The effect of contextual plausibility on word skipping during reading

Aaron Veldre[1], Erik D. Reichle[2], Roslyn Wong[1], & Sally Andrews[1]

[1] The University of Sydney, School of Psychology

[2] Macquarie University, Department of Psychology

Short title: Plausibility effects on word skipping

Contact details

Aaron Veldre

School of Psychology, The University of Sydney

Sydney, NSW 2006

Australia

aaron.veldre@sydney.edu.au

+61 2 9351 8990

Abstract

Recent eye-movement evidence suggests readers are more likely to skip a high-frequency word than a low-frequency word independently of the semantic or syntactic acceptability of the word in the sentence. This has been interpreted as strong support for a serial processing mechanism in which the decision to skip a word is based on the completion of a preliminary stage of lexical processing prior to any assessment of contextual fit. The present large-scale study was designed to reconcile these findings with the plausibility preview effect: higher skipping and reduced first-pass reading times for words that are previewed by contextually plausible, compared to implausible, sentence continuations that are unrelated to the target word. Participants' eye movements were recorded as they read sentences containing a short (3-4 letters) or long (6 letters) target word. The boundary paradigm was used to present parafoveal previews which were either higher or lower frequency than the target, and either plausible or implausible in the sentence context. The results revealed strong, independent effects of all three factors on target skipping and early measures of target fixation duration, while frequency and plausibility interacted on later measures of target fixation duration. Simulations using the E-Z Reader model of eye-movement control in reading demonstrated that plausibility effects on skipping are potentially consistent with the assumption that higher-level contextual information only affects post-lexical integration processes. However, no current model of eye movements in reading provides an explicit account of the information or processes that allow readers to rapidly detect an integration failure.

Keywords
Reading, eye movements, word skipping, plausibility effects, parafoveal processing, E-Z Reader

Word count: 10,314 (excluding Abstract, References, Footnotes, Tables, and Figures)

1. Introduction

Skilled readers process upcoming information in the parafoveal region of vision. Parafoveal processing facilitates reading fluency because the identification of a word can begin before it is directly fixated. Readers can also make use of parafoveal information to decide to *skip* a word entirely. Approximately 30% of words do not receive a direct fixation during reading (Rayner, 2009). By far the strongest predictor of skipping is word length because short words are much more likely to be skipped than long words: 2-3–letter words are skipped approximately 75% of the time, while skipping is rare for 8-letter words (Rayner & McConkie, 1976). The strong effect of word length on skipping is consistent with the distinction between *when* and *where* processes in oculomotor control (e.g., Inhoff, Radach, Eiter, & Juhasz, 2003; Rayner & Pollatsek, 1981): skipping is determined by the extraction of low-level visual information from the parafovea that is used to determine *where* to land the eyes next.

These effects of coarse parafoveal processing were initially assumed to be independent of the fine-grained processing required for the lexical retrieval processes that determine *when* to move the eyes, which depend on foveal vision (e.g., Rayner & Pollatsek, 1981). However, subsequent empirical data demonstrated that parafoveal processing extends beyond coarse, low-level features. There is robust evidence that word skipping is also affected by linguistic factors. Demonstrating that lexical processing affects the likelihood that a word will receive a direct fixation, high-frequency words are skipped more than low-frequency words (e.g., Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner & Duffy, 1986; Schilling, Rayner, & Chumbley, 1998). Characteristics of the sentence or discourse context also impact skipping: Words that are highly predictable from the context—as indexed by their cloze probability—are skipped more than unpredictable words (e.g., Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek,

2005; Ehrlich & Rayner, 1981). Brysbaert, Drieghe, and Vitu's (2005) meta-analysis estimated that, controlling for word length, the effect sizes of frequency and predictability on skipping were approximately 5% and 8%, respectively, providing clear evidence that skipping is affected by lexical information extracted from parafoveal vision.

Even stronger evidence of the early impact of deep parafoveal processing on oculomotor control processes is provided by recent evidence that the *plausibility* of a word in a sentence context also influences the earliest measures of reading, including word skipping, even for words that are very low in cloze probability. The depth of parafoveal preprocessing is typically indexed by estimating the *preview benefit* using the *gaze-contingent boundary paradigm* (Rayner, 1975). In the boundary paradigm, a critical target word is either available for parafoveal processing or is masked by an invalid preview (e.g., an illegal consonant string, or an unrelated word) until the reader moves their eyes to the target word (see Schotter, Angele, & Rayner, 2012, for a review). The reader's saccade across an invisible boundary located at the end of the pre-target word triggers a rapid display change that replaces the preview stimulus with the target word so that the reader only ever fixates on the correct version of the target word. A difference in fixation duration between targets preceded by valid (i.e., identical to the target) and invalid (i.e., unrelated to the target) previews is therefore taken as evidence that parafoveal processing occurred, and manipulations of the similarity/relatedness of the preview to the target can be used to assess what attributes of parafoveal words were processed (Schotter et al., 2012).

Vasilev and Angele's (2017) meta-analysis estimated the size of the *preview effect* for an identical preview on gaze duration (the sum of first-pass fixations on a word) to range from 30-50 ms, depending the type of baseline used (i.e., unrelated word, pseudoword, random consonant string, or x-string mask). The fact that the size of the preview effect depends on the baseline

condition implies that it is composed of both a benefit from preprocessing the target word in the identical condition and a cost from previewing unrelated—and often orthographically illegal—information in the invalid condition (e.g., Hutzler, Schuster, Marx, & Hawelka, 2019; Kliegl, Hohenstein, Yan, & MacDonald, 2013; Veldre & Andrews, 2018b).

Further evidence of a direct influence of properties of the preview on oculomotor planning is provided by the *plausibility-preview effect*. Veldre and Andrews (2016, 2017, 2018a, 2018b, 2018c) used the boundary paradigm to compare parafoveal word previews that were either contextually plausible or implausible continuations of the sentence. These studies showed that first-pass fixation duration on an unrelated target word was approximately 20 ms shorter when the preview was contextually plausible than when it was implausible (see also Schotter & Jia, 2016; Yang, Li, Wang, Slattery, & Rayner, 2014; Yang, Wang, Tong, & Rayner, 2014). The effect of preview plausibility was found to be independent of any contribution of semantic relatedness of the preview to the target word, i.e. a plausible unrelated word yielded as much benefit as a synonym of the target (Veldre & Andrews, 2016; see also Schotter & Jia, 2016). The effect was also demonstrated to be independent of the well-established orthographic preview effect observed when the preview and target share overlapping orthographic features (Schotter et al., 2012) —the size of the plausibility-preview effect was equivalent for both orthographic neighbors of the target and orthographically unrelated words (Veldre & Andrews, 2017). The plausibility preview benefits described above occurred for words that were low in cloze predictability, but the early benefit from previewing a plausible word was observed even when the sentence context strongly constrained towards a different word (Veldre & Andrews, 2018c). Importantly, Veldre and Andrews (2018a) demonstrated that the plausibility-preview effect reflected both semantic and syntactic acceptability (see Andrews & Veldre, 2019, for a review).

Critically, in addition to the preview effect on fixation duration, several studies have also shown an effect of preview plausibility on skipping of the target word, while carefully controlling the frequency and contextual predictability of the parafoveal items (Veldre & Andrews, 2017, 2018a, 2018b, 2018c). For example, Veldre and Andrews (2018a) manipulated the semantic and syntactic acceptability of parafoveal previews in two experiments and found that, relative to a plausible preview, readers were less likely to skip either a semantically or syntactically anomalous word. Although the two types of violation yielded qualitatively different patterns of effects on fixation duration and regressions out of the target words, skipping rates for these previews did not differ. Several independent studies have reported similar effects of plausibility, syntactic acceptability, or grammaticality on skipping rates (e.g., Brothers & Traxler, 2016; Matsuki et al., 2011; Staub, 2011).

This evidence of plausibility-preview effects on fixation duration and word skipping has important implications for understanding the role of parafoveal processing in reading. First, the data briefly reviewed above add to a growing body of evidence that preview effects are not solely due to the trans-saccadic integration of parafoveal and foveal information (e.g., Risse & Kliegl, 2012, 2014; Risse & Seelig, 2018; Schotter & Leinenger, 2016; Schotter, Leinenger, & von der Malsburg, 2018; Schotter, von der Malsburg, & Leinenger, 2018), as assumed in previous accounts of preview effects (e.g., Cutter, Drieghe, & Liversedge, 2015; Pollatsek, Lesch, Morris, & Rayner, 1992; Rayner, 1975). Secondly, the evidence that plausibility effects are observed on very early measures, including the oculomotor planning processes reflected in skipping, appears to challenge the common assumption that contextual influences on sentence processing arise from post-lexical integration processes because it seems unlikely that integration

would occur rapidly enough to influence decisions about whether or not to fixate an upcoming word.

However, it is important to reconcile the evidence of plausibility-preview effects with several recent studies that have failed to find an influence of plausibility on skipping rates. Abbott and Staub (2015) factorially manipulated the frequency and plausibility of a critical target word in normally presented sentences like 1(a) and 1(b).

1. (a) The professor invited the *writer*/*orator* to an important meeting.

   (b) The professor repaired the *writer*/*orator* with a trusty old wrench.

While frequency yielded a 3.5% effect on target skipping, there was no influence of plausibility. Direct evidence supporting this null effect of plausibility was provided by a Bayesian analysis. The most obvious explanation for the apparent contradiction between the plausibility-preview effects reviewed above and Abbott and Staub's (2015) findings is that the latter study manipulated the plausibility of targets in foveal vision, rather than using parafoveal previews that are never directly fixated. Foveal presentation of implausible targets alerts readers to the critical manipulation. Abbott and Staub's participants were instructed before the experiment that some sentences would be "a little weird" (p. 80), which may have caused them to adopt a more cautious reading strategy from the very beginning of the task, selectively reducing skipping rates. The discrepant findings may also arise because plausibility effects on skipping are limited to short words, which are more likely to be fully processed in parafoveal vision. Abbott and Staub's target words ranged from 5 to 8 letters in length (M = 6.51) while the majority of the target/preview words used by Veldre and Andrews (2017, 2018a, 2018b, 2018c) were 5 letters in length (range: 4 to 7 letters). Whether because of slightly longer words or a different reading

strategy, the very low rate of skipping in Abbott and Staub's experiment (6.7%, on average) may have obscured a small effect of plausibility.

The possibility that plausibility effects on skipping are restricted to short words is, however, contradicted by evidence that readers are more likely to skip a preview of a short function word, even when it is not a syntactically legal continuation of the sentence. Angele and Rayner (2013) found that readers were more likely to skip a preview of the article *the* than a contextually plausible word (see Zang et al., 2018, for similar evidence in Chinese). Skipping rates for a preview of *the* were also higher than for a highly predictable target word, but sentence constraint also yielded an effect on skipping rates independent of the preview word (Abbott, Angele, Ahn, & Rayner, 2015). Taken together, these data suggest that readers "do not seem to take syntactic information into account when planning an eye movement" (Angele & Rayner, 2013, p. 656). While it is perhaps unsurprising that readers skip an extremely high-frequency word such as *the* more often than a low-frequency content word, the absence of any contribution of syntactic acceptability or semantic plausibility to skipping seems incompatible with the robust evidence of *predictability* effects on skipping (e.g., Brysbaert et al., 2005), because cloze probability encompasses the syntactic and semantic 'fit' of a word.

To assess whether the lack of plausibility effects on skipping of short words generalized beyond very high-frequency function words, Angele, Laishley, Rayner, and Liversedge (2014) compared identical (plausible) previews to higher or lower frequency previews that were syntactically and/or semantically anomalous continuations of sentences like in 2(a) and 2(b).

2. (a) The excitable (*dog/dim*) dog was eager to go for his walk.

   (b) The increasingly (*dog/dim*) dim light made it hard to see.

Their critical finding was an interaction between target frequency and preview validity that reflected increased skipping of higher frequency words relative to lower frequency words regardless of syntactic fit. While Angele et al.'s data suggest that frequency outweighs plausibility in skipping of short words, the critical manipulation in their study occurred between items—that is, the sentence context was not held constant across all conditions. This is a problem for interpreting their effects because the targets/previews were not matched on cloze probability, which ranged from 0 to 91%. The authors also failed to collect validation data to confirm that the low- and high-frequency targets were equally plausible continuations of the context. More recently, Zang, Du, Bai, Yan, and Liversedge (2019) showed that readers of Chinese skip high-frequency words more than low-frequency words, but found that syntactic fit also affected the skipping of high-frequency words. Further evidence is therefore necessary to determine the generalizability of plausibility-preview effects, and the extent and manner in which they are modulated by target length and frequency. This is the goal of the present research.

Disentangling the relative contribution of these stimulus attributes to parafoveal preview effects is important because of their significance for current models of eye-movement control in reading. A limited impact of plausibility on word skipping rates appears to be compatible with the assumptions of the *E-Z Reader* model of eye movements (Reichle, Pollatsek, & Rayner, 2012; Reichle, Pollatsek, Fisher, & Rayner, 1998). As illustrated in Figure 1, E-Z Reader assumes that lexical processing proceeds in a strictly serial fashion and that attention is allocated to one word at a time. The planning of an eye movement to the next word is triggered by the completion of an initial *familiarity check* stage of lexical processing (L_1) whose duration is directly determined by the word's frequency and predictability. Word length also indirectly modulates the duration of L_1 because words that extend further from the point of fixation are

processed more slowly due to reduced visual acuity. During the delay between the initial planning of the saccade and its execution, lexical processing of the fixated word completes (L_2) and covert attention shifts to the parafovea to begin lexical processing of the upcoming word. Critically, if L_1 for the parafoveal word completes during the initial labile stage of saccadic planning (M_1), the saccade to that word is cancelled and a new saccade is programmed to the following word, resulting in a skip.

Thus, according to E-Z Reader, the decision to skip a word is based on the completion of an early stage of lexical processing that is determined by a word's frequency, cloze probability, and length.[1] E-Z Reader 10 introduced a post-lexical *integration* stage (*I*) to account for the impact of a word's plausibility on eye movements (Reichle, Warren, & McConnell, 2009). Post-lexical integration difficulty causes the cancellation of planned forward saccades and the programming of refixations and regressions back to the source of the processing difficulty. According to the model, post-lexical integration difficulty reflects two sources (see Figure 2 of Reichle et al., 2009). The first occurs whenever word$_N$ has not been integrated prior to the identification of word$_{N+1}$. This situation is assumed to be problematic for the incremental construction of a sentence representation and the probability of it occurring reflects the relative times required to complete *I* for word$_N$ (mean = 25 ms) versus the sum of the durations of L_1 and L_2 for word$_{N+1}$. The second source of integration difficulty is assumed to reflect problems with sentence processing (e.g., syntactic parsing failure). Because E-Z Reader does not actually simulate sentence processing, the model is limited to specifying the probability of this occurring (p_F = 0.01) for any given word.

[1] A non-negligible amount of word skipping also occurs due to saccadic error; saccades directed towards an upcoming word overshoot their target, resulting in word skipping.

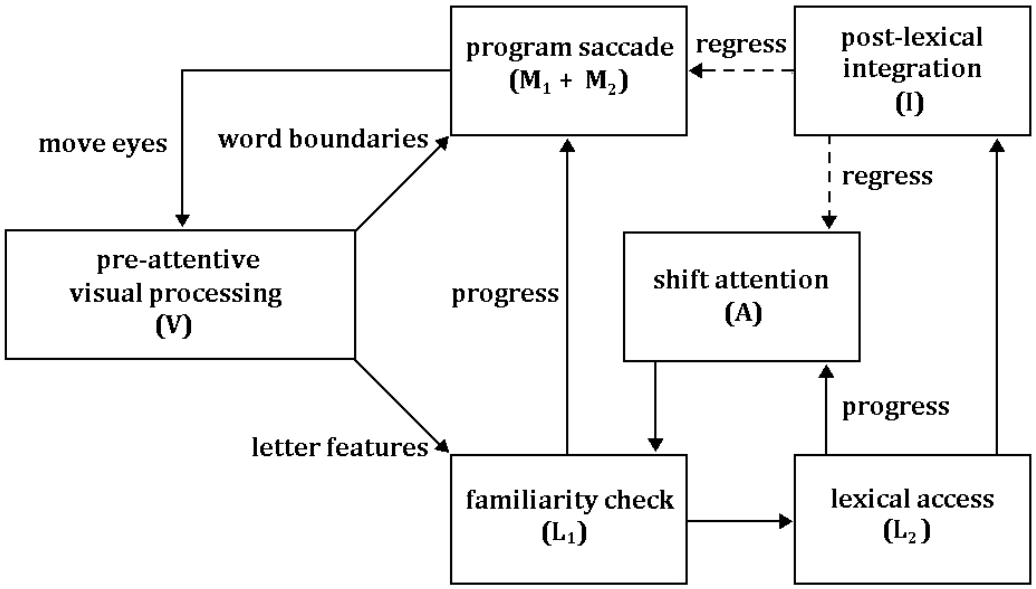

Figure 1. A schematic diagram of the *E-Z Reader* model. During pre-attentive visual processing (*V*), word boundaries are used for saccadic programming (M_1 + M_2) and letter features are used for lexical processing. The latter is completed in two stages: an initial familiarity check (L_1) that triggers saccadic programming, and the completion of lexical access (L_2), which shifts attention (*A*) and initiates post-lexical processing (*I*). Failure of the latter (e.g., because a word is implausible) can produce a pause or regression back to the problematic word.

Because of the timing constraints imposed by the coordination of serial lexical processing and saccade planning, the model would seemingly predict that post-lexical integration should rarely occur for parafoveal words. For plausibility to affect skipping, for example, lexical access and post-lexical processing for an upcoming word (i.e., L_2 + I for word$_{N+1}$) would have to be completed during the labile stage of saccadic planning (i.e., M_1 to move the eyes to word$_{N+2}$) to allow the saccade to skip word$_{N+1}$ to be cancelled and a new saccade to move the eyes to the parafoveal word to be programmed. Simulations reported by Abbott and Staub (2015) suggest that this sequence of events should only occur rarely, even for short words, and that plausibility should have a negligible impact on skipping of words of 5 or more letters in length. Both Abbott

and Staub and Veldre and Andrews (2018a) conducted post-hoc analyses that revealed no evidence that plausibility effects were restricted to short words. However, no clear conclusions can be drawn from these analyses given that frequency, predictability, and plausibility were not carefully controlled for different word lengths, and that differences in these factors may have contributed to the skipping effects.

Thus, the aim of the present study was to clarify conflicting results in the literature by factorially manipulating the length, frequency, and plausibility of non-identical parafoveal preview words presented in the same unconstraining sentence frames. The use of the boundary paradigm allowed us to conduct a further test of the plausibility-preview effect and provided novel evidence about the relative contribution of length, frequency, and plausibility to word skipping and parafoveal-preview effects that has important implications for refining theories of eye-movement control in reading.

To anticipate the critical outcomes, the data revealed significant plausibility-preview effects on skipping for both short and long words that are low in cloze probability which appear to be incompatible with the intuitive predictions of accounts that attribute contextual plausibility effects solely to post-lexical integration, such as the E-Z Reader model. However, direct simulation of the present experiment using E-Z Reader showed that it successfully predicted the effects of preview plausibility obtained in the empirical data.

2. Method

2.1 Participants

The final sample comprised 104 undergraduate students (mean age: 19.84 years) from The University of Sydney.[2] All had normal or corrected-to-normal vision and reported that

[2] An additional two participants were excluded because their comprehension accuracies (70% and 72.5%) were more than 2.5 SDs below the mean.

English was the first language they learned to read and write. Participants received partial course credit as compensation.

2.2 Materials and Design

The stimuli were 160 sentences, 8-18 words in length (*M*=12.29), in which a critical target word was embedded. In 80 of the sentences the target was a *short* 3- or 4-letter word, and in the remaining 80 sentences the target was a *long* 6-letter word. The target word was located towards the middle of the sentence and was preceded by a high-frequency (*M*=653.04 occurrences per million) pre-target word (*M*=6.16 letters in length). The boundary paradigm was used to compare four non-identical parafoveal preview conditions that factorially manipulated frequency and plausibility (see Figure 2). All previews were orthographically and semantically unrelated to the target word. The plausible previews were acceptable continuations of the pre-target sentence context. The implausible previews produced semantic and/or syntactic violations. The sentence materials were constructed in pairs so that the same preview was plausible in one sentence and implausible in the other sentence. Although participants were only exposed to each target word once, all sentences appeared in all preview conditions across four counterbalanced lists.

1.(a) The first course was very hot but was a good start to the meal.
  (b) The first course was very icy but was a good start to the meal.
  (c) The first course was very boy but was a good start to the meal.
  (d) The first course was very imp but was a good start to the meal.

2.(a) The pilot was able to easily change the direction of the plane.
  (b) The pilot was able to easily invert the direction of the plane.
  (c) The pilot was able to easily person the direction of the plane.
  (d) The pilot was able to easily trader the direction of the plane.

Figure 2. Example sentences in each of the four preview conditions: (a) high frequency/plausible, (b) low frequency/plausible, (c) high frequency/implausible, (d) low frequency/implausible. When the reader made a saccade across the invisible boundary, indicated by the dotted line, the preview was replaced by the medium-frequency target word. In sentence (1) the target word was *raw*; in sentence (2) the target word was *assess*.

2.2.1 Stimulus validation. A separate group of 25 participants from the same population of undergraduate students provided cloze norming data. They were given each sentence frame up to and including the pre-target word and asked to provide the word that was most likely to come next. The results of the cloze task showed that the target word and the high- and low-frequency plausible preview words were on average generated less than 4% of the time (range: 0-44%) confirming that they were not highly predictable from the sentence context. The high- and low-frequency implausible preview words were never generated as responses in the cloze task (see Table 1).

To confirm the plausibility manipulation, a further group of 25 participants provided plausibility ratings on a 5-point scale for the sentence frames up to and including each of the preview words. The sentence fragments ending with the target, high-frequency plausible, and low-frequency plausible previews were rated as higher in acceptability than high-frequency

implausible and low-frequency implausible previews, which did not differ. The stimulus characteristics and norming data are reported in Table 1.

2.3 Apparatus

An SR Research EyeLink 1,000 Plus system was used to record participants' eye movements at 2,000 Hz as they read sentences on a 21-inch CRT monitor, which had a refresh rate of 140 Hz. The sentences occupied a single line and were presented in black monospaced font on a gray background. Viewing was binocular but fixation position was monitored from the right eye. Participants were seated 60 cm from the monitor and a chin and forehead rest was used to minimize head movements. At this distance 2.5 characters subtended 1 degree of visual angle.

2.4 Procedure

Participants were instructed to read the sentences for meaning and to respond to occasional comprehension questions. The experiment began with a three-point calibration procedure followed by three practice trials and the 160 experimental trials, which were presented in an individually randomized order and intermixed with 22 filler sentences containing no display changes. At the beginning of each trial a fixation point appeared at the location of the first letter of the sentence. Once the participant made a stable fixation on this point, the sentence was displayed, or a new calibration procedure was performed if necessary. Mean calibration error was less than 0.3 degrees of visual angle. The participant pressed a key when they finished reading the sentence. On all practice trials and approximately 30% of experimental and filler trials, the sentence was followed by a multiple-choice comprehension question that required a moderate understanding of the meaning of the sentence.

Table 1

Mean (and Standard Error) Stimulus Characteristics and Norming Data

Variable	Short target word				Long target word			
	HF Plausible	LF Plausible	HF Implausible	LF Implausible	HF Plausible	LF Plausible	HF Implausible	LF Implausible
Target frequency[a]	39.73 (3.48)	39.73 (3.48)	39.73 (3.48)	39.73 (3.48)	38.75 (3.15)	38.75 (3.15)	38.75 (3.15)	38.75 (3.15)
Preview frequency[a]	226.98 (16.84)	6.13 (0.55)	226.98 (16.84)	6.13 (0.55)	224.54 (15.67)	3.03 (0.29)	224.54 (15.67)	3.03 (0.29)
Letter overlap with target[b]	.09 (.02)	.10 (.02)	.11 (.02)	.10 (.02)	.10 (.02)	.12 (.02)	.09 (.01)	.08 (.01)
Target cloze probability	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)
Preview cloze probability	.04 (.01)	.01 (.00)	.00 (.00)	.00 (.00)	.02 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Target plausibility[c]	4.21 (0.08)	4.21 (0.08)	4.21 (0.08)	4.21 (0.08)	4.03 (0.12)	4.03 (0.12)	4.03 (0.12)	4.03 (0.12)
Preview plausibility[c]	4.30 (0.07)	4.15 (0.12)	1.54 (0.08)	1.74 (0.11)	4.12 (0.09)	3.97 (0.10)	1.38 (0.08)	1.66 (0.11)

[a] Frequency values extracted from the CELEX corpus (Baayen, Piepenbrock, & Van Rijn, 1993).

[b] Letter overlap calculated as the percentage of letters shared with the target in the same position.

[c] Plausibility rating (1-5 scale) for sentence up to and including the target/preview word.

3. Results

Prior to the calculation of reading measures, all fixations below 80 ms that were within one letter space of an adjacent fixation were merged with that fixation, and remaining fixations below 80 ms or above 1,000 ms were eliminated. Trials were excluded if the participant blinked immediately before or after fixating the target word (3.17% of trials), or if the display change completed more than 10 ms into a fixation or was triggered by a saccade that ultimately landed to the left of the boundary (9.45% of trials). Target gaze durations above 1,200 ms, and go-past and total durations above 2,000 ms were also excluded (<0.5% of trials). These exclusions left 14,486 trials (87.06% of the data) available for analysis. Mean comprehension accuracy was very high (95.36%), indicating that the participants were reading for meaning.

The following reading measures were analyzed: the probability of *skipping* the target during first-pass reading; *first-fixation duration* (the duration of the first fixation on the target word regardless of the number of first-pass fixations it receives); *single-fixation duration* (the fixation duration in cases when only one first-pass fixation is made on the target word); and *gaze duration* (the sum of all first-pass fixations on the target word). We also analyzed two late measures of reading: *go-past duration* (the sum of all fixations from the first fixation on the target word until a word to the right is fixated, i.e., including fixations on the target and any subsequent fixations on words earlier in the sentence); and *total duration* (the sum of all fixations on the target word including first-pass reading and any later rereading). The probability of the reader making a *first-pass refixation* prior to leaving the target (see Reingold, Yang, & Rayner, 2010); *regressions out* of the target to words earlier in the sentence; and *regressions in* to the target from words later in the sentence were also analyzed. Means for each condition on each of these measures are presented in Table 2.

Table 2

Mean (and Standard Error) Reading Measures Aggregated by Subjects for the Target Word across Preview Conditions

Measure	Short target word				Long target word			
	HF Plausible	LF Plausible	HF Implausible	LF Implausible	HF Plausible	LF Plausible	HF Implausible	LF Implausible
Fixation Duration Measures (ms)								
First fixation	241 (3)	249 (3)	261 (3)	261 (3)	242 (3)	251 (2)	255 (3)	260 (2)
Single fixation	241 (3)	251 (3)	267 (3)	266 (3)	245 (3)	256 (3)	265 (3)	272 (3)
Gaze duration	256 (3)	270 (4)	285 (3)	286 (3)	275 (3)	290 (3)	298 (3)	303 (3)
Go-past duration	300 (5)	333 (7)	354 (5)	362 (6)	321 (5)	354 (6)	370 (6)	389 (6)
Total duration	341 (5)	364 (6)	352 (6)	366 (5)	390 (6)	405 (6)	407 (5)	417 (6)
Fixation Probability Measures (%)								
Skipping	40.26 (1.24)	37.28 (1.01)	35.09 (1.09)	33.31 (1.00)	15.58 (0.95)	11.99 (0.91)	12.57 (0.82)	10.37 (0.75)
First-pass refixation	7.02 (0.84)	9.28 (0.73)	11.81 (1.06)	11.13 (0.80)	15.27 (0.91)	18.52 (1.00)	19.78 (1.01)	20.17 (1.07)
Regressions out	12.07 (1.03)	15.49 (1.33)	18.89 (1.07)	20.93 (1.07)	10.80 (0.75)	14.19 (1.06)	16.32 (0.93)	20.20 (1.09)
Regressions in (target fixated)	18.60 (1.04)	18.50 (1.12)	12.13 (1.04)	15.95 (1.02)	24.90 (1.22)	22.30 (1.00)	20.24 (1.02)	17.48 (0.98)
Regressions in (target skipped)	33.80 (2.22)	30.91 (2.06)	37.21 (2.17)	35.86 (2.32)	44.32 (3.75)	32.92 (3.45)	57.33 (3.97)	37.95 (3.48)

The data were analyzed with (generalized) linear mixed-effects models (LMM) using the *lme4* package (Version 1.1-17; Bates, Maechler, Bolker, & Walker, 2015) in *R* (Version 3.5.0; R Core Team, 2018). In each model, sum-coded contrasts tested the fixed effects of preview frequency and plausibility, target-word length, and each of their interactions. The models included subject and item random intercepts, subject random slopes for the effects of frequency, plausibility, and length, and item random slopes for the effects of frequency and plausibility.[3] Estimates 1.96 times larger than their standard errors were interpreted as significant at the .05 alpha level. The (G)LMM estimates for coefficients, standard errors, and *t*/z values for the fixed effects are reported in Table 3. Parallel analyses conducted on log-transformed duration measures yielded an identical pattern of significant results to the analyses of the untransformed data reported below. Following the experiment, 47 participants reported some awareness of display changes; on average, these participants reported noticing changes on 14% of trials, substantially fewer than the actual proportion of trials with a display change of 86%. Excluding these participants did not change the pattern of results reported below.

3.1 Pre-boundary Word

There were no significant effects of preview frequency, plausibility, length, or interactions, on gaze duration on the pre-boundary word [all $|t|$s<1.82], i.e., there were no parafoveal-on-foveal effects, consistent with the findings of a recent Bayesian meta-analysis of such effects (Brothers, Hoversten, & Traxler, 2017).

[3] The LMMs for first-fixation, single-fixation, and go-past duration also included a subject random slope for the Frequency × Plausibility interaction but the remaining models failed to converge with the inclusion of interaction random slopes.

Table 3

Results of the (Generalized) Linear Mixed-Effects Models for Fixation Duration and Probability Measures on the Target Word. Significant Effects are Indicated in Bold.

Measure	Fixed effect	*b*	*SE*	*t*/*z*
Skipping	Intercept	**-1.45**	**0.09**	**-16.24**
	Frequency effect	**-0.19**	**0.05**	**-3.64**
	Plausibility effect	**-0.25**	**0.05**	**-4.97**
	Length effect	**-1.64**	**0.09**	**-17.71**
	Frequency × Plausibility	0.09	0.09	0.98
	Frequency × Length	-0.15	0.10	-1.51
	Plausibility × Length	-0.04	0.10	-0.42
	Frequency × Plausibility × Length	0.08	0.18	0.46
First fixation duration	Intercept	**252.76**	**3.90**	**64.75**
	Frequency effect	**4.79**	**1.97**	**2.43**
	Plausibility effect	**13.17**	**2.24**	**5.87**
	Length effect	-1.07	2.58	-0.42
	Frequency × Plausibility	-4.93	3.55	-1.39
	Frequency × Length	4.26	3.80	1.12
	Plausibility × Length	-6.15	4.19	-1.47
	Frequency × Plausibility × Length	3.01	7.02	0.43
Single fixation duration	Intercept	**258.24**	**4.13**	**62.51**
	Frequency effect	**6.05**	**2.12**	**2.86**
	Plausibility effect	**18.46**	**2.56**	**7.21**
	Length effect	3.08	2.83	1.09
	Frequency × Plausibility	-6.44	3.88	-1.66
	Frequency × Length	6.04	4.20	1.44
	Plausibility × Length	-4.40	4.79	-0.92
	Frequency × Plausibility × Length	5.47	7.59	0.72
Gaze duration	Intercept	**283.05**	**5.20**	**54.44**
	Frequency effect	**8.39**	**2.71**	**3.09**
	Plausibility effect	**20.33**	**2.56**	**7.95**
	Length effect	**17.39**	**3.56**	**4.89**
	Frequency × Plausibility	**-10.99**	**4.41**	**-2.49**
	Frequency × Length	4.36	5.38	0.81
	Plausibility × Length	-4.67	4.79	-0.98
	Frequency × Plausibility × Length	3.31	8.82	0.38
Go-past duration	Intercept	**347.95**	**8.21**	**42.38**
	Frequency effect	**22.05**	**4.69**	**4.70**

	Plausibility effect	**41.50**	**4.99**	**8.32**
	Length effect	**22.15**	**7.14**	**3.10**
	Frequency × Plausibility	**-17.74**	**7.86**	**-2.26**
	Frequency × Length	8.70	8.48	1.03
	Plausibility × Length	0.45	9.25	0.05
	Frequency × Plausibility × Length	15.87	15.50	1.02
Total duration	Intercept	**380.25**	**10.82**	**35.14**
	Frequency effect	**14.41**	**4.46**	**3.23**
	Plausibility effect	**9.88**	**3.71**	**2.66**
	Length effect	**48.93**	**8.36**	**5.86**
	Frequency × Plausibility	-5.46	7.18	-0.76
	Frequency × Length	-2.76	8.42	-0.33
	Plausibility × Length	8.77	7.20	1.22
	Frequency × Plausibility × Length	8.15	14.36	0.57
First-pass refixation	Intercept	**-2.06**	**0.09**	**-22.75**
	Frequency effect	0.05	0.06	0.78
	Plausibility effect	**0.36**	**0.06**	**5.83**
	Length effect	**0.81**	**0.08**	**10.07**
	Frequency × Plausibility	**-0.29**	**0.12**	**-2.44**
	Frequency × Length	0.06	0.12	0.50
	Plausibility × Length	-0.17	0.12	-1.44
	Frequency × Plausibility × Length	0.21	0.24	0.87
Regressions out	Intercept	**-1.87**	**0.09**	**-21.11**
	Frequency effect	**0.27**	**0.06**	**4.39**
	Plausibility effect	**0.50**	**0.07**	**7.31**
	Length effect	-0.09	0.10	-0.92
	Frequency × Plausibility	-0.07	0.11	-0.59
	Frequency × Length	0.12	0.11	1.01
	Plausibility × Length	0.01	0.14	0.08
	Frequency × Plausibility × Length	0.10	0.23	0.44
Regressions in (target fixated)	Intercept	**-1.63**	**0.08**	**-20.62**
	Frequency effect	0.01	0.06	0.13
	Plausibility effect	**-0.37**	**0.06**	**-6.03**
	Length effect	**0.36**	**0.10**	**3.54**
	Frequency × Plausibility	0.18	0.11	1.75
	Frequency × Length	**-0.31**	**0.12**	**-2.48**
	Plausibility × Length	0.11	0.11	0.93
	Frequency × Plausibility × Length	-0.33	0.21	-1.55
Regressions in (target skipped)	Intercept	**-0.46**	**0.09**	**-4.96**
	Frequency effect	**-0.28**	**0.09**	**-2.99**
	Plausibility effect	**0.28**	**0.09**	**2.97**
	Length effect	**0.50**	**0.12**	**3.99**
	Frequency × Plausibility	-0.04	0.17	-0.22
	Frequency × Length	**-0.39**	**0.18**	**-2.17**

Plausibility × Length	0.10	0.19	0.54
Frequency × Plausibility × Length	-0.11	0.35	-0.32

3.2 Target Word

3.2.1 Skipping. Readers were more likely to skip the target when the preview was a high-frequency word than when it was a low-frequency word [z=-3.64]. There was also a significant plausibility effect on skipping: readers were more likely to skip a plausible than an implausible parafoveal word [z=-4.97]. Interestingly, frequency and plausibility yielded similar sized effects on skipping: 2.67 vs. 3.34%, respectively. Target-word length yielded a substantially larger effect on skipping because readers were more likely to skip short than long words [23.81%; z=-17.71]. There were no significant interactions between frequency, plausibility, and length [all $|z|$s<1.51]. Follow-up analyses, conducted separately for short and long words, confirmed significant effects for both target lengths of frequency [Short: b=-0.12, SE=.05, z=-2.43; Long: b=-0.27, SE=.10, z=-2.54] and plausibility [Short: b=-0.22, SE=.05, z=-4.32; Long: b=-0.31, SE=.10, z=-3.10].

3.2.2 Fixation measures. Across all measures of reading time, there was a significant effect of preview frequency because readers spent less time fixating the target word when the preview was a high-frequency word than when it was a low-frequency word [all $|t|$s>2.43]. Readers were also less likely to make a regression from the target when the preview was high frequency [z=4.39], but preview frequency did not affect first-pass refixations on the target, or regressions in, for trials on which the target was fixated during first-pass reading [both $|z|$s<1].

There was a significant effect of preview plausibility on all measures of fixation time because readers spent less time fixating the target following a preview of a plausible word compared to an implausible word [all $|t|$s>2.66]. Readers were also less likely to regress from the target and less likely to refixate the target following a plausible preview [both $|z|$s>5.83].

While the size of the plausibility effect was larger than the frequency effect on first-pass reading measures (e.g., 20 vs. 8 ms on gaze duration), it was smaller on total duration (10 vs. 14 ms) because regressions in to the target word were more likely following a *plausible* preview compared to an implausible preview for trials on which the target was fixated during first-pass reading [z=-6.03]. In contrast, for trials on which the reader skipped the target, regressions in were more likely following an implausible preview [z=2.97].

There was a significant effect of target word length on gaze, go-past, and total duration [all |t|s>3.10] but not on first- or single-fixation duration [both |t|s<1.09]. Target length did not affect regressions out [|z|<1]. The fact that length did not affect the duration of the first fixation on the target word suggests that the later effects were specifically due to refixations and regressive fixations. This was confirmed by the significant effects of length on first-pass refixations and regressions in, both when the target was fixated and when it was skipped during first-pass reading: readers were less likely to refixate or regress to short words compared to long words [all |z|s>3.54]. However, the length effect on regressions in was qualified by a significant interaction with frequency [both |z|s>2.17]. Follow-up GLMMs revealed that the length effect was restricted to high-frequency previews [target fixated in first-pass: b=0.48, SE=0.12, z=3.95; target skipped in first-pass: b=0.73, SE=0.14, z=5.17]; word length did not significantly affect regressions in following low-frequency previews [target fixated: b=0.20, SE=0.11, z=1.81; target skipped: b=0.30, SE=0.15, z=1.93].

Finally, on gaze duration and go-past duration there was a significant interaction between frequency and plausibility [both |t|s>2.26]. Follow-up analyses were conducted separately for the plausible and implausible conditions. For plausible previews, there was a significant effect of frequency on both gaze duration [b=13.82, SE=3.29, t=4.20] and go-past duration [b=26.84, SE=6.69, t=4.01]. In contrast, for implausible previews, there was no effect of frequency on gaze duration [b=3.15, SE=3.09, t=1.02] and the frequency effect was

reduced, but still significant, on go-past duration [b=16.58, SE=6.60, t=2.51]. The Frequency × Plausibility interaction was not significant on either first-fixation or single-fixation duration [both $|t|$s>1.66] suggesting that it was specifically due to refixations on the target word. This was confirmed by the significant interaction between frequency and plausibility on first-pass refixations [z=-2.44]. Follow-up GLMMs showed that, for plausible previews, readers were significantly less likely to refixate the target following high-frequency than low-frequency previews [b=0.24, SE=0.09, z=2.54]. In contrast, for implausible previews, there was no effect of frequency on refixation likelihood [z<1].

3.3 Bayesian Analyses

The most critical findings of the present study are null effects: the absence of significant interactions between word length and either preview frequency or plausibility on skipping. To address the limitations of relying on Null Hypothesis Significance Testing we also conducted Bayesian analyses to directly assess the strength of evidence for the null interaction. Bayesian LMMs were conducted on target-word skipping and gaze duration in order to assess the relative likelihood of additive versus interactive effects of frequency, plausibility, and length. The BayesFactor package (version 0.9.12-2; Morey & Rouder, 2015) was used to compute Bayes factors for each of the nested linear models compared to a random-effects–only model. Because logistic regression has not yet been implemented in the BayesFactor package, the skipping data were analyzed by separate ANOVA models treating subjects and items as random effects. The analyses used the default scale value (0.5) for the Cauchy priors on effect size and 100,000 Monte Carlo iterations. Bayes factors for each nested model were computed relative to a random-effects–only model and models were then compared by taking the ratio of their Bayes factors. A Bayes factor greater than 1 indicates evidence in favor of the more complex model and a Bayes factor of less than 1 indicates evidence in favor of the simpler model.

3.3.1 Skipping. Relative to random-effects–only models, there was extreme evidence in favor of models including the effects of frequency, plausibility, and length on skipping: BF_s=4.37 × 10^{151}; BF_i=7.87 × 10^{45}. The critical test of whether the models including the interactive terms were favored over the models including only the additive effects yielded evidence favoring the additive models: Frequency × Plausibility (BF_s=0.17; BF_i=0.17); Frequency × Length (BF_s=0.11; BF_i=0.12); and Plausibility × Length (BF_s=0.40; BF_i=0.38). That is, the Bayesian analysis confirmed the null interaction effects on skipping.

3.3.2 Gaze duration. Relative to a random-effects–only model, the evidence extremely favored a model including the effects of frequency, plausibility, and length on target gaze duration: BF=4.92 × 10^7. Confirming the LMM analysis, the model including the Frequency × Plausibility interaction was favored over the additive model (BF=20.06), while the additive model was favored over models including the Frequency × Length interaction (BF=0.13) or the Plausibility × Length interaction (BF=0.06).

3.4 Supplementary Analyses

3.4.1 Saccade launch distance. Supplementary analyses were conducted to investigate whether the effects of parafoveal frequency, plausibility, and length depended on the distance of the prior fixation. Longer words extend further into the parafovea and therefore may not be processed as efficiently as short words due to the reduced visual acuity and the greater impact of crowding at increased eccentricity. The absence of interactions between word length and either frequency or plausibility in the main analyses may therefore actually be due to systematic differences in saccadic launch distances for short versus long words.

An LMM analysis of saccadic launch distance revealed no significant main effects or interactions of frequency, plausibility, or length [all $|t|$s<1.34], suggesting that parafoveal information did not significantly affect fixation locations prior to crossing the boundary,

consistent with the lack of parafoveal-on-foveal effects on reading time of the pre-target word.

To explore whether launch distance affected the skipping effects, we specified separate GLMMs on the target-skipping data based on whether the pre-target word was fixated or skipped prior to the reader's eyes crossing the boundary. For trials on which the pre-target word was fixated, there were significant effects of frequency [b=-0.24, SE=0.06, z=-4.09], plausibility [b=-0.26, SE=0.06, z=-4.66], and length [b=-1.87, SE=0.11, z=-17.30] on target skipping. Importantly, as was observed in the main analysis, none of the interactions were significant [all $|z|$s<1.67]. In contrast, when the pre-target word was skipped, the only significant effect on target skipping was word length [b=-0.61, SE=0.17, z=-3.56]. Thus, evidence of lexical and contextual processing on word skipping, indexed by frequency and plausibility effects, respectively, was restricted to cases in which the reader was fixated on the pre-target word, i.e., at a close launch site. When readers fixated further to the left, target skipping was only affected by length, presumably reflecting the impact of low-level visual processing and/or saccadic error.

3.4.2 Restricted subset. The main analyses revealed a discrepancy between the effects of frequency and plausibility on initial target fixations (first-fixation and single-fixation duration) and later measures (gaze and go-past duration). Specifically, the former measures showed additive effects of the two factors while the latter measures showed interactive effects. This discrepancy may reflect the impact of differences in the likelihood of regressions out of the target word based on the frequency and plausibility of the preview word (see also Veldre & Andrews, 2018a). Immediate regressions out of the target prematurely terminate an initial fixation and it is therefore possible that early effects of these variables on fixation durations were obscured by the inclusion of trials on which the reader made a regression after crossing the boundary (see e.g., Veldre & Andrews, 2016). We

therefore ran a supplementary analysis of target word first-pass reading measures (first-fixation, single-fixation, and gaze duration) on the subset of trials on which readers fixated the pre-target word immediately prior to fixating the target word and made no regression from the target word (6,673 trials; 65.97% of trials with first-pass fixation on the target word).

Consistent with the unrestricted analysis reported above, this supplementary analysis showed significant effects of preview frequency [all *t*s>2.45] and preview plausibility [all *t*s>6.29] on all three measures. Also paralleling the unrestricted analyses, the effect of target length was not significant on first-fixation duration [$t<1$] but target length significantly affected both single-fixation duration [$b=7.84$, $SE=3.75$, $t=2.09$] and gaze duration [$b=21.26$, $SE=4.75$, $t=4.48$].

However, in contrast with the main analysis, the Frequency × Plausibility interaction was significant on all first-pass measures in the restricted subset analysis: first-fixation duration [$b=-10.85$, $SE=4.56$, $t=-2.38$], single-fixation duration [$b=-13.50$, $SE=4.86$, $t=-2.78$], and gaze duration [$b=-18.41$, $SE=5.69$, $t=-3.24$]. Follow-up comparisons showed a significant frequency effect for plausible previews on all measures [all *t*s>3.03] but no effect of frequency for implausible previews [all *t*s<1]. Finally, there was a significant Frequency × Length interaction on single-fixation duration [$b=11.91$, $SE=5.91$, $t=2.01$]. Follow-up comparisons revealed a significant preview frequency effect for long targets [$b=13.06$, $SE=3.74$, $t=3.49$] but not for short targets [$t<1$].

3.4.3 Quantile analysis. Schotter and Leinenger (2016) provided an account of preview difficulty effects based on the concept of *forced fixations*. They argued that, not only are easy-to-process parafoveal words more likely to be skipped, but that there will also be a subset of trials on which an easy parafoveal word will be identified too late in the saccadic programming window to yield a skip. In these cases, a saccade from the target to the post-

target word will be planned on the basis of the preview information and an intervening short fixation on the target will be 'forced'. This account predicts that the effect of preview difficulty should be concentrated in short fixations because longer fixations will be more likely to reflect properties of the target word extracted after it is fixated. Schotter and Leinenger confirmed both of these predictions with a quantile regression: preview frequency affected earlier quantiles while target frequency only affected later quantiles of single fixation duration on the target.

To compare the time course of parafoveal frequency and plausibility effects in the present dataset we conducted a quantile regression analysis on single-fixation duration on the target word using the *quantreg* package (Koenker, 2018) in *R*. The quantile regression was based on the 7,737 trials (63.41%) on which the reader made a single fixation and no regression out of the target. In order to test the effects across the entire time course of processing, the analysis was conducted on the full range of values for single-fixation duration, prior to the merging and deleting of short and long fixations.[4] The analysis tested for the effect of target length, preview frequency, preview plausibility, and the Frequency × Plausibility interaction at each quantile (see Table 4). The results showed that the effects of target length and preview frequency were significant only in the earlier quantiles until the .49 quantile (all *p*s<.003). In contrast, preview plausibility was significant from the earliest quantile until the .89 quantile (all *p*s<.001). Finally, the Frequency × Plausibility interaction was only significant in the .69 and .79 quantiles (both *p*s<.05).

These findings converge with the skipping data which showed that parafoveal frequency and plausibility effects emerged early enough to influence the decision to skip the preview word before the reader fixated the target word. For trials on which target word *was*

[4] LMM analyses conducted on these data yielded an identical pattern of results to those conducted on the trimmed data.

fixated, frequency and plausibility independently affected the fixation-duration distribution from the shortest fixations. Preview plausibility had a prolonged impact across almost the full distribution of fixation durations. However, in later quantiles, the two factors interacted because the frequency effect disappeared for implausible previews while, for plausible previews, the frequency effect persisted across almost the entire distribution of fixation durations on the target. The effects of preview frequency and length on the early quantiles of the fixation distribution are potentially compatible with Schotter and Leinenger's (2016) *forced fixation* account of preview effects, but the extended impact of preview plausibility and the plausibility by frequency interaction is more difficult to reconcile with a single mechanism. We will return to these issues in the Discussion.

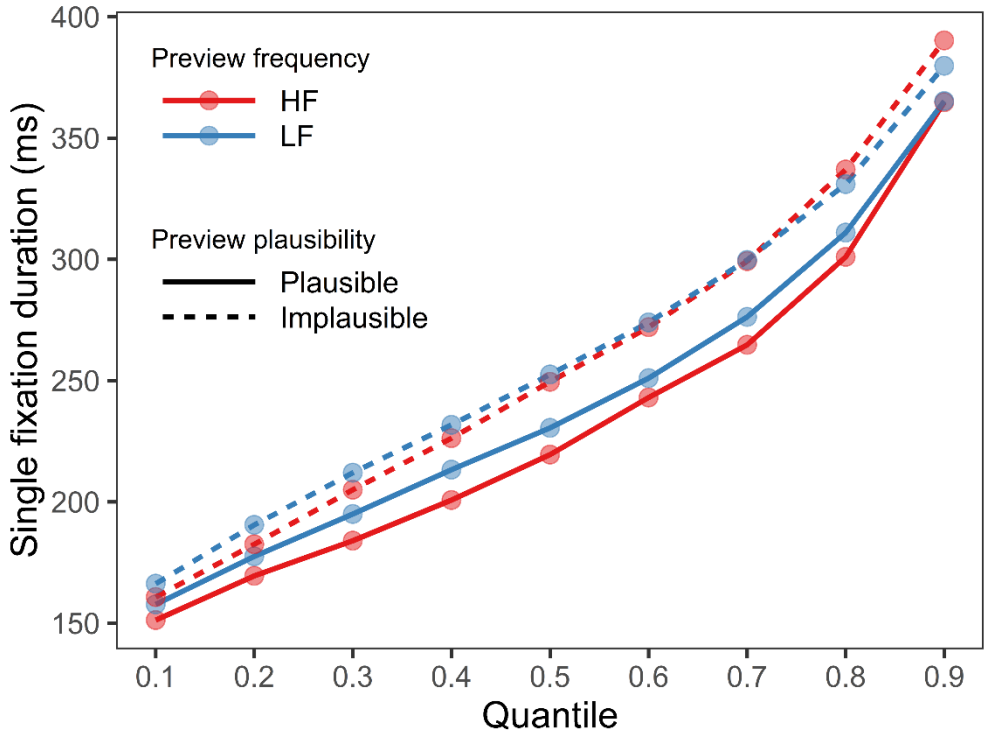

Figure 3. Single-fixation duration on the target by preview type across quantiles.

Table 4

Quantile regression analysis for the effects of target length, preview frequency, preview plausibility, and the Frequency × Plausibility interaction at each of 10 quantiles of single fixation duration on the target word. Significant effects are indicated in bold.

Quantile	Intercept			Target length			Preview frequency			Preview plausibility			Frequency × Plausibility		
	b	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*
0-.09	**159**	**189.48**	**<.001**	**9**	**5.50**	**<.001**	**6**	**3.29**	**.001**	**9**	**5.09**	**<.001**	-1	-0.30	.765
.1-.19	**180**	**199.85**	**<.001**	**10**	**5.80**	**<.001**	**8**	**4.47**	**<.001**	**13**	**7.26**	**<.001**	0	0.00	1.000
.2-.29	**199**	**195.09**	**<.001**	**10**	**5.09**	**<.001**	**9**	**4.41**	**<.001**	**19**	**9.31**	**<.001**	-4	-0.98	.327
.3-.39	**218**	**199.16**	**<.001**	**11**	**5.13**	**<.001**	**9**	**3.91**	**<.001**	**22**	**9.90**	**<.001**	-7	-1.61	.107
.4-.49	**238**	**195.01**	**<.001**	**10**	**4.16**	**<.001**	**7**	**3.01**	**.003**	**26**	**11.17**	**<.001**	-8	-1.72	.086
.5-.59	**260**	**187.38**	**<.001**	4	1.45	.146	5	1.86	.064	**26**	**9.65**	**<.001**	-6	-1.11	.266
.6-.69	**285**	**202.49**	**<.001**	1	0.36	.720	6	1.96	.050	**29**	**10.20**	**<.001**	**-11**	**-1.97**	**.049**
.7-.79	**320**	**157.22**	**<.001**	-5	-1.24	.216	2	0.50	.616	**28**	**7.02**	**<.001**	**-16**	**-2.00**	**.045**
.8-.89	**375**	**121.77**	**<.001**	-9	-1.47	.141	-5	-0.75	.456	**20**	**3.23**	**.001**	-11	-0.91	.362
.9-.99	**615**	**31.76**	**<.001**	-35	-0.99	.324	9	0.22	.823	17	0.43	.665	-53	-0.70	.486

3.5 E-Z Reader Simulations

The empirical data provide evidence that whatever type of higher-level linguistic information is required to evaluate a parafoveal word's within-sentence plausibility can be acquired rapidly enough to—in some instances—influence the decision about whether to fixate or skip the word. Such findings have been assumed (e.g., Abbott & Staub, 2015; Veldre & Andrews, 2018a) to be incompatible with the post-lexical integration account of plausibility effects implemented in the E-Z Reader model of eye-movement control in reading (Reichle et al., 1998, 2012). As previously discussed, the model's assumption that words are identified in a strictly serial manner imposes severe timing constraints on both lexical and post-lexical processing.

To account for the present evidence of preview plausibility effects on skipping within the framework of E-Z Reader, the sequence shown schematically in Figure 3 would have to occur. As shown, the completion of lexical processing of the pre-target word (i.e., word$_N$) causes attention to shift to the target word (word$_{N+1}$) and initiates a labile saccadic program, M_1, to move the eyes to word$_{N+1}$. However, the subsequent completion of the first stage of lexical processing, L_1, for word$_{N+1}$ cancels the saccade that would have otherwise caused the eyes to (barring saccadic error) move to word$_{N+1}$ (labelled (1) in Figure 3). At this point, the second stage of lexical processing, L_2, completes for word$_{N+1}$, initiating both post-lexical integration of word$_{N+1}$, and the programming of a saccade to move the eyes to word$_{N+2}$. If integration completes, then the most likely sequence of events will be the completion of the saccadic program to move the eyes to word$_{N+2}$, causing word$_{N+1}$ to be skipped. However, if integration of word$_{N+1}$ fails, then this interrupts the saccadic program (labelled (2) in Figure 3) and results in attention and the eyes being directed towards the source of integration difficulty, word$_{N+1}$, causing that word to be fixated.

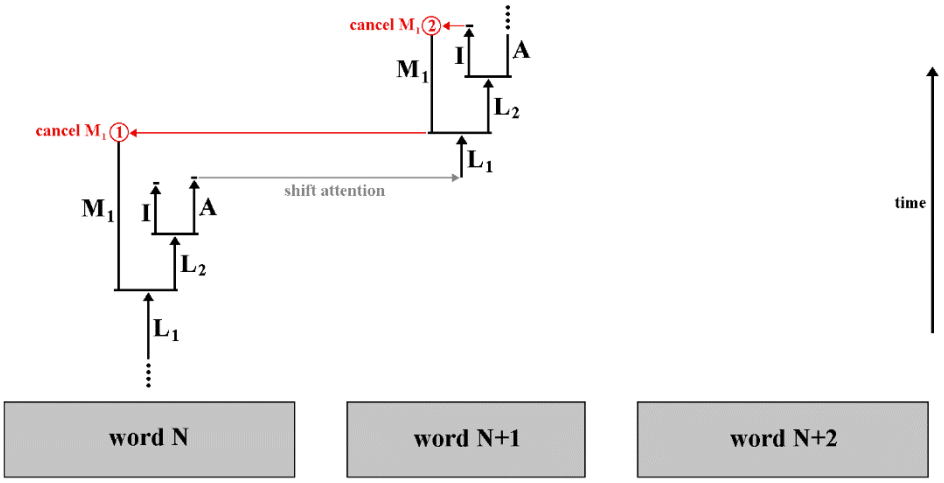

Figure 3. Time course of critical events that allow the plausibility of word$_{N+1}$ to influence whether or not it will be skipped. The gray boxes denote the locations of word$_N$, word$_{N+1}$, and word$_{N+2}$, with the time course of the completion of the key processes progressing upwards. (*L1* = familiarity check; *L2* = lexical access; *M1* = labile saccadic programming; *A* = attention shift; and *I* = post-lexical integration.)

Previous simulations using E-Z Reader have shown that its serial-attention architecture affords significant parafoveal processing (e.g., see Schotter, Reichle, & Rayner, 2014). However, it is not obvious that this affordance is sufficient to allow integration failure to modulate skipping rates a reasonable proportion of the time to accommodate the present evidence of plausibility effects on skipping of both short and long words. To directly investigate this issue, we conducted simulations using parameters derived from the present stimulus materials.

Simulations of E-Z Reader were completed using 1,000 statistical subjects per condition and parameter values that had previously been selected to maximize the model's goodness-of-fit of various fixation duration and probability measures of eye-movement data from the Schilling et al. (1998) sentence corpus. Following the methods used in previous E-Z Reader simulations (e.g., Schotter et al., 2014), these sentences were used as "frames" to evaluate how properties of our pre-target words and the target-word previews influenced the

probability of skipping the target words and three fixation-duration measures on the target words (first-fixation duration, gaze duration, and total duration). To do this, the lengths and frequencies of the pre-target words were set equal to the mean values used in the present study (i.e., length = 6 letters; frequency = 591 per million), as were the cloze probabilities of the target-word previews (i.e., cloze = 0). The length, frequency, and plausibility of the target-word previews were then manipulated (see Table 1).[5] Because it is not possible to know the correspondences between the values of the parameter that modulates the probability of post-lexical integration failure of the target-word previews, p_F, and the normative plausibility ratings of our actual stimuli, we compared five values that spanned the domain of possible values: p_F = 0, 0.25, 0.5, 0.75, and 1. The value of *I*, the time required to complete post-lexical integration, was not manipulated because the question of interest is whether or not post-lexical integration of the preview word completes rapidly enough (i.e., using the default value of *I*) to influence skipping of the preview word.

The simulation results are shown in Table 5, as a function of the properties of the target-word previews, with manipulations of preview plausibility being simulated using different values of the p_F parameter. As shown, all three properties of the target-word preview had marked effects on the simulated fixation-duration and skipping rate measures. As expected, target words were more likely to be the recipients of fewer and/or shorter fixations if they were high frequency and/or short in length. For example, the mean gaze durations were shorter for high- as compared to low-frequency previews (*M* = 305.2 vs. 317.9 ms, respectively), as well as for short as compared to long previews (*M* = 293.4 vs. 329.7 ms, respectively). Both of these findings are consistent with our results and prior simulations using E-Z Reader (e.g., see Reichle, 2011). Our simulations also clearly show, however, that

[5] The discrepancy between the frequency values in Table 1 and those used in the simulations is because word frequency in E-Z Reader is taken from the Kučera and Francis (1967) norms whereas the stimulus materials were matched on CELEX frequency.

target-word preview *can* rapidly influence even the early decisions about whether or not to fixate a word, as evidenced by decreased skipping rates for implausible target-word previews. For example, compared to completely plausible previews (i.e., $p_F = 0$), even the modestly implausible previews (i.e., $p_F = 0.25$) were skipped less often ($M = 0.173$ vs. 0.163, respectively). This result demonstrates that, despite the non-trivial constraints imposed by the times required to both identify words and program saccades (e.g., see Reichle & Reingold, 2013), the amount of parafoveal processing that can be completed is, on average, sufficient to reach the integration stage so that contextual plausibility can then influence decisions about whether to fixate or skip the word. This demonstration therefore shows that even the severe constraints imposed by the serial allocation of attention and a staged architecture can nonetheless accommodate our findings that parafoveal word plausibility can rapidly modulate even the earliest markers of lexical processing—decisions to skip an upcoming word.

4. Discussion

The present study was motivated by discrepancies in the literature concerning the relative contribution of word frequency and plausibility to skipping effects for words of different lengths. The study was designed to directly compare the effects of parafoveal frequency and plausibility for short and long target words to fill important gaps in the existing evidence. The results showed robust, additive effects of the frequency, plausibility, and length of the parafoveal word on skipping rates and target fixation duration. Frequency and plausibility also yielded significant interactions on late eye-movement measures because there was no frequency effect for implausible previews. The quantile analysis of single-fixation trials confirmed that this interaction emerged late in the time course of target word processing.

Table 5

E-Z Reader simulations using five target-word plausibility values (i.e., values of p_F).

Target Word Characteristics		DVs	Target-Word Plausibility (p_F Values)				
Length	Frequency		0	0.25	0.5	0.75	1
Short = 3 & 4 Letters	High Frequency = 197 per Million	PrS	0.23	0.22	0.21	0.20	0.19
		FFD	237	239	240	239	238
		GD	245	268	289	310	330
		TD	249	275	299	321	343
	Low Frequency = 5 per Million	PrS	0.22	0.22	0.21	0.19	0.19
		FFD	245	246	246	246	245
		GD	254	278	300	321	343
		TD	258	285	310	334	356
Long = 6 Letters	High Frequency = 230 per Million	PrS	0.12	0.11	0.09	0.08	0.07
		FFD	240	242	242	241	240
		GD	268	296	323	350	375
		TD	273	305	336	366	395
	Low Frequency = 4 per Million	PrS	0.12	0.11	0.09	0.08	0.07
		FFD	248	250	249	249	247
		GD	279	309	338	366	393
		TD	284	318	350	382	414

Note: *PrS* = skipping probability; *FFD* = first-fixation duration; *GD* = gaze duration; and *TD* = total duration.

The results of the present study have several important theoretical implications. Firstly, these data challenge the conclusions of Angele and colleagues (2014) who argued against the idea "that syntactic and semantic information about the upcoming word and its fit with the sentence affect the timing of the skipping decision" (p. 1190). They found that readers were more likely to skip three-letter words when the parafoveal preview was high frequency compared to when it was low frequency, regardless of the preview's contextual acceptability in the sentence. Similarly, Abbott and Staub (2015) found that frequency, but not plausibility, affected skipping of 5-8–letter target words embedded in neutral sentence contexts. The current study was designed to address methodological limitations of these two studies. Abbott and Staub's data showed low overall skipping rates that may have been due to

a cautious reading strategy caused by the foveal presentation of anomalous words. The use of the boundary paradigm in the current study allowed for the manipulation of parafoveal frequency and plausibility without readers' awareness. The sentence frames were also carefully constructed so that each of the previews were compared within the same unconstraining context, and none of the target words or previews were contextually predictable, addressing a limitation of Angele et al.'s materials. In contrast to the results of Abbott and Staub and Angele et al., the current study found that frequency and plausibility yielded similarly sized, additive effects on skipping rates.

The data also showed that word length yielded a strong effect on skipping because short words were more likely to be skipped than long words, but there was little evidence that length modulated the effects of frequency or plausibility. The length effects converge with previous findings that word length does not modulate predictability effects on skipping (Drieghe, Brysbaert, Desmet, & De Baecke, 2004; Rayner, Slattery, Drieghe, & Liversedge, 2013), which have been interpreted as evidence that visual and linguistic information exert independent influences on parafoveal processing. The only evidence of a combined impact of length and frequency on skipping in the present data came from the analysis of regressions in following skips. When readers skipped long words, they were more likely to regress back to high-frequency words than low-frequency words, whereas frequency did not affect regressions in for short words. Thus, word length appears to primarily impact the low-level oculomotor planning processes that are reflected in fixation location, rather than directly modulating the accumulation of lexical information from parafoveal words, at least for the range of word lengths used in this study.

The fixation duration results replicate previous evidence of preview difficulty effects on first-pass reading of the target word (Risse & Kliegl, 2012, 2014; Schotter & Leinenger, 2016; Schotter et al., 2018): Readers made shorter fixations on the target when the preview

was a high-frequency word compared to a low-frequency word. Replicating previous demonstrations of the plausibility-preview effect, fixation durations were also shorter when the preview was a plausible continuation of the sentence than when it was implausible (Schotter & Jia, 2016; Veldre & Andrews, 2016, 2017, 2018a, 2018b, 2018c; Yang et al., 2012, 2014). These robust effects of preview frequency and plausibility were observed from the duration of the first fixation on the target word and the quantile analysis showed that these factors both affected the shortest fixation durations. In contrast, apart from the robust main effect on skipping, word length primarily influenced the likelihood of refixating and regressing to the target, resulting in inflated gaze, go-past, and total durations for longer words. Such effects are likely to be due, at least in part, to the differential impact of oculomotor errors on short versus long words. Specifically, readers are more likely to overshoot a short target, resulting in a skip, whereas, for a long target, saccadic error is more likely to result in a non-optimal fixation location, requiring a refixation. Like the skipping data, the effects of word length were found to be independent of preview frequency and plausibility, with the Bayesian analysis confirming equivalent effects for short and long words.

In contrast to the independent effects of frequency and plausibility on skipping, a significant Frequency × Plausibility interaction was observed on gaze and go-past duration in the main analyses, and was evident on all first-pass reading measures for trials with no regressions out of the target word. In all cases, the interaction reflected the absence of a frequency effect for implausible word previews. This may be because high-frequency words are more likely to be processed sufficiently deeply to trigger post-lexical processing during the labile stage of saccadic planning. This would result in a greater proportion of trials on which post-lexical integration failure occurred early enough to affect first-pass reading measures for high-frequency relative to low-frequency previews, leading to increased average

fixation durations following high-frequency implausible previews. While this interactive pattern of effects is intriguing, we acknowledge that it was not observed on all measures in the unrestricted analysis and should therefore be interpreted with some caution (von der Malsburg & Angele, 2017).[6]

If future research confirms that the interactive effects of frequency and plausibility observed in the present study are robust, it will be important to reconcile them with the results of studies that have jointly manipulated frequency and *predictability.* The latter studies have overwhelmingly observed additive effects of the two factors on fixation durations (see Staub, 2015, for a review).[7] A possible reason for the inconsistency is that, in contrast to previous research, the present study manipulated the frequency and plausibility of the parafoveal *preview word*, and the interaction was observed on fixation duration on an orthographically and semantically unrelated *target word*. Thus, the duration of the fixation on the target word in the present study was affected by multiple influences: the properties of the preview word, the display change from the preview to the target word, and the properties of the target word.

Most critically, the present study failed to replicate previous findings that frequency affects both skipping and fixation duration while plausibility selectively affects fixation duration. Such evidence has been argued to support a "staged" reading architecture (Staub, 2011) in which a word's plausibility only affects the ease of post-lexical integration, through a mechanism like that incorporated in E-Z Reader 10 (Reichle et al., 2009). This post-lexical account of plausibility effects has been assumed to predict little, if any, influence of

[6] Abbott and Staub (2015) found that frequency and plausibility yielded additive effects on gaze duration in the by-subjects ANOVA and an interaction in the by-items analysis. Supplementary Bayesian analyses favored the additive model in the by-subjects analysis but provided only equivocal evidence in favor of the additive model in the by-items analysis. Thus, further research is necessary to provide conclusive evidence in favor or against an interaction of frequency and plausibility on fixation duration.

[7] The pattern is less consistent for the influence of frequency and predictability on skipping: some studies have observed additive contributions of frequency and predictability, while others have found interactive effects but in opposite directions (see Staub, 2015).

plausibility on skipping (Abbott & Staub, 2015). As can be deduced from Figure 1, for skipping rates to be reduced for implausible compared to plausible words that are matched on frequency, length, and cloze probability, E-Z Reader 10 requires that both lexical access (L_2) and post-lexical integration (I) must complete during labile saccadic planning so that integration failure can cancel the skipping saccade. We have previously assumed that this sequence of events occurs too rarely to yield the robust effects of plausibility on skipping observed in previous studies (Veldre & Andrews, 2017, 2018a, 2018b), and replicated in the present data.

Contrary to this intuition, the simulations of E-Z Reader conducted using the present stimuli and the model's default parameters revealed that the model successfully captured the broad features of the skipping data: the probability of skipping an implausible relative to a plausible preview word was reduced to a similar degree to the 3.4% average reduction observed in the empirical data. The simulation results are also broadly consistent with the empirical evidence of statistically equivalent effects of parafoveal plausibility on the skipping rates for 3-4 letter words and 6 letter words and converges with previous evidence of plausibility effects on skipping of longer words (e.g., Brothers & Traxler, 2016; Matsuki et al., 2011; Staub, 2011; Veldre & Andrews, 2017, 2018a, 2018b). Thus, despite the timing constraints on lexical and post-lexical processing due to the E-Z Reader's staged, sequential architecture, the simulation results obtained from the model are more consistent with the present data than with claims that any effect of preview plausibility on skipping should be limited to short words (Abbott & Staub, 2015).

It should be noted that other aspects of the simulation results were less consistent with our experimental findings. While the E-Z Reader estimates of the skipping rates for long words were similar to the empirical data, the model dramatically underestimated the observed skipping rates of short words (20.8% vs 35.6%). Furthermore, in contrast to the

approximately 3% effect of frequency on skipping in the empirical data, the E-Z Reader simulations showed very little impact of word frequency on skipping for short or long words. Such differences may reflect the fact that E-Z Reader's free parameters were "tuned" to maximize the fit of the Schilling et al. corpus data and no adjustments were made to the parameter values to fit the present data (see Table 1 of McGowan & Reichle, 2018). [8] Differences between the average skill level of the two samples, sentence materials, and/or the comprehension demands of the reading tasks may have influenced the overall skipping rates and average reading speeds in the two datasets. It is also important to recognize that, because the model does not directly simulate the boundary paradigm, the simulated fixation-duration measures are not directly comparable to the empirical data: they only reflect the influence of properties of the preview words and not the contribution of either the display change or the properties of the target words. Importantly, this issue does not affect the interpretability of the simulated skipping data because skipping occurs prior to the display change. Nevertheless, the simulated fixation-duration measures are useful to contextualize the skipping rates and to demonstrate that they are of a reasonable duration and show effects of word frequency and length, despite the simulated plausibility manipulation. Most critically, the outcome of these

[8] The model's underestimation of the skipping rates for our short targets is likely due to its assumptions about saccadic error. Simulations (not reported in this article) that disabled both the systematic and random error components of saccade execution showed a greatly reduced effect of word length on skipping, suggesting that most of the variance is due to saccadic error rather than modulation of $t(L_1)$ and the cancelling of one labile saccadic program via the initiation of another.
Additional simulations indicate that the limited impact of frequency on the simulated skipping rates of our target words is likely due to the model's assumptions regarding the duration of L_1 and the current parameter values. Specifically, the model is likely assigning too much of the skipping variance to the assumption that word$_N$ can be guessed, setting $t(L_1)$ equal to 0 ms, with a probability equal to the cloze probability of word$_N$. Consistent with this assessment, simulations using the default model parameter values that disabled the "word guessing" mechanism showed an attenuated effect of frequency on skipping across the full range of word-frequency values. However, simulations using the best-fitting parameter values of this non-word-guessing version of E-Z Reader demonstrated that the model *did* produce the expected effect of word frequency on skipping and was sufficient to account for the key benchmark findings related to eye-movement control in reading. Importantly, this version of the model also replicated the critical finding of the simulations reported in this paper—the probability of skipping was reduced with increasing probability of post-lexical integration failure, p_F.

simulations provides an "existence proof" that post-lexical integration failure could occur early enough to affect skipping in a staged, serial architecture.

Importantly, however, although E-Z Reader makes assumptions about the duration of post-lexical integration (I) and the likelihood that such integration fails (p_F), the model does not specify the information or processes that allow readers to detect, and react to, post-lexical integration failure so quickly. For such failures to occur early enough to affect skipping behavior, readers must be able to extract sufficient semantic and syntactic information from parafoveal words to determine whether or not they are contextually acceptable in the developing sentence or discourse representation. E-Z Reader assumes that the speed of both L_1 and L_2 stages of lexical processing is a function of the frequency and cloze probability of the target word, but that cloze probability only affects the duration of lexical processing of word$_N$ if post-lexical integration for word$_{N-1}$ has completed. This assumption implies that words are incrementally integrated into the developing sentence representation process, affording the potential for contextual predictability to rapidly contribute to the ease of identifying upcoming words. However, on average, the plausible previews used in the present study were very low in cloze probability—although rated as highly plausible, they were virtually never produced as likely continuations of the sentence. Thus, if cloze probability was the only source of contextual influence on lexical retrieval, all of the present previews would be equally likely to yield post-lexical integration failure because none of them are highly predictable continuations of the sentence.[9]

[9] Although the mean cloze probability of the items in the present study was very low, some individual preview words were moderately predictable. In order to confirm that the results were not due to small differences in cloze probability between the conditions, we analyzed the skipping data for the subset of items with exactly zero cloze probability across all preview conditions (104 of 160 items; 9410 data points). The results of this analysis were identical to the main analyses: significant effects of length, frequency, and plausibility (all $|z|$s>2.50), and no significant interactions (all $|z|$s<1). Thus, the critical results of the present study cannot be explained by the effect of small differences in cloze probability between the plausible and implausible conditions.

Reichle et al. (2009) acknowledged the limitations of E-Z Reader's implementation of post-lexical integration. This stage was added to the model to demonstrate that its capacity to simulate the influence of higher level language processes on the pattern of eye movements during reading could be enhanced by a "few simple assumptions", but the mechanism was acknowledged to be a "placeholder" for a more detailed computational account of sentence processing that specifies the knowledge and processes that contribute to post-lexical integration failure. The simulations show that *if* the information required to detect an integration failure was available following 25 ms of post-lexical processing for a given word, then post-lexical integration failure could occur early enough to influence skipping. However, they beg the question of *how* the contextual plausibility of words that are low in cloze probability is computed quickly enough to trigger post-lexical integration failure.

Readers' sensitivity to contextual plausibility implies that they may compute the probability of multiple possible continuations rather than generating a single prediction. Evidence indicating that readers' response to post-lexical integration failure yields qualitatively different forms of reanalysis (e.g., refixation vs regressions) according to whether an implausible preview is a semantic or a syntactic violation (e.g., Veldre & Andrews, 2018a) further suggests that these two dimensions of contextual plausibility make independent contributions to post-lexical integration. Thus, readers may make use of all available sources of information—semantic and syntactic constraints in the sentence in combination with their language experience and general world knowledge—to narrow the pool of potential lexical candidates and activate a set of possible continuations for the sentence given the prior context (e.g., Levy, 2008; Luke & Christianson, 2016). While many of these items may be unpredictable in the sense that they are produced by few, if any, individual participants in the cloze task (Staub, Grant, Astheimer & Cohen, 2015), previewing a word from this set may confer a processing benefit that facilitates lexical

processing and/or post-lexical integration quickly enough to increase skipping rates relative to a word that does not fall into this set. Such a view is not inherently incompatible with E-Z Reader. Reichle et al. (2009) acknowledge that the use of cloze probability to operationalize 'predictability' is motivated by sustaining comparability with previous data, rather than a conceptual commitment to this metric. However, recognizing that the effects of predictability and plausibility may reflect multiple, potentially independent probabilistic influences rather than the single index of the success of integration implemented in E-Z Reader 10 raises the possibility that, rather than being entirely due to costs associated with post-lexical integration failure, plausibility effects may also reflect facilitated processing of contextually plausible words.

This account of plausibility preview effects suggests that words that are semantically and/or syntactically compatible with the preceding context will be easier to process parafoveally than those that are not. Schotter and Leinenger (2016) argued that preview benefits can arise when the signal to skip an easy to process parafoveal word (i.e., completion of L_1) is triggered too late to cancel the planned saccade to that word. In these cases, the reader initiates a parallel saccade program away from the parafoveal word resulting in a brief "forced" fixation on the target word that is not affected by trans-saccadic integration failure. This framework can explain preview effects of word frequency because high-frequency parafoveal words are processed more easily, and therefore more quickly, and so yield more skips and forced fixations than low-frequency words. This "preview difficulty effect" (Risse & Kliegl, 2014) can potentially be extended to account for early effects of contextual plausibility. If plausible words are pre-activated, then plausibility may also influence the relative difficulty of preview processing and therefore modulate the rate of accumulation of lexical information from the parafovea, such that target words preceded by plausible

previews would also be expected to receive more forced fixations than targets preceded by implausible previews.

Consistent with this interpretation, Schotter and Leinenger's (2016) quantile analysis showed that preview frequency effects were restricted to early quantiles of target fixation duration, whereas later quantiles were only affected by target frequency (but see Risse & Seelig, 2018). This suggested that forced fixations represent a subset of trials with very short fixations on the target word that is distinct from trials with long fixations, which are more likely to reflect properties of the target word and be impacted by the disruption caused by the display change from the preview to the target word. The present quantile analysis replicated Schotter et al.'s finding that preview frequency only affected short fixation durations, but showed a plausibility effect across the full time-course of fixation durations. The forced fixation account does not, therefore, appear to provide a complete account of plausibility-preview effects.

The combination of additive effects of frequency and plausibility on skipping, and interactive effects emerging in later quantiles of fixation duration in the present data, point to the contribution of multiple mechanisms underlying the plausibility preview effect. Early benefits on skipping and fixation duration from plausible parafoveal words may reflect the pre-activation of contextually plausible sentence continuations. The interaction of preview frequency and plausibility, reflecting the absence of a frequency effect for implausible previews, may be due to the impact of post-lexical integration failure for implausible parafoveal words: Because high-frequency words are retrieved more quickly, there is a higher probability that integration processes will be initiated—and fail—in time to affect eye-movement planning by cancelling forward saccades. Re-analysis processes triggered by such integration failures may also increase the probability of observing preview costs from the discrepancy between the preview and the target.

4.1 Conclusion

This study provided clear evidence that the plausibility of an upcoming word affects the decision to fixate or skip the word. This plausibility effect was independent of, and of a similar size to, the effect of parafoveal word frequency on skipping. Importantly, the plausibility effect did not differ for short and long words, suggesting that it may, in part, reflect contextual pre-activation of semantically and syntactically plausible words that is not adequately captured by the metric of cloze probability. The interactive effect of preview frequency and plausibility that emerged on fixation duration highlights the contribution of both preview benefits from plausible words due to contextual pre-activation and forced fixations, and preview costs from post-lexical integration failure of implausible words. Finally, our simulations of the E-Z Reader model revealed that, contrary to previous assumptions, the depth of parafoveal processing within a serial, staged architecture can be sufficient to allow post-lexical integration of parafoveal words to occasionally occur early enough to yield plausibility effects on skipping. These data add to a growing body of literature demonstrating the importance of incorporating more detailed accounts of the processes underlying higher-level language processing to provide a complete model of eye movements during online reading.

Author Note

We thank Morgan Weaving for her assistance with data collection, and Adrian Staub, Elizabeth Schotter, and an anonymous reviewer for helpful comments on an earlier version of this manuscript. Portions of these data were presented at the 59th Meeting of the Psychonomic Society, New Orleans, LA. This research was supported under Australian Research Council's *Discovery Projects* funding scheme (project numbers DP160103224 and DP190100719). The data and stimulus materials used in the experiment are available at https://osf.io/93phz/.

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.

Abbott, M. J., Angele, B., Ahn, Y. D., & Rayner, K. (2015). Skipping syntactically illegal *the* previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1703-1714.

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading? *Language and Linguistics Compass, 13*, e12344.

Angele, B., & Rayner, K. (2013). Processing *the* in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 649-662.

Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high- and low-frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1181-1203.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364-390.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Brothers, T., & Traxler, M. J. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition, 42*, 1894-1906.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language, 96*, 9-22.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance* (pp. 53-77). Oxford, UK: Oxford University Press.

Drieghe, D., Brysbaert, M., Desmet, T., & Da Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology, 16*, 79-103.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 954-969.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20,* 641-655.

Hutzler, F., Schuster, S., Marx, C., & Hawelka, S. (2019). An investigation of parafoveal masks with the incremental boundary paradigm. *PLOS One, 14,* e0203013.

Inhoff, A. W., Radach, R., Eiter, B. M., & Juhasz, B. J. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology, 56A,* 803-827.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262-284.

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *Quarterly Journal of Experimental Psychology, 66,* 581-600.

Koenker, R. (2018). *Quantile Regression.* (*R* package version 5.38). Retrieved from https://CRAN.R-project.org/package=quantreg

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126-1177.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88,* 22-60.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition, 37*, 913-934.

McGowan, V. A. & Reichle, E. D. (2018). The "risky" reading strategy revisited: New simulations using E-Z Reader. *Quarterly Journal of Experimental Psychology*, 71, 179-189.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for common designs*. (*R* package version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14*, 191-201.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research, 16,* 191-201.

Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology, 33*, 351-373.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2013). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 514-528.

Reichle, E. D. (2011). Serial attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements* (pp. 767-786). Oxford, UK: Oxford University Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in non-reading tasks: A unified framework for understanding the eye-mind link. *Psychological Review*, *119*, 155-185.

Reichle, E. D. & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience, 7*:*361*, doi: 10.3389/fnhum.2013.00361.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin and Review, 16,* 1-21.

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*, 1677-1683.

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word n + 2 in reading. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1026-1042.

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word n+1 during reading. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 653-667.

Risse, S., & Seelig, S. (2018). Stable preview difficulty effects in reading with an improved variant of the boundary paradigm. *Quarterly Journal of Experimental Psychology*.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition, 26*, 1270-1281.

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

Schotter, E. R., Leinenger, M., & von der Malsburg, T. (2018). When your mind skips what your eyes fixate: How forced fixations lead to comprehension illusions in reading. *Psychonomic Bulletin & Review, 25*, 1884-1890.

Schotter, E. R., Reichle, E. D., & Rayner, K. (2014). Rethinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N+2 preview effects. *Visual Cognition, 22*, 309-333.

Schotter, E. R., von der Malsburg, T., & Leinenger, M. (2018). Forced fixations, trans-saccadic integration, and word recognition: Evidence for a hybrid mechanism of saccade triggering in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1839-1866.

Schotter, E. R., & Leinenger, M. (2016). Reversed preview benefit effects: Forced fixations emphasize the importance of parafoveal vision for efficient reading. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 2039-2067.

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General, 140*, 407-433.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*, 311-327.

Staub, A., Grant, M., Astheimer, L. & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1-17.

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.

Veldre, A., & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception and Performance, 42*, 939-952.

Veldre, A., & Andrews, S. (2017). Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. *Psychonomic Bulletin & Review, 24*, 519-528.

Veldre, A., & Andrews, S. (2018a). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language, 100*, 1-17.

Veldre, A., & Andrews, S. (2018b). How does foveal processing difficulty affect parafoveal processing during reading? *Journal of Memory and Language, 103*, 74-90.

Veldre, A., & Andrews, S. (2018c). Parafoveal preview effects depend on both preview plausibility and target predictability. *Quarterly Journal of Experimental Psychology, 71*, 64-74.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language, 94*, 119-133.

Yang, J., Li, N., Wang, S., Slattery, T. J., & Rayner, K. (2014). Encoding the target or the plausible preview word? The nature of the plausibility preview benefit in reading Chinese. *Visual Cognition, 22*, 193-213.

Yang, J., Wang, S., Tong, X., & Rayner, K. (2012). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading & Writing, 25*, 1031-1052.

Zang, C., Du, H., Bai, X., Yan, G., & Liversedge, S. P. (2019). Word skipping in Chinese reading: The role of high-frequency preview and syntactic felicity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S. P. (2018). Skipping of the very-high-frequency structural particle de (的) in Chinese reading. *Quarterly Journal of Experimental Psychology, 71*, 152-160.