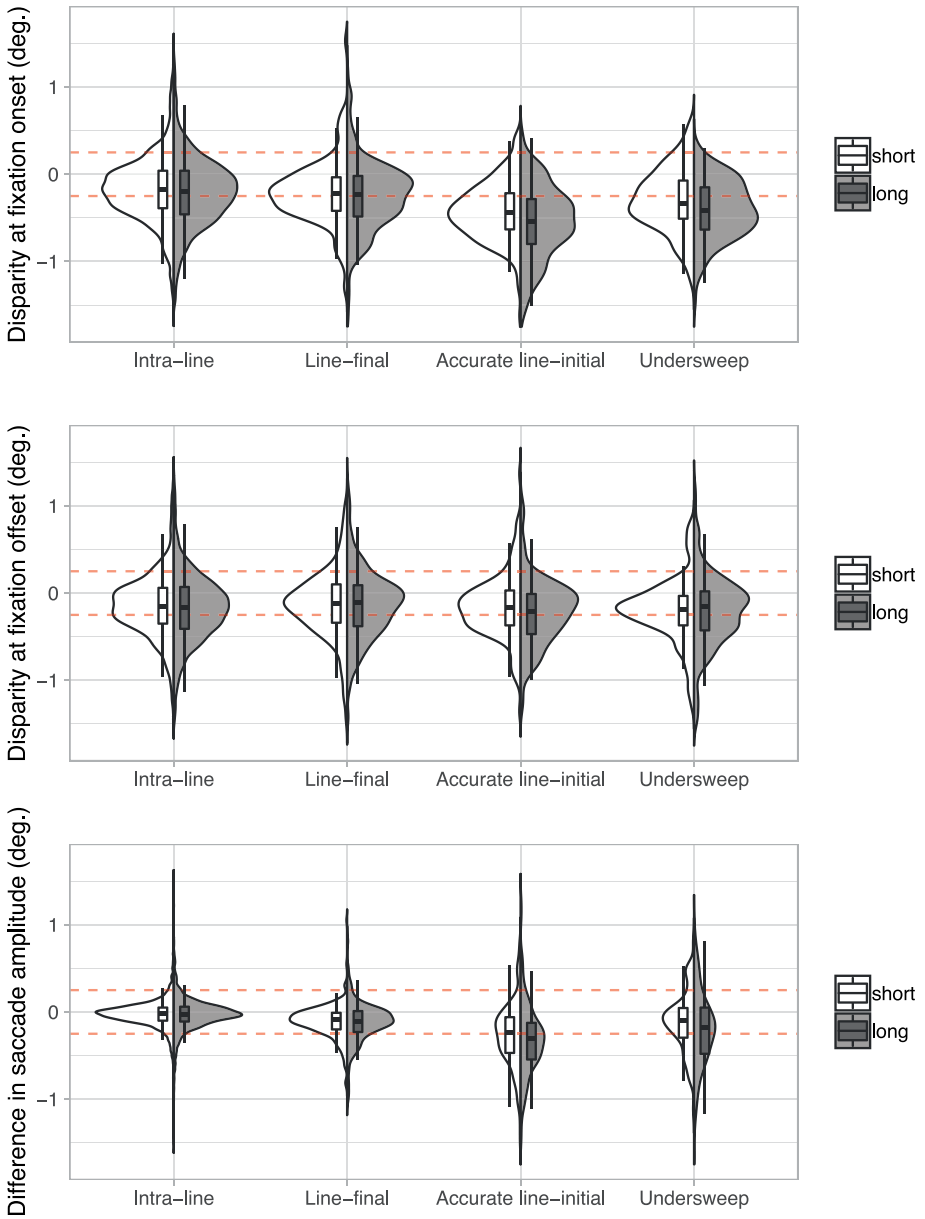

Figure 2. Box plots for fixation disparities observed at fixation onset and offset for each fixation population, with violin plots behind. Disparity is reported in degrees of visual angle; positive values indicate crossed disparities, and negative values indicate uncrossed disparities. The two dashed lines at $\pm 0.25^{\circ}$ represent a maximum disparity of a character in either direction, and fixations were considered aligned when the observed disparity fell within that range.

that these line-initial fixations will have, on average, traveled farther in the long-line condition than the short-line condition, it would appear that the increased length of the return-sweep will have led to more divergence during a saccade. This then results in a larger difference in saccade amplitude between the two eyes.

The proportions of alignment prior to and following a return-sweep

Recall that aligned fixations are those where both fixation points fall within one character of each other within a word (cf. Liversedge, White, et al., 2006). Unaligned fixations are those where disparity exceeds one character space. Crossed fixations are unaligned fixations where the left eye fixates farther to the right than the right eye. Uncrossed fixations are unaligned fixations where the left eye fixates farther to the left than the right eye. From Figure 3, it is evident that the majority of fixations were aligned at fixation onset and offset. For unaligned fixations, the majority were uncrossed, and only a small proportion were crossed. This pattern of alignment is comparable with previous research using similar stimulus conditions (Blythe et al., 2010; Blythe et al., 2006; Jainta et al., 2015; Juhasz et

	Onset disparity (°)			Offset disparity (°)			Difference in saccade amplitude (°)		
Predictor	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	0.341	0.028	**12.37**	0.321	0.025	**12.77**	0.121	0.011	**11.43**
Line-final	0.020	0.011	1.72	0.001	0.010	0.12	0.043	0.017	**2.61**
Accurate line-initial	0.174	0.013	**12.96**	0.003	0.015	0.17	0.217	0.034	**6.44**
Undersweep	0.072	0.016	**4.63**	0.024	0.016	1.48	0.027	0.012	**2.15**
Line length	−0.007	0.012	−0.54	−0.008	0.013	−0.62	0.319	0.228	1.40
Line-final × Line length	−0.001	0.011	−0.07	0.009	0.010	0.88	0.013	0.017	0.77
Accurate line-initial × Line length	−0.046	0.013	**−3.43**	0.008	0.015	0.54	0.085	0.026	**3.22**
Undersweep × Line length	−0.007	0.016	−0.42	−0.003	0.016	−0.21	0.162	0.030	**5.40**

	Onset disparity (°)		Offset disparity (°)		Difference in saccade amplitude (°)	
Random effects	Variance	*SD*	Variance	*SD*	Variance	*SD*
Item: Intercept	0.014	0.119	0.016	0.125	<0.001	0.014
Item: Line length	0.023	0.151	0.024	0.156		
Participant: Intercept	0.008	0.092	0.007	0.081	0.001	0.037
Participant: Line-final					0.002	0.041
Participant: Accurate line-initial					0.011	0.106
Participant: Undersweep fixation					0.717	0.847
Participant: Line length	0.002	0.047	0.002	0.045	0.002	0.043

Table 1. Results of linear mixed-effects model for absolute fixation disparity at onset and offset and difference in saccade amplitude for both eyes. *Notes*: Significant *t* values ($|t| \geq 1.96$) are in bold. *SE* = standard error; *SD* = standard deviation.

al., 2006; Kirkby et al., 2010; Kirkby et al., 2013; Kirkby, Blythe et al., 2011; Liversedge, Rayner, et al., 2006; Liversedge, White, et al., 2006; Nikolova et al., 2015; Nikolova et al., 2017, 2018; Schotter, Blythe et al., 2012). The proportion of fixation alignment for each fixation population can be seen in Figure 2 as a function of line length.

To examine the extent to which the proportion of alignment differed between line-length condition and fixation type, we fit a series of GLMMs to fixation-alignment data at both fixation onset and offset. The coefficients for each analysis are shown in Table 2. The first model was fit to a categorical variable which coded whether the fixation was aligned or not aligned at fixation onset: glmer(dv∼ fixation population * line length + (1+ fixation population * line length | participants) + (1+ line length | items)). For intraline reading fixations (intercept), there was no difference in the proportion of aligned and unaligned fixations. A similar pattern was observed for line-final fixations. Following a return-sweep, there were significantly more unaligned fixations for both accurate return-sweeps and undersweep fixations in comparison to intraline reading fixations. There was no effect of line length on the proportion of alignment for any fixation population. These results are in line with our predictions.

Next we fit a GLMM to unaligned fixations to assess whether there were more crossed or uncrossed fixations for each population: glmer(dv∼ fixation population * line length + (1+ line length | participants) + (1+ line length | items)). For all fixation populations, there were more uncrossed than crossed fixations. There was a significant interaction between line length and fixation population, such that there was a higher proportion of uncrossed fixations in the long-line relative to the short-line condition for undersweep fixations. There was no reliable interactive effect of line length with any of the remaining fixation populations in relation to the proportion of crossed and uncrossed fixations.

To assess the proportion of fixations that were aligned and unaligned at fixation offset, we fit a model to a categorical variable which coded whether the fixation was aligned or not at fixation onset: glmer(dv∼ fixation population * line length + (1+ line length | participants) + (1+ line length | items)). Analysis revealed that the proportion of aligned and unaligned fixations did not differ between fixation population or line length condition. Finally, we fit a model to unaligned fixation data to evaluate the effect of fixation population and line length on the proportion of crossed and uncrossed fixations at fixation offset: lmer(dv∼ fixation population * line length + (1+ fixation population * line length | participants) + (1 | items)). As predicted, there were more uncrossed than crossed fixations. However, the proportion of crossed and uncrossed fixations did not differ between fixation population or line-length condition.

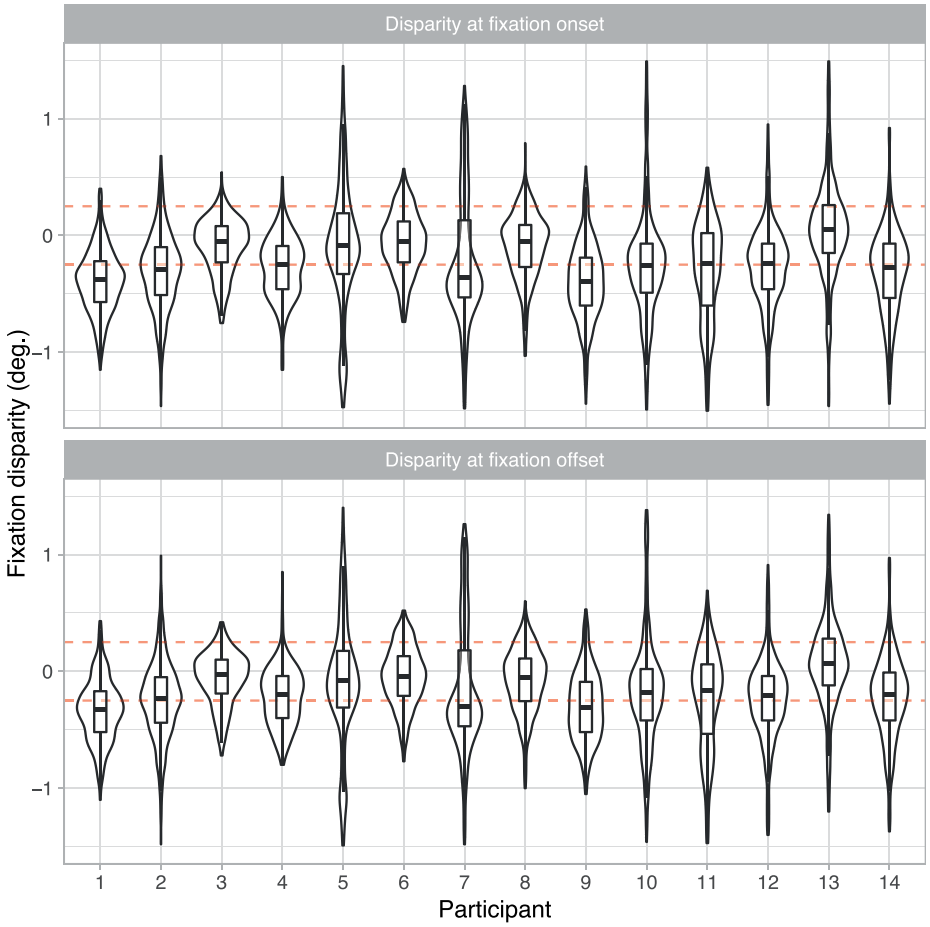

Figure 3. Box plots for fixation disparities observed at fixation onset and offset for each participant, with violin plots behind. Disparity is reported in degrees of visual angle; positive values indicate crossed disparities, and negative values indicate uncrossed disparities. The two lines at ±0.25° represent a maximum disparity of a character in either direction, and fixations were considered aligned when the observed disparity fell within that range.

Does fixation disparity modulate line-initial fixation duration?

Here we address the question: is the duration of a line-initial fixation related to binocular disparity and coordination processes? Recall Stern's (1978) assertion that accurate line-initial fixations are longer as a result of divergence during a return-sweep, which results in the need to reconverge at the beginning of the line. Consistent with part of this assertion, previous analyses have indicated an increased magnitude of disparity at fixation onset following return-sweeps. At fixation offset, the magnitude of disparity did not differ between intraline reading fixations and those following a return-sweep. Therefore, it is possible that an increased magnitude of disparity at fixation onset may be responsible for the longer duration of line-initial fixations. To examine this possibility, we report two LMM analyses. The first examined whether line-initial fixations differed between line-length conditions. As the magnitude of fixation disparity was larger at fixation onset for accurate line-initial fixations in the long-line condition than in the short-line condition, Stern's assertion would predict longer line-initial fixations in the long-line condition. The second examined whether line-initial fixation durations were modulated by the magnitude of disparity at fixation onset. In order to infer the extent to which our data reflect null effects for these analyses, as opposed to type II error, we computed Bayes factors for these analyses.

Fixation durations across multiline texts

The distributions of fixation durations for each fixation population in each line-length condition are shown in Figure 4. Following Brysbaert and Stevens (2018), fixation times were inverse transformed for normality prior to analysis. The model included fixed effects for fixation population, line-length condition, and their interaction: lmer(dv~ fixation population + line length + (1+ fixation population * line length | participants) + (1 | items)).

Fixed effects	Fixation onset						Fixation offset					
	Proportion of alignment			Proportion of uncrossed fixations			Proportion of alignment			Proportion of uncrossed fixations		
Predictor	*b*	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*
Intercept	0.119	0.216	0.55	2.490	0.494	**5.04**	−0.008	0.190	−0.04	−0.008	0.401	**4.95**
Line-final	0.151	0.175	0.86	0.586	0.222	**2.64**	−0.024	0.089	−0.27	−0.024	0.210	−1.71
Accurate line-initial	1.156	0.294	**3.93**	3.245	0.491	**6.61**	−0.080	0.132	−0.61	−0.080	0.343	0.99
Undersweep	0.7071	0.219	**3.20**	2.219	0.451	**4.92**	0.070	0.147	0.48	0.070	0.410	0.93
Line length	0.040	0.096	0.41	0.084	0.296	0.28	0.080	0.097	0.82	0.0799	0.173	0.33
Line-final × Line length	−0.020	0.101	−0.20	−0.227	0.222	−1.02	−0.005	0.089	−0.06	−0.005	0.172	−0.33
Accurate line-initial × Line length	0.133	0.144	0.93	0.460	0.491	0.94	−0.179	0.132	−1.36	−0.179	0.290	0.24
Undersweep × Line length	0.077	0.147	0.53	0.813	0.451	1.80	0.034	0.147	0.23	0.034	0.327	−0.73

Random effects	Fixation onset				Fixation offset			
	Proportion of alignment		Proportion of uncrossed fixations		Proportion of alignment		Proportion of uncrossed fixations	
	Variance	*SD*	Variance	*SD*	Variance	*SD*	Variance	*SD*
Item: Intercept	0.304	0.551	4.590	2.142	0.323	0.568	0.898	0.948
Item: Line length	0.670	0.819	10.481	3.237	0.814	0.902		
Participant: Intercept	0.678	0.823	2.636	1.624	0.485	0.700	1.740	1.319
Participant: Line-final	0.280	0.529					0.110	0.332
Participant: Accurate line-initial	0.918	0.958					0.140	0.374
Participant: Undersweep fixation	0.302	0.550					0.335	0.579
Participant: Line length	0.273	0.523	1.762	1.328	0.255	0.505	1.274	1.129

Table 2. Results of linear mixed-effects model for the proportion of alignment and crossed and uncrossed fixations at fixation onset and offset. *Notes*: Significant *z* values ($|z| \geq 1.96$) are in bold. *SE* = standard error; *SD* = standard deviation.

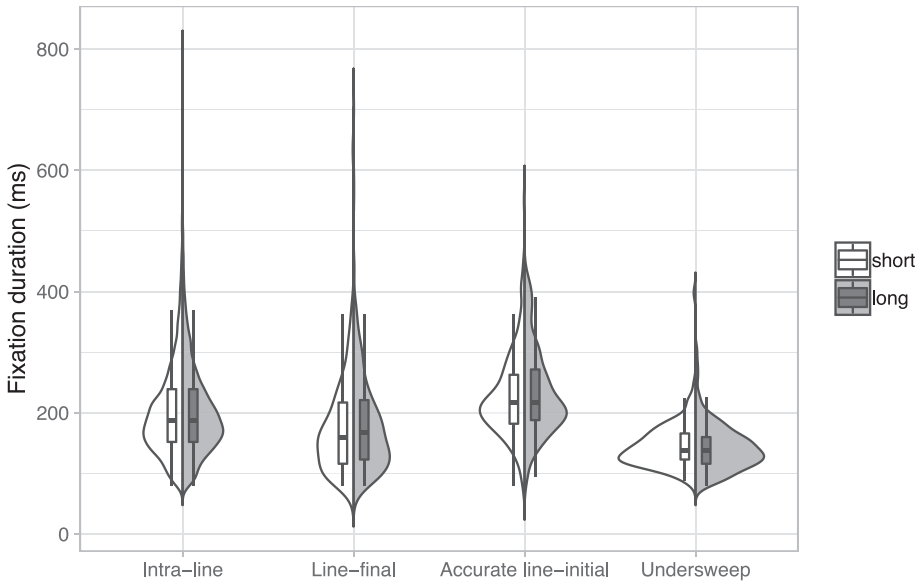

Figure 4. Split violin plot for fixation duration (ms) as a function of fixation population and line-length condition. The distribution of return-sweep landing position in the short-line condition is shown in white, and the distribution of fixation durations in the long-line condition is shown in gray. Box plots show the first quartile, median, and third quartile per fixation population.

Fixed effects	Inverse fixation duration		
Predictor	*b*	*SE*	*t*
Intercept	−5.622	0.108	**−51.92**
Line-final	−0.930	0.163	**−5.72**
Accurate line-initial	0.891	0.131	**6.78**
Undersweep	−1.784	0.205	**−8.71**
Line length	−0.030	0.050	−0.60
Line-final × Line length	−0.090	0.092	−0.98
Accurate line-initial × Line length	−0.033	0.108	−0.30
Undersweep × Line length	0.102	0.128	0.80

Random effects	Variance	*SD*
Item: Intercept	0.008	0.087
Participant: Intercept	0.218	0.467
Participant: Line-final	0.247	0.498
Participant: Accurate line-initial	0.077	0.278
Participant: Undersweep fixation	0.342	0.585
Participant: Line length	0.111	0.333

Table 3. Results of linear mixed-effects model for inverse fixation durations per fixation population (−1,000/ms). *Notes*: Significant *t* values ($|t| \geq 1.96$) are in bold. *SE* = standard error; *SD* = standard deviation.

As indicated in Table 3, the pattern of fixation durations was as predicted for each fixation population. In line with previous work (e.g., Abrams & Zuber, 1972), line-final fixations were significantly shorter than intraline reading fixation by 23 ms. Similarly, undersweep-fixation durations were dramatically shorter than those of intraline reading fixations (61 ms). This is consistent with findings reported by Heller (1982). Conversely, accurate line-initial fixations were significantly longer than intraline reading fixations. This 27-ms difference was similar to the 30-ms difference reported by Rayner (1977). These differences between return-sweep fixations and intraline fixations were identical to those reported by Parker et al. (2019). However, there was no effect of line length on fixation durations, nor did line length interact with any fixation population. Such a pattern of results argues against Stern's (1978) assertion. Analysis of raw fixation times revealed an identical pattern of results.[4]

The relationship between fixation disparity and duration

The analysis of fixation duration indicated that while line-initial fixations were longer than intraline reading fixations, this duration did not differ between line-length conditions. This is interpreted as evidence against Stern's (1978) account of longer line-initial fixations. To further examine Stern's conjecture, we conducted an analysis that examined the relationship between fixation duration and the magnitude of disparity at fixation onset (see Figure 5). The LMM, fit to inverse fixation duration, included fixed categorical effects for line-initial fixation population and line length. The absolute magnitude of disparity at fixation onset was included as a centered continuous predictor. All possible interactions were included: lmer(dv∼ fixation population * line length * fixation disparity + (1 + fixation disparity | participants) + (1 | item)).

Fixed effects	Inverse fixation duration		
Predictor	*b*	*SE*	*t*
Intercept	−4.712	0.134	**−35.08**
Undersweep	−2.654	0.134	**−19.74**
Line length	−0.063	0.087	−0.72
Fixation disparity	0.020	0.326	0.06
Undersweep × Line length	0.171	0.135	1.26
Undersweep × Fixation disparity	0.040	0.449	0.09
Line length × Fixation disparity	−0.003	0.262	−0.01
Undersweep × Line length × Fixation disparity	−0.027	0.436	−0.06

Random effects	Variance	*SD*
Item: Intercept	0.001	0.001
Participant: Intercept	0.144	0.379
Participant: Fixation disparity	0.310	0.557

Table 4. Results of linear mixed-effects model for inverse line-initial fixation durations (−1,000/ms). *Notes*: Significant *t* values ($|t| \geq 1.96$) are in bold. *SE* = standard error; *SD* = standard deviation.

Examination of the regression coefficients (see Table 4) indicated that undersweeps were significantly shorter than accurate line-initial fixations. Furthermore, fixation duration did not vary as a function of line length, disparity magnitude at fixation onset, or interactions between fixed effects. This analysis provides some of the strongest evidence that line-initial fixation durations are not related to the magnitude of disparity at fixations onset. Analysis of raw fixation times revealed an identical pattern of results.[5]

Bayes-factor analysis

To assess the evidence for the critical null effects, we supplemented our analyses of fixation duration with Bayes-factor analysis (for a detailed review, see Wagenmakers, 2007). Bayes factors were computed using the lmBF() function from the BayesFactor package in the R environment (Version 0.9.12-4.2; Morey & Rouder, 2018), with 100,000 Monte Carlo iterations. In all analyses, we assumed the default Cauchy prior for effect size (for discussion, see Abbott & Staub, 2015). The Bayes factor for the model outlined earlier under Fixation durations across multiline texts, when compared against a denominator model that included only fixed effects for fixation

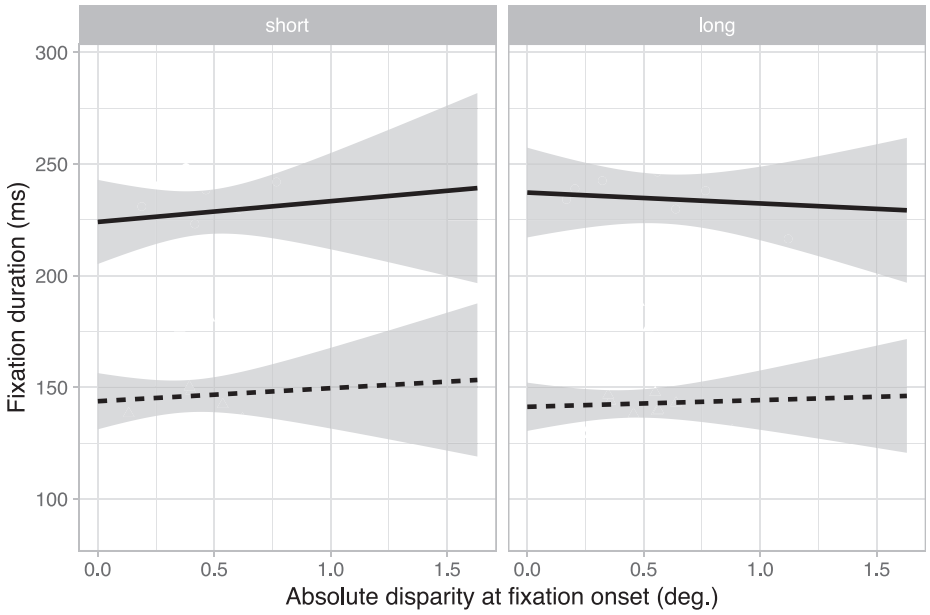

Figure 5. Fixation duration (ms) as a function of absolute disparity at fixation onset for line-initial fixations in the short- (left panel) and long-line (right panel) conditions. Accurate line-initial fixations are represented by the solid black line. Undersweep fixations are shown by the dashed black line. Gray bands show 95% confidence intervals.

population and line length (no interaction), was 0.007. Based on Jeffreys' (1961) evidence categories for Bayes factors, this provides very strong evidence in favor of the denominator model that did not include the interaction between fixation population and line length. The Bayes factor for the model outlined earlier under The relationship between fixation disparity and duration, when compared against a denominator model that used only line-initial fixation population and line-length condition as predictors of fixation duration, was <0.001. This provides extreme evidence in favor of the denominator model that did not include the magnitude of disparity at fixation onset as a predictor of fixation duration. Together, these analyses support the conclusion that line-initial fixation duration is not influenced by the magnitude of disparity at fixation onset.

Discussion

In recent years, there has been an increase in research on binocular coordination during normal continuous reading. However, this body of work is largely concerned with the reading of single-line text. The novel contributions of our current work can be summarized in two general points. First, we extended the existing literature on binocular coordination during reading by examining binocular coordination measures over multiline texts. This enabled us to assess how binocular disparity is influenced by long-distance return-sweep saccades. Second, we empirically examined Stern's (1978) assertion that longer line-initial fixations are the result of binocular coordination processes which compensate for divergence during the return-sweep. We discuss both points in turn.

At fixation onset all participants had, on average, 0.34° of disparity between the points of fixation of the two eyes. This corresponded to a disparity of 1.36 character spaces. Consistent with our predictions, when the magnitude of fixation disparity was examined for each fixation population there was an increased magnitude of disparity for line-initial fixations (both accurate and undersweep) in comparison to intraline reading fixations. We suggest that this increased magnitude of disparity at fixation onset results from transient divergence occurring during the return-sweep saccade. This is consistent with the body of literature reporting a linear relationship between the length of the preceding saccade and the magnitude of fixation disparity at fixation onset (Collewijn et al., 1988; Kirkby et al., 2010; Kirkby, Blythe et al., 2011). The magnitude of fixation disparity being greatest for accurate line-initial fixations in the long-line condition further supports this, as these return-sweeps would, on average, have traveled farther than any other saccade in the current reading study. It is also important to note that the magnitude of disparity at onset and offset did not differ between intraline and line-final fixations. If there were an increase in disparity across the line, line-final fixations may act to resolve this divergence in preparation for a return-sweep. However, our data lend no support to such a claim. Instead, it would appear that readers maintain a tight coupling of the two eyes'

visual axes across the line, with only saccade length influencing the disparity at onset.

Over the fixation period, fine-grained vergence movements maximize the degree of correspondence between the two disparate retinal images (Jainta & Jaschinski, 2012; Leigh & Zee, 2006). Thus, it is not surprising that despite clear differences in disparity for line-initial fixations at onset, the magnitude of fixation disparity was similar between intraline and return-sweep fixations at fixation offset. Blythe et al. (2010) reported that word identification was impaired when dichoptic presentation of single words induced disparities of 0.74° but not 0.37°. This led to the conclusion that retinal inputs of a word that are disparate by up to 0.37° fall within the effective fusional range, such that lexical decisions are unimpaired. Recall that in the short- and long-line conditions, accurate line-initial fixations were, on average, 0.46° and 0.58° disparate at fixation onset. At fixation offset, the magnitude of disparity for this population was 0.34°, comparable to the disparity at offset for intraline fixations. This indicates that readers attempt to maximize the correspondence between the two visual axes to fall within this range by fixation offset. However, given that the magnitude of disparity was greater at fixation onset and there was no relationship between the magnitude of disparity at fixation onset and fixation duration, it appears that the effective fusional range may be slightly larger than 0.37° during natural reading. Alternatively, assuming a reasoning similar to that of Hendriks (1996), readers may rely on ongoing linguistic and contextual processing when the magnitude of disparity falls outside of the effective fusional range, to ensure continuous processing of the text. The proportion of alignment for each fixation population further quantifies the role of the vergence system during the reading of multiline texts. Consistent with Liversedge, White, et al. (2006; see also Blythe et al., 2010; Blythe et al., 2006; Jainta et al., 2015; Juhasz et al., 2006; Kirkby et al., 2010; Kirkby et al., 2013; Kirkby, Blythe et al., 2011; Liversedge, Rayner, et al., 2006; Nikolova et al., 2015; Nikolova et al., 2017, 2018), we observed more aligned than unaligned fixations for intraline reading. However, a higher proportion of fixations were unaligned following a return-sweep saccade at fixation onset. The majority of these fixations were uncrossed. By fixation offset, the majority of fixations were aligned.

Stern (1978) argued that longer fixations at the start of the line result from vergence movements during line-initial fixations. That is, the eyes diverge during the return-sweep and then must reconverge during the first fixation on a line before normal reading processes can begin, thereby increasing the duration of these fixations. We conducted two analyses aimed at examining this explanation of longer line-initial fixations. Since the magnitude of disparity at fixation onset was greater for line-initial fixations in the long compared with the short condition, fixation durations in the long-line condition should be longer if Stern is correct. This was not the case. For line-initial fixations, the magnitude of disparity increased by 0.12° in the long-line condition. Yet this translated to a nonsignificant 6.1-ms difference in fixation duration between conditions. Bayes-factor analysis indicated that the data are more likely to be observed under the null hypothesis that fixation duration does not differ between line-length conditions. In the second analysis we directly examined the relationship between the magnitude of disparity at fixation onset and the duration of line-initial fixations. Given Stern's assertion, we predicted that an increased magnitude of disparity at fixation onset would result in increased durations of line-initial fixations. Again, our data show no such pattern. These null effects are of course not surprising. Kliegl et al. (2006) reported that durations of intraline reading fixations were unaffected by the magnitude of fixation disparity. However, what is novel from our data is that this holds for a specific population of fixations for which the magnitude of disparity at fixation onset far exceeds that observed in single-line reading. Together, these findings have implications for explanations of longer line-initial fixation durations and indicate that even when the magnitude of disparity is large at fixation onset (i.e., 2.3 characters on average), lexical processing can continue unhindered by this large disparity.

With regard to explanations of line-initial fixation durations, the current study provides evidence against a binocular account of longer line-initial fixations. Instead there exist two alternate explanations. The first relies on the premise that readers engage in a strategic oculomotor program of saccade planning over the line (Kuperman et al., 2010; Pynte & Kennedy, 2006). Rayner (1977) argued that when readers arrive at the start of the line, they have not yet had the opportunity to program their subsequent saccades over the line. Thus, longer fixation durations at the very start of the line result from establishing a grouped saccade program across the line. As of yet, there is a lack of empirical evidence to support this. Kuperman et al. have provided evidence of differential oculomotor behavior over the line during paragraph reading, yet their analysis of paragraph data excluded fixations either side of the return-sweep. As a result, it is difficult to draw conclusions based on their data. Instead, we advocate an account which argues that line-initial fixations are longer as a result of a lack of preview for information at the start of the line. That is, information at the very start of a line is unavailable for parafoveal preprocessing during the prior (line-final) fixation as it lies outside of the perceptual span. Therefore, it can only be processed following the return-sweep. This parallels studies in which accurate preview of an

upcoming word is prevented by masking the text. In a recent Bayesian meta-analysis, Vasilev and Angele (2017) estimated the preview-benefit effect to be 29 ms for the first-fixation duration. This is very similar to the 28-ms difference that we observed between intraline and accurate line-initial fixations in the current study. Additionally, Parker et al. (2017) observed longer gaze durations on line-initial words, where preview cannot be obtained, than on words presented midline, where they could be preprocessed prior to direct fixation. These results add weight to an account of longer line-initial fixations resulting from a lack of preview.

Disparity at fixation onset for accurate line-initial fixations was greatest in the long-line condition. However, these fixations did not differ statistically in duration from those in the short-line condition. This can be taken as evidence that the rate of lexical processing for line-initial fixations is largely unaffected by the magnitude of disparity at fixation onset. This has implications for the modeling of return-sweeps. Suppes (1994) argued that models of eye-movement control during reading would have to incorporate return-sweep saccades to be truly comprehensive. If the duration of accurate line-initial fixations did systematically vary as a function of the magnitude of disparity at fixation onset, models attempting to simulate the reading of multiline texts would have to account for the effects of binocular coordination on lexical processing. However, data from the current study warrant no such requirement. Instead, we argue that additional research is needed to explore how visual, attentional, and linguistic processes are influenced by return-sweeps. This research would provide benchmark findings from which models of eye-movement control (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle & Sheridan, 2015; Snell, van Leipsig, Grainger, & Meeter, 2018) could begin to simulate the reading of multiline texts.

Conclusions

The present data complement our current knowledge of binocular eye movements during the reading of single-line texts. To address the question, "Is binocular disparity larger following a return-sweep saccade than it is following an intraline saccade?", we first examined the basic characteristics of binocular coordination during the reading of multiline texts. Second, by directly examining the influence of binocular disparity on line-initial fixation durations we were able to empirically examine Stern's (1978) hypothesis that longer line-initial fixations are the result of binocular coordination processes. This enabled us to answer the question: Is the duration of a line-initial fixation related to binocular coordination processes? We report that following return-sweeps, there is an increased magnitude of fixation disparity at fixation onset, with the majority of fixations being uncrossed. This is taken as evidence for the linear relationship between the length of the preceding saccade and the magnitude of disparity at fixation onset. Despite being increased at fixation onset, the magnitude of disparity was similar between fixation populations at fixation offset. While Stern argued that increased divergence during a return-sweep would result in longer line-initial fixation durations, the duration of these fixations was unrelated to disparity at fixation onset. That is, the magnitude of disparity at fixation onset did not influence line-initial fixation durations. We instead propose that longer line-initial fixations result from a lack of parafoveal preview for information at the start of the line during the preceding fixation.

Keywords: binocular coordination, return-sweeps, eye movements, reading

Acknowledgments

This work was presented at the 2018 Meeting of the Psychonomic Society in New Orleans. It was made possible in part by a research grant from the Microsoft Corporation to TJS, which was matched by Bournemouth University to fund AJP's doctoral studies.

Commercial relationships: none.
Corresponding author: Adam J. Parker.
Email: parkera@bournemouth.ac.uk.
Address: Department of Psychology, Faculty of Science & Technology, Bournemouth University, Talbot Campus, Poole, UK.

Footnotes

[1] Initial fixations are also longer in free viewing and visual-search tasks. One candidate explanation for this increased duration is that viewers program a sequence of saccades during the initial fixation on a task (Zingale & Kowler, 1987).

[2] Presenting stimuli as black text on a white background would have increased the brightness of the stimuli. This increased brightness would have led to pupil shrinkage, making it more difficult to track each eye's Purkinje reflection. Red text was chosen to be consistent with prior experimental procedures and for consistency with calibration and validation points. Red circles were presented in order to minimize dichoptic

cross talk when calibrating the trackers with the shutter goggles.

[3] To assess the extent of data loss for each fixation population across line-length conditions, a generalized linear mixed-effects model was fit to all data points prior to data exclusion. Relative to the intercept (intraline fixations), data exclusion was more likely for line-final fixations, $b = 0.483$, $SE = 0.069$, $z = 5.23$. Compared to intraline fixations, data exclusion was reduced for accurate line-initial fixations, $b = -1.105$, $SE = 0.195$, $z = -5.66$, and undersweep fixations, $b = -0.446$, $SE = 0.171$, $z = -2.60$. Line-length condition did not influence data exclusion for any population ($zs < 1$).

[4] Test statistics for nontransformed data are as follows—intercept (intraline reading fixation), $b = 203.008$, $SE = 3.693$, $t = 54.97$; line-final: $b = -23.048$, $SE = 5.688$, $t = -4.05$; accurate line-initial: $b = 28.798$, $SE = 5.755$, $t = 5.004$; undersweep: $b = -60.662$, $SE = 6.737$, $t = -9.01$; line length: $b = -0.405$, $SE = 1.861$, $t = 0.22$; Line-final × Line length: $b = -4.166$, $SE = 3.481$, $t = -1.20$; Accurate line-initial × Line length: $b = -2.032$, $SE = 4.171$, $t = -0.49$; Undersweep × Line length: $b = 0.046$, $SE = 4.774$, $t = 0.01$.

[5] Test statistics for nontransformed data are as follows—intercept (accurate line-initial fixation): $b = 232.571$, $SE = 5.029$, $t = 46.25$; undersweep: $b = -89.424$, $SE = 4.925$, $t = -18.16$; line length: $b = -2.897$, $SE = 3.176$, $t = -0.91$; fixation disparity: $b = 8.317$, $SE = 12.072$, $t = 0.69$; Undersweep × Line length: $b = 3.334$, $SE = 4.955$, $t = 0.67$; Undersweep × Fixation disparity: $b = -5.214$, $SE = 16.451$, $t = -0.32$; Line length × Fixation disparity: $b = 4.130$, $SE = 9.604$, $t = 0.43$; Undersweep × Line length × Fixation disparity: $b = -9.315$, $SE = 15.944$, $t = -0.58$.

References

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. *Journal of Memory and Language*, *85*, 76–87.

Abrams, S. G., & Zuber, B. L. (1972). Some temporal characteristics of information processing during reading. *Reading Research Quarterly*, *8*, 40–51.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*: 328.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2019). Parsimonious mixed models. arXiv1506.04967 2015

Bates, et al. (2018). lme4: Linear Mixed-Effects Models using 'Eigen' and S4. R package version 1.1-17.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.

Becker, W. (1976). Do corrective saccades depend exclusively on retinal feedback? A note on the possible role of non-retinal feedback. *Vision Research*, *41*, 425–427.

Beymer, D., Russell, D. M., & Orton, P. Z. (2005). Wide vs. narrow paragraphs: An eye tracking analysis. In F. Costabile & F. Paternò (Eds.), *IFIP Conference on Human-Computer Interaction* (pp. 741–752). Berlin, Germany: Springer.

Blythe, H. I., Liversedge, S. P., & Findlay, J. M. (2010). The effective fusional range for words in a natural viewing situation. *Vision Research*, *50*(16), 1559–1570.

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision Research*, *46*(22), 3898–3908.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(9), 1–20.

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *Journal of Physiology*, *404*, 157–182.

Coltheart, M., & Freeman, R. (1974). Case alternation impairs word identification. *Bulletin of the Psychonomic Society*, *3*(2), 102–104.

Deubel, H., & Bridgeman, B. (1995). Perceptual consequences of ocular lens overshoot during saccadic eye movements. *Vision Research*, *35*(20), 2897–2902.

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

Hawley, T. T., Stern, J. A., & Chen, S. C. (1974). Computer analysis of eye movements during reading. *Literacy Research and Instruction*, *13*, 307–317.

Heller, D. (1982). Eye movements in reading. In R. Groner & P. Fraisse (Eds.), *Cognition and eye movements* (pp. 487–498). Berlin, Germany: Deutscher Verlag der Wissenschaften.

Heller, D., & Radach, R. (1999). Eye movements in reading: Are two eyes better than one? In W. Becker, H. Deubel, & T. Mergner (Eds.), Current oculomotor research: Physiological and psychological aspects (pp. 341–348). New York, NY: Plenum Press.

Hendriks, A. W. (1996). Vergence eye movements during fixations in reading. *Acta Psychologica*, *92*, 131–151.

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*(8), 180502.

Hofmeister, J. (1997). On corrective saccades in reading and reading-like tasks (Unpublished doctoral dissertation). Rheinisch-Westfälische Technische Hochschule, Aachen, Germany.

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker, H. Deubel, & T. Mergner (Eds.), Current oculomotor research: Physiological and psychological aspects (pp. 349–357). New York, NY: Plenum Press.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York, NY: Oxford University Press.

Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth: Vol. 2, Stereoscopic vision*. Oxford, UK: Oxford University Press.

Jainta, S., Blythe, H. I., & Liversedge, S. P. (2014). Binocular advantages in reading. *Current Biology*, *24*(5), 526–530.

Jainta, S., Blythe, H. I., Nikolova, M., Jones, M. O., & Liversedge, S. P. (2015). A comparative analysis of vertical and horizontal fixation disparity in sentence reading. *Vision Research*, *110*, 118–127.

Jainta, S., Hoormann, J., Kloke, W. B., & Jaschinski, W. (2010). Binocularity during reading fixations: Properties of the minimum fixation disparity. *Vision Research*, *50*(18), 1775–1785.

Jainta, S., & Jaschinski, W. (2012). Individual differences in binocular coordination are uncovered by directly comparing monocular and binocular reading conditions. *Investigative Ophthalmology & Visual Science*, *53*, 5762–5769.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York, NY: Oxford University Press.

Juhasz, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: Word frequency and case alternation affect fixation duration but not fixation disparity. *The Quarterly Journal of Experimental Psychology*, *59*, 1614–1625.

Kirkby, J. A., Blythe, H. I., Benson, V., & Liversedge, S. P. (2010). Binocular coordination during scanning of simple dot stimuli. *Vision Research*, *50*(2), 171–180.

Kirkby, J. A., Blythe, H. I., Drieghe, D., Benson, V., & Liversedge, S. P. (2013). Investigating eye movement acquisition and analysis technologies as a causal factor in differential prevalence of crossed and uncrossed fixation disparity during reading and dot scanning. *Behavior Research Methods*, *45*(3), 664–678.

Kirkby, J. A., Blythe, H. I., Drieghe, D., & Liversedge, S. P. (2011). Reading text increases binocular disparity in dyslexic children. *PLoS One*, *6*(11), 1–7.

Kirkby, J. A., Webster, L. A. D., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin*, *134*(5), 742–763.

Kirkby, J. A., White, S. J., & Blythe, H. I. (2011). Binocular coordination during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 801–818). Oxford, UK: Oxford University Press.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35.

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology*, *63*(9), 1838–1857.

Leigh, R. J., & Zee, D. S. (2006). *The neurology of eye movements*. New York, NY: Oxford University Press.

Liversedge, S. P., Rayner, K., White, S. J., Findlay, J. M., & McSorley, E. (2006). Binocular coordination of the eyes during reading. *Current Biology*, *16*(17), 1726–1729.

Liversedge, S. P., White, S. J., Findlay, J. M., & Rayner, K. (2006). Binocular coordination of eye movements during reading. *Vision Research*, *46*(15), 2363–2374.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-move-

ments in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*(3), 266–293.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Nikolova, M., Jainta, S., Blythe, H. I., Jones, M. O., & Liversedge, S. P. (2015). Vergence responses to vertical binocular disparity during lexical identification. *Vision Research*, *106*, 27–35.

Nikolova, M., Jainta, S., Blythe, H. I., & Liversedge, S. P. (2017). Using a dichoptic moving window presentation technique to investigate binocular advantages during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 265–280.

Nikolova, M., Jainta, S., Blythe, H. I., & Liversedge, S. P. (2018). Binocular advantages for parafoveal processing in reading. *Vision Research*, *145*, 56–63.

Nuthmann, A., & Kliegl, R. (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision*, *9*(5):31, 1–28, https://doi.org/10.1167/9.5.31. [PubMed] [Article]

Parker, A. J., Kirkby, J. A., & Slattery, T. J. (2017). Predictability effects during reading in the absence of parafoveal preview. *Journal of Cognitive Psychology*, *29*(8), 902–911.

Parker, A. J., Slattery, T. J., & Kirkby, J. A. (2019). Return-sweep saccades during reading in adults and children. *Vision Research*, *155*, 35–43.

Paterson, D. G., & Tinker, M. A. (1940). *How to make type readable*. Oxford, UK: Harper.

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013). Aging and the control of binocular fixations during reading. *Psychology and Aging*, *28*(3), 789–795.

Pratt-Johnson, J. A., & Tillson, G. (2001). *Management of strabismus and amblyopia: A practical guide*. New York, NY: Thieme.

Pynte, J., & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research*, *46*(22), 3786–3801.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, *5*, 443–448.

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

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, *12*(2), 206–226.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Reichle, E. D., & Sheridan, H. (2015). E-Z reader: An overview of the model and two recent applications. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 277–290). Oxford, UK: Oxford University Press.

Reingold, E. M., Yang, J., & Rayner, K. (2010). The time course of word frequency and case alternation effects on fixation times in reading: Evidence for lexical control of eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1677–1683.

Schneps, M. H., Thomson, J. M., Sonnert, G., Pomplun, M., Chen, C., & Heffner-Wong, A. (2013). Shorter lines facilitate reading in those who struggle. *PLoS One*, *8*(8), 1–16.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, *74*(1), 5–35.

Schotter, E. R., Blythe, H. I., Kirkby, J. A., Rayner, K., Holliman, N. S., & Liversedge, S. P. (2012). Binocular coordination: Reading stereoscopic sentences in depth. *PLoS One*, *7*(4), 1–12.

Shebilske, W. L. (1975). Reading eye movements from an information-processing point of view. In W. Massaro (Ed.), *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. New York, NY: Academic Press.

Shillcock, R., Roberts, M., Kreiner, H., & Obregón, M. (2010). Binocular foveation in reading. *Attention, Perception, & Psychophysics*, *72*(8), 2184–2203.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*, 969–984.

Stern, J. A. (1978). Eye movements, reading and cognition. In J. W. Denders, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological functions* (pp. 145–155). Hillsdale, NJ: Lawrence Erlbaum Associates.

Suppes, P. (1994). Stochastic models of reading. In J. Ygge & G. Lennerstrand (Eds.), *Eye movements in reading* (pp. 349–364). Oxford, UK: Pergamon.

Tinker, M. A. (1963). *Legibility of print*. Ames, IA: Iowa State University Press.

van Heuven, W. J. B., Mandera, P., Keuleers, E., &

Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N + 1 and word N + 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, *24*, 666–689.

Vernet, M., & Kapoula, Z. (2009). Binocular motor coordination during saccades and fixations while reading: A magnitude and time analysis. *Journal of Vision*, *9*(7):2, 1–13, https://doi.org/10.1167/9.7.2. [PubMed] [Article]

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Worth, C. (1921). *Squint: Its causes, pathology and treatment*. Philadelphia, PA: Blakiston.

Zee, D. S., Fitzgibbon, E. J., & Optican, L. M. (1992). Saccade-vergence interactions in humans. *Journal of Neurophysiology*, *68*, 1624–1641.

Zingale, C. M., & Kowler, E. (1987). Planning sequences of saccades. *Vision Research*, *27*(8), 1327–1341.