The role of preview validity in predictability and frequency effects on eye movements in reading

Adrian Staub[a]

Kirk Goddard[b]

[a]Department of Psychological and Brain Sciences

University of Massachusetts Amherst

[b]Basque Center on Cognition, Brain and Language

Keywords: eye movements; predictability; word frequency; parafoveal preview

Address correspondence to:

Adrian Staub, Department of Psychological and Brain Sciences, University of Massachusetts, 430 Tobin Hall, Amherst, MA 01003; email: astaub@psych.umass.edu; phone: (413) 545-5925; fax: (413) 545-0996.

The role of preview validity in predictability and frequency effects on eye movements in reading

Abstract

A word's predictability, as measured by its cloze probability, has a robust influence on the time a reader's eyes spend on the word, with more predictable words receiving shorter fixations. However, several previous studies using the boundary paradigm have found no apparent effect of predictability on early reading time measures when the reader does not have valid parafoveal preview of the target word. The present study directly assesses this pattern in two experiments, demonstrating evidence for a null effect of predictability on first fixation and gaze duration with invalid preview, supported by Bayes Factor analyses. While the effect of context independent word frequency is shown to survive with invalid preview, consistent with previous studies, the effect of predictability is eliminated with both unrelated word previews and random letter string previews. These results suggest that a word's predictability influences early stages of orthographic processing, and does so only when perceptual evidence is equivocal, as is the case when the word is initially viewed in parafoveal vision. Word frequency may influence not only early orthographic processing, but also later processing stages.

Introduction

It is well established that a word's predictability, as measured by cloze probability, influences how long the eyes spend on that word in reading, with predictable words receiving shorter fixations than less predictable words (e.g., Ehrlich & Rayner, 1981; Staub, 2011; see Staub, 2015, for a review). Predictable words are also more likely than unpredictable words to be skipped rather than directly fixated. The fact that these effects are in evidence in the earliest possible eye movement measures suggests that a word's predictability influences visual word recognition itself, rather a post-lexical process of integration. It appears that these effects are not due to discrete predictions of specific words, but rather to broad pre-activation of potential continuations (e.g., Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016).

The time the eyes spend on a word, once it is directly fixated, also depends on whether the word was visible parafoveally during the eyes' fixation on the previous word. This has been demonstrated in numerous studies using the boundary paradigm (Rayner, 1975). In this paradigm, a critical target word may be replaced by some other word, pseudoword, or nonword until the reader directly fixates the target. Therefore, while the reader is still fixating the previous word, the target word is not visible parafoveally; this is referred to as an *invalid preview* condition. Readers are usually not consciously aware of the change from preview string to target word, as this change is executed during a saccade. However, invalid preview reliably increases reading time on the target word, confirming that useful information is extracted from a word before it is directly fixated. A large literature has investigated the question of what kinds of information are, and are not, extracted during parafoveal viewing (see Schotter, Angele, & Rayner, 2012, for a review).

The present study addresses an apparent interaction between these two effects that was first observed by Balota, Pollatsek, and Rayner (1985). This study focused on the effect of preview type and predictability on the probability of word skipping. Balota et al. found that the high rate of word skipping that is typically seen with predictable words was present only with valid preview (i.e., when there is no change from preview to target) or when the preview string was an orthographically similar nonword (e.g., *cahc* as a preview for *cake*). This finding argues against an account of predictability-based skipping according to which it is due to a pure guess, without parafoveal visual input, as to the identity of an upcoming word. However, Balota et al. also found that the predictability effect on first fixation duration (the duration of the first eye fixation on the target word) and gaze duration (the sum of all first-pass fixation durations) was present only with valid preview or, to a lesser extent, with preview of a nonword that was orthographically similar to the target. When the preview was an orthographically different word or nonword, the predictability effect on early eye movement measures was entirely eliminated. Balota et al. did not directly comment on this aspect of the fixation duration pattern, as their main interest was in assessing the relative degree of benefit conferred by different preview types, rather than in assessing how the predictability effect itself might depend on preview type.

Several recent studies (Schotter, Lee, Reiderman, & Rayner, 2015, Choi, Lowder, Ferreira, Swaab, & Henderson, 2017; Veldre & Andrews, in press) have included similar manipulations. The relevant results of these studies are summarized in Table 1.[1] With the possible exception of two conditions from Schotter et al. (2015), these studies have not obtained significant predictability effects in invalid preview conditions. (Schotter et al. do not directly report the significance of these effects. Effects of predictability in this study are also somewhat difficult to interpret, as predictability was strongly related to sentence position.) Moreover, Choi et al. (2017) reported a significant reversed effect in first fixation duration, i.e., longer reading time for the predictable word. Notably, this reversed effect was obtained with a nonword preview that, like in Balota et al. (1985), was orthographically similar to the target word. (As we discuss below, none of these studies used nonword previews that are not wordlike, e.g., random letter strings.) Again, the influence of preview validity on the predictability effect has not been explicitly addressed in these studies, which have focused on preview benefit effects, rather than predictability effects.

[1] Table 1 does not include two studies (Juhasz, White, Liversedge, & Rayner, 2008; White, Rayner, & Liversedge, 2005) that have crossed target word predictability with the validity of the word length information that is available parafoveally, because the incorrect length previews in these experiments did preserve the letter identities of the target word.

We also omit from discussion a recent study (Sereno, Hand, Shahid, Yao, & O'Donnell, 2018) that included both predictability and word frequency manipulations and a parafoveal preview manipulation. The design of this study is sufficiently non-standard that it is difficult to compare its results to the results of other studies. First, unlike (to our knowledge) all other studies using the boundary paradigm, the preview manipulation was between experiments; subjects in one experiment read normally presented sentences, while subjects in the other experiment read only sentences with invalid parafoveal preview of a target word. Second, in contrast to all other experimental studies of the effect of predictability on eye movements, this manipulation involved neither a comparison of the same target word in different contexts, nor a comparison of different words in the same context. Rather, the predictability effect was assessed by comparing *different* words in *different* contexts. In other words, the test of the predictability effect did not involve experimental control over either target words or sentence contexts.

		predictability effect, valid preview (ms)		predictability effect, invalid preview (ms)		invalid preview type
	n	first fix	gaze	first fix	gaze	
Balota et al. (1985)	30	9	32	-2	15	visually similar nonword (e.g., cake → cahc)
				-4	7	semantically related word (e.g., cake → pies)
				2	-3	visually dissimilar nonword (e.g., cake → picz)
				3	-2	anomalous word (e.g., cake → bomb)
Schotter et al. (2015, Experiment 2)	72	5	10	8	11	synonym (e.g., lousy → awful)
				6	15	related word (e.g., lousy → great)
				1	0	unrelated word (e.g., lousy → rated)
Veldre & Andrews (2018)	95	15	24	-17	-20	contextually plausible word preview (e.g., guilty → insane)
				-7	5	contextually implausible, related word preview (e.g., guilty → courts)
				-6	0	contextually implausible, unrelated word preview (e.g., guilty → mirror)
Choi et al. (2017), younger adults	24	8	13	-8	1	nonword; single letter substitution (e.g., heart→ heant)
Choi et al. (2017), older adults	24	17	30	-13	-11	nonword; single letter substitution (e.g., heart→ heant)

Table 1. Predictability effects on first fixation and gaze duration (in milliseconds) with valid and invalid preview, in studies that have factorially manipulated predictability and preview validity.

One possible account of the elimination of the predictability effect with invalid preview might proceed as follows. While most eye fixations in normal reading terminate due to the completion of some stage of lexical processing, as proposed by eye movement models such as E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003; Reichle, Warren, & McConnell, 2009), there may also be a 'time out' mechanism, such that a fixation is terminated when it reaches a certain duration even if the lexical processing that would normally trigger a saccade program to the next word has not completed (see, e.g., Henderson & Ferreira, 1990). Indeed, such a 'time out' mechanism must be operative in some cases, or inspection of a word that is not known to the reader would continue indefinitely. Denying parafoveal preview creates a situation in which the first fixation on the target word is very long, and as a consequence, inspection of the target word may often be terminated by this 'time out' mechanism, rather than by the completion of the stage of lexical processing that would normally trigger a saccade program to the next word. Thus, the predictability of the target has little effect on first fixation or gaze duration with invalid preview because, whether the word is predictable or not, most target word inspections are terminated by this 'time out' rather than by normal mechanisms.

An obvious prediction of such an account, however, is that other lexical variables should also have little or no effect on first fixation and gaze duration when the reader is denied valid preview. This prediction has been tested for the word frequency variable which, like predictability, has a sizable and reliable influence on early reading time measures (Rayner & Duffy, 1986; Staub, White, Drieghe, Hollway, & Rayner, 2010). Several studies (Choi & Gordon, 2013; Inhoff & Rayner, 1986; Kennison & Clifton, 1995; Reingold, Reichle, Glaholt, & Sheridan, 2012, Risse & Kliegl, 2014) have manipulated both word frequency and the validity of the target word preview. These studies are summarized in Table 2. While the results are not entirely consistent across studies, some frequency effect on first fixation and gaze duration typically survives with invalid preview. This persistence of the effect of target frequency has been obtained in studies with unrelated word previews (e.g., Risse & Kliegl, 2014, Experiment 1) and in studies with word-like (Choi & Gordon, 2013) and non word-like (Kennison & Clifton, 1995, high-span readers) nonword previews. Particular weight should be given to Reingold et al. (2012), which is by far the highest-powered of these studies, both in terms of the number of subjects (n = 60) and in terms of the number of items; each subject read 240 items, 60 at each level of target word frequency and preview validity. (As we discuss below, this unusually large number of items was included in order to fit the ex-Gaussian distribution to individual subjects' fixation duration distributions in each condition.) Reingold et al., using pronounceable nonword previews, found that invalid preview attenuated the first fixation effect of target word frequency, reducing this effect from 20 ms to 9 ms. However, there was no additional attenuation of the gaze duration effect, which was reduced from 58 ms to 47 ms, i.e., exactly the same 11 ms reduction that is present in first fixation duration.

		frequency effect, valid preview (ms)		frequency effect, invalid preview (ms)		invalid preview type
	n	first fix	gaze	first fix	gaze	
Inhoff & Rayner (1986)	24	20	35	0	25	strings of letters of same shape, or x-masks (e.g., music → nacle; music → xxxxx), not reported separately; effect sizes estimated from figure
Kennison & Clifton (1995), high-span readers	24	13	38	33	23	random consonant letter string (e.g., animal → ngpfmx)
Kennison & Clifton (1995), low-span readers	24	17	28	-4	4	random consonant letter string (e.g., animal → ngpfmx)
Choi & Gordon (2013, Experiment 1)	24	21	32	25	44	nonword constructed by transposition of internal letters of target (e.g., north → nroth)
Reingold et al. (2012)	60	20	58	9	47	pronounceable nonword (e.g., table → purty)
Risse & Kliegl (2014, Experiment 1)	29	22	22	13	11	LF preview word (e.g., rope) for HF target (e.g., book), or vice-versa
Risse & Kliegl (2014, Experiment 2)	31	10	17	5	-2	LF preview word (e.g., rope) for HF target (e.g., book), or vice-versa

Table 2. Frequency effects on first fixation and gaze duration (in milliseconds) with valid and invalid preview, in studies that have factorially manipulated word frequency and preview validity.

Thus, while invalid parafoveal preview appears to eliminate the effect of target word predictability on both first fixation duration and gaze duration, it does not appear to eliminate the effect of target word frequency on these measures. Though the empirical picture is not perfectly clear, invalid preview may somewhat attenuate, but not eliminate, the effect of frequency on first fixation duration, and there may be no additional attenuation of the gaze duration effect. If this preservation of the frequency effect with invalid preview is robust, it provides a strong argument against the 'time out' account of the elimination of the predictability effect with invalid preview.

This apparent dissociation between the effects of predictability and frequency with invalid preview is surprising from the perspective of the best known model of eye movements in reading, E-Z Reader (Reichle et al., 1998, 2003, 2009). This model proposes that frequency and predictability additively influence the duration of each of two stages of lexical processing, denoted L1 and L2 in the model. It is the completion of the L1 stage for the currently fixated word that triggers the initiation of a saccade to the next word. Fixation durations on a target word will be longer when parafoveal preview is absent, as lexical processing begins later in this case. However, both frequency and predictability still exert an influence on the duration of L1, and it is still the duration of L1 that determines how long the eyes remain on the word. This architecture does not predict a pattern in which frequency does exert an influence, but predictability does not.

This dissociation is also not predicted by recent findings (Risse & Kliegl, 2014; Veldre & Andrews, 2016, 2017, 2018) emphasizing the role of preview processing difficulty in invalid preview conditions. Reading times on a target word appear to be especially lengthened when the preview is a low frequency word (Risse & Kliegl, 2014), or a contextually anomalous word (Veldre & Andrews, 2016, 2017, 2018). These findings suggest that a parafoveal word may receive more complete processing than has been previously believed, and that the cost of invalid preview is not due only to the lack of parafoveal processing of the target. However, such findings do not explain why a difficult-to-process preview should eliminate the influence of target word variables on fixation durations, or why this elimination should be restricted specifically to the effect of predictability.

The first goal of the present study is to directly assess the evidence for the lack of a predictability effect with invalid preview, a null effect that has previously appeared only incidentally in other studies, and the evidence for the presence of a frequency effect with invalid preview. This dissociation is assessed in a within-subject design in Experiment 1. We compute Bayes Factors (Kass & Raftery, 1995; see Abbott & Staub, 2015, for application to eye movement data) to determine the strength of the evidence for the null effect of predictability. To anticipate, we find strong evidence for the null effect of predictability with invalid preview, coupled with positive evidence for a frequency effect.

Experiment 2 of the present study is designed to rule out another possible explanation of the elimination of the predictability effect on early reading time measures with invalid preview; a version of this hypothesis, which we refer to as the *lexical suppression* hypothesis, is advanced by Parker, Kirkby, and Slattery (2017). In most of the previous studies listed in Table 1, as well as in Experiment 1 of the present study, the invalid preview string is a word. In the few experimental conditions listed in Table 1 in which the preview is not a word, it is a highly word-like nonword (Balota et al., 1985; Choi et al., 2017). Thus, in all of these studies, processing of the the invalid parafoveal preview string is likely to activate a lexical representation. When this lexical representation is inconsistent with the reader's expectations, this activation may cause the reader's lexical expectations to be rapidly suppressed. These expectations would then no longer be in force when the target word is directly fixated, and as a result no predictability effect would be in evidence in first fixation duration or gaze duration on that word.

The lexical suppression hypothesis predicts that a random letter string as invalid preview should not suppress activation of a predictable word, or at least, it should not have as great a suppressive effect, as parafoveal processing of this string would cause little activation of any specific unexpected word. If a predictability effect survives when the preview is a random letter string, this result would support the lexical suppression hypothesis. On the other hand, if the predictability effect is also eliminated when the preview is a random letter string, this would provide evidence against this hypothesis. Experiment 2 directly compares the effects of unrelated word previews and random letter string previews on processing of a predictable word. To anticipate once again, we do not find evidence of a difference between these two preview conditions; the predictability effect on first fixation and gaze duration is eliminated in both cases.

The experiments, then, place on firm empirical ground the conclusion that invalid preview eliminates the predictability effect on first fixation and gaze duration, and provide evidence against both the time out hypothesis and the lexical suppression hypothesis as explanations for this phenomenon. After presenting the experiments, we propose a tentative novel explanation. We adopt a Bayesian model of word recognition (Norris, 2006; Smith & Levy, 2013), whereby the effect of predictability is understood as reflecting the influence of a prior probability distribution that is combined with bottom-up perceptual input. We consider how some modifications to the assumptions of Norris' (2006) Bayesian Reader model, which we argue are independently motivated, would account for the dependence of predictability effects, but not frequency effects, on valid parafoveal preview.

Experiment 1

The goal of Experiment 1 was to directly test the dissociation between predictability and frequency effects, with regard to their persistence with invalid preview: While invalid preview eliminates the effect of predictability on early reading time measures, it appears not to eliminate the effect of context-independent word frequency on the same measures (e.g., Reingold et al., 2012). However, this dissociation has never been demonstrated in a single experiment. Rather, a number of experiments have failed to find predictability effects on first fixation and gaze duration with invalid preview, and separate experiments have found significant frequency effects on the same measures, with invalid preview. Here we assess whether, in a design in which word frequency, predictability, and preview validity are fully crossed, the pattern will emerge in which both frequency and predictability influence the critical early reading time measures with valid preview, but only frequency does so with invalid preview.

Method

Subjects. Participants were 80 students at UMass Amherst who received course credit for their participation; both Experiment 1 and Experiment 2 were approved by the UMass IRB. All participants were speakers of English as a first language, and none reported any history of reading or language disorder. Eight subjects were excluded based on a criterion of losing more than 33% of experimental trials to track loss or blink on first pass reading of the critical word, or to incorrectly timed display change; the criterion for display change timing was that the change was initiated during the initial saccade into the target word, and completed no more than 7 ms into the first fixation on that word. These exclusions leave 72 subjects in the analysis.

Materials. Eighty of the 160 items used by Kretzschmar, Schlesewsky, and Staub (2015), in which frequency and predictability were both manipulated, were adopted for this experiment. (The materials for both Experiments 1 and 2 are available from the authors upon request.) Each subject read all 80 of the critical sentences. The sentences made use of 20 high frequency (HF) target words, with mean Zipf frequency (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) based on the Subtlex Corpus (Brysbaert & New, 2009) of 4.87 (sd = .47), and 20 low frequency (LF) target words (mean Zipf frequency = 3.17, sd = .44); the frequency distributions were non-overlapping. The HF and LF words were permitted to differ slightly in length (mean of 4.75 characters for HF words, sd = .79; mean of 5.25 characters for LF words, sd = 1.07; $t(71.62) = 2.41$, $p < .02$). Each target word was used in two sentences, once following a context that rendered this word highly predictable and once following a context that rendered it unpredictable. In the high predictability conditions, predictability differed slightly between HF and LF words (mean cloze of .84 for HF words, sd = .14; mean cloze of .69 for LF words, sd = .17; $t(35.99) = 3.09$, $p < .01$). There was no difference in the low predictability condition (mean cloze of .007 for HF words, sd = .01; mean cloze of .004 for LF words, sd = .01; $p > .5$). The position of the target word in the sentence did not significantly vary based on either frequency or predictability (*p*s > .7).

Preview was manipulated such that the subject had either valid preview of the target, or preview of an unrelated, contextually anomalous word matched in length to the target, with matched positions of ascending and descending letters. The invalid preview words for HF and LF targets did not differ significantly in frequency (HF mean 3.92, sd = .96; LF mean 3.64, sd = .86; $t(77.05) = 1.39$, $p = 0.17$). Each subject read each target word once with valid preview and once with invalid preview; if the target was predictable when read with valid preview, it was unpredictable when read with invalid preview, and vice versa. In sum, each subject read 10 sentences in each of the eight cells of the design defined by the combinations of frequency, predictability, and preview type. Example items with HF and LF targets are in (1) and (2), respectively, with the preview strings in parentheses; in the (a) versions the target is predictable, in the (b) versions unpredictable.

(1) a. Sylvie's favorite part about Christmas was adding ornaments to the (tree/bear) tree with her family.

b. The other day, Mr. Hudson made an official complaint about our (tree/bear) tree that shades some of his property.

(2) a. Once he got to the bar, the man ordered a gin and (tonic/backs) tonic with a lime wedge on the side.

b. Mary told her friends that she doesn't like the taste of (tonic/backs) tonic water, but that club soda was okay.

The 80 critical sentences were randomly intermixed with 20 filler sentences of various types, which were followed by two-alternative comprehension questions. The 100 sentences were preceded by eight practice sentences.

Procedure. Eye movements were recorded using an EyeLink 1000 (SR Research, Toronto, ON, Canada) eyetracker, interfaced with a PC computer. The sampling rate was 1000 Hz. Subjects were seated 55 cm from a CRT monitor, with 1024 x 768 resolution and a screen refresh rate of 120 Hz. At this distance the resolution of the eyetracker was substantially less than one character. Only the movement of the right eye was recorded.

All sentences were displayed on a single line in 11-point Monaco font. Before the experiment began, each subject was instructed to read for comprehension. A three-point calibration procedure was performed at the start of the experiment and as needed between trials. The subject triggered each sentence by fixating a box at the left edge of the monitor. The experiment lasted approximately 30 minutes. The experiment was implemented with the EyeTrack software, and initial stages of data analysis were carried out with Robodoc and EyeDry (http://blogs.umass.edu/eyelab/software/).

All subjects obtained at least 78% correct on the comprehension questions, with a mean of 91.6%. Trials were excluded if there was a blink or track loss during first pass reading of the target word, or if the display change did not trigger or complete during the correct saccade. As noted above, eight subjects were excluded due to excessive data loss. For the remaining 72 subjects, blinks, track loss, or display change error resulted in deletion of 17.4% of trials, leaving 4757 trials for inclusion in the analysis. Individual eye fixations less than 80 ms in duration and within one character of a previous or subsequent fixation were incorporated into this neighboring fixation. No other data trimming was carried out.

Subjects were asked in a post-experiment debriefing if they noticed anything 'strange or unusual about the text,' and if so, what they noticed. Thirty-eight of the 72 included subjects (53%) reported no awareness of the display change in the post-experiment debriefing. Of the 34 subjects who reported some awareness of a change or flicker, 23 estimated that this occurred on 10% of trials or fewer; in fact, there were changes on 40% of trials in the experiment. Analyses that included subjects' awareness of display changes as a predictor did not reveal any reliable main effects in the two experiments, or reliable interactions with the experimental manipulations, so we leave this factor out of the models reported below.

Results

Analyses focused on the target word, where we report the probability that the word was skipped rather than fixated on first pass reading, as well as four eye movement measures that are contingent on the word being fixated; for all of these measures, a trial is excluded from analysis if the region was skipped on first pass reading. *First fixation duration* is the duration of the first eye fixation on the word, on the first pass through the sentence. *Gaze duration* is the sum of the duration of all first pass fixations on the word, before the eyes first leave the word to either the left or right. If the reader made only a single first pass fixation before leaving the word, first fixation duration and first pass time are identical for that trial. *Go-past time* is the sum of all fixation durations beginning with the first on the word, and including all fixations until the reader moves past the word to the right; this measure includes the durations of any fixations made after a regressive eye movement to the left. Finally, *regression proportion* is the proportion of trials on which first pass inspection of the word ended with an eye movement to the left rather than the right. Means for each measure on the target word are shown in Figure 1.

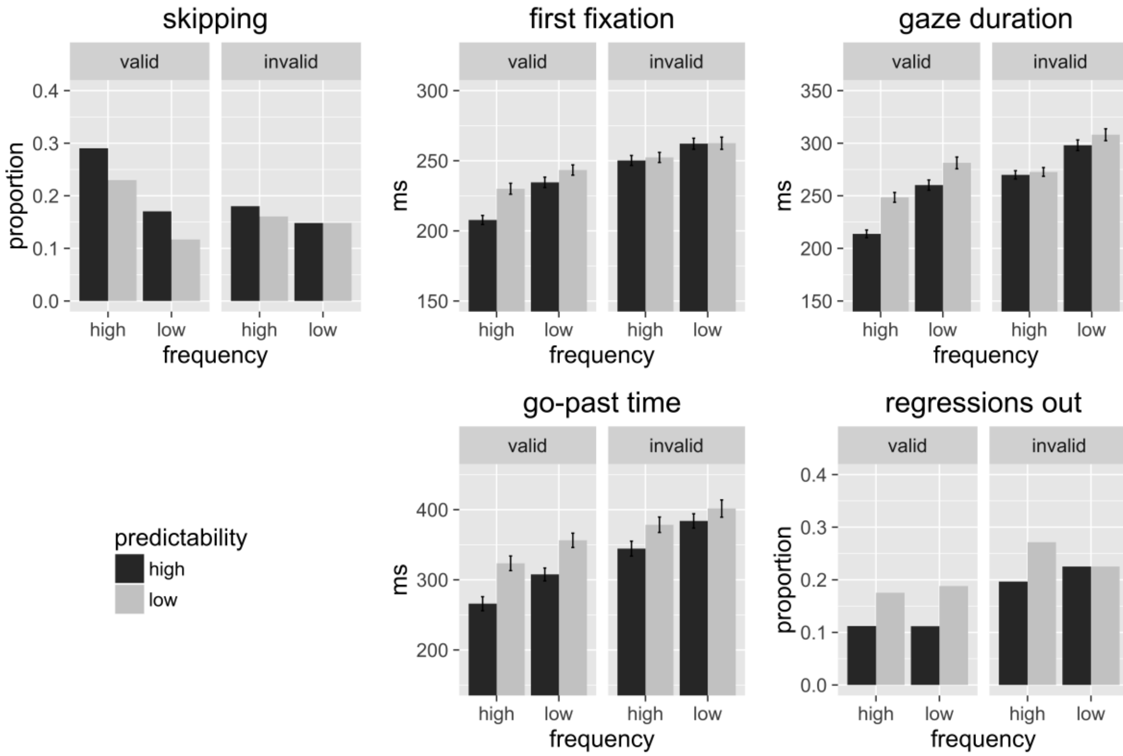

Figure 1. Experiment 1 target word mean first fixation duration, gaze duration, and go-past time, with standard error; skipping proportion, and regression proportion.

Our initial analysis was carried out by means of linear mixed effects models of reading times and mixed effects logistic regression models of skipping and regression probability, implemented using the lme4 package (version 1.1-7; Bates, Maechler, Bolker, & Walker, 2015) for the R statistical programming environment (version 3.1.2; R Core Team, 2014). Fixed effects in these models were centered, with the factor levels coded as -.5 and .5. Random factors included intercepts for subjects and items, as well as random subject and item slopes for each of the three fixed effects (predictability, frequency, and preview validity; the frequency manipulation was a between-item manipulation, so models did not include random by-item slopes for frequency). Models including random interaction slopes did not converge. We treat $|t|$ or $|z| > 2$ as indicating statistical significance.

The results of these models are shown in Table 3, and are easily summarized. As expected, the effects of both frequency and predictability on skipping were eliminated when the target word was not actually visible parafoveally, resulting in significant interactions between both factors and preview validity. All three reading time measures show an identical pattern: (a) there are significant main effects of each of the three factors (longer reading times for low frequency words, low predictability words, and with invalid preview); (b) there are significant predictability-by-preview interactions, such that the effect of predictability is not as pronounced with invalid preview as with valid preview; (c) the frequency-by-preview interaction never approached significance; and (d) the frequency-by-predictability interaction also did not approach significance, consistent with many previously findings of additive effects of these variables (e.g., Rayner, Ashby, Pollatsek, & Reichle, 2004, see Staub, 2015, for a review), and replicating the additive effects found with a superset of the same items by Kretzschmar et al. (2015). Both preview validity and predictability influenced the probability of a regression from the target word, but their interaction did not reach significance. [2]

[2] To rule out the possibility that the critical frequency effects in this experiment were actually due to the .5 character difference in mean length between HF and LF words, we also constructed models of both first fixation duration and gaze duration in which word length was included as a fixed effect, along with frequency. The effect of length never approached significance ($|t| < .5$), while the effects of frequency were essentially identical to the effects in the models we report below.

Measure	Effect	Estimate	SE	*t/z*-value
Skipping Probability	Frequency	-0.64	0.16	**-4.00**
	Predictability	-0.31	0.17	-1.75
	Preview	-0.24	0.10	**-2.34**
	Freq x Pred	-0.07	0.33	-0.21
	Freq x Preview	0.71	0.19	**3.82**
	Pred x Preview	0.43	0.17	**2.52**
	Freq x Pred x Preview	0.31	0.33	0.95
First Fixation Duration	Frequency	14.82	3.31	**4.47**
	Predictability	7.62	3.58	**2.13**
	Preview	27.90	2.64	**10.58**
	Freq x Pred	-7.55	7.06	-1.07
	Freq x Preview	-8.71	4.98	-1.75
	Pred x Preview	-16.81	4.87	**-3.45**
	Freq x Pred x Preview	10.55	9.73	1.08
Gaze Duration	Frequency	34.29	6.26	**5.48**
	Predictability	15.40	5.87	**2.62**
	Preview	36.43	3.46	**10.52**
	Freq x Pred	-4.41	11.55	-0.38
	Freq x Preview	-7.99	6.71	-1.19
	Pred x Preview	-25.99	6.15	**-4.23**
	Freq x Pred x Preview	18.59	12.28	1.51
Go-Past Time	Frequency	29.88	14.70	**2.03**
	Predictability	39.05	17.06	**2.29**
	Preview	64.57	8.69	**7.43**
	Freq x Pred	-14.25	33.63	-0.42
	Freq x Preview	-9.61	16.77	-0.57
	Pred x Preview	-30.71	13.94	**-2.20**
	Freq x Pred x Preview	-13.36	27.86	-0.48
Regression Probability	Frequency	-0.01	0.14	-0.09
	Predictability	0.37	0.17	**2.25**
	Preview	0.67	0.13	**5.32**
	Freq x Pred	-0.19	0.32	-0.60
	Freq x Preview	-0.17	0.22	-0.76
	Pred x Preview	-0.37	0.19	-1.92
	Freq x Pred x Preview	-0.72	0.37	-1.97

Table 3. Experiment 1 statistical results from mixed-effects models for the target word, with all three factors included. Model specification is described in the text.

However, the specific hypotheses that motivated this experiment, namely that the predictability effect on early reading time measures should be absent with invalid preview, while the frequency effect should be present, are not tested by this full model. In particular, the finding of significant predictability-by-preview interactions in the reading time measures does not demonstrate the absence of a predictability effect in the invalid preview conditions. Table 4 presents the results of separate models of the valid and invalid preview conditions. (Note that unlike for the full model, the reading time models were able to converge with the maximal random effect structure, including random slopes for the interaction effect.) In the valid preview conditions, frequency and predictability additively influenced both the probability that the target word was skipped and the early reading time measures of first fixation and gaze duration. The results are also consistent with previous demonstrations that predictability may affect the probability of a regression (e.g., Staub, 2011; Kretzschmar et al., 2015), while frequency generally does not (Abbott & Staub, 2015). The critical result, however, is that in the invalid preview models, the only effects that reached (or even approached) significance were the effects of frequency on first fixation duration and gaze duration; the effects of predictability on these measures were estimated to be -0.35 ms and 3.02 ms, respectively.

Measure	Effect	Estimate	SE	*t/z*-value
Valid Preview				
Skipping Probability	Frequency	-0.99	0.20	**-4.91**
	Predictability	-0.42	0.17	**-2.40**
	Freq x Pred	-0.16	0.32	-0.50
First Fixation Duration	Frequency	19.40	4.19	**4.63**
	Predictability	15.71	4.63	**3.39**
	Freq x Pred	-13.92	9.69	-1.44
Gaze Duration	Frequency	38.23	7.85	**4.87**
	Predictability	28.70	7.85	**3.66**
	Freq x Pred	-13.44	15.89	-0.85
Go-Past Time	Frequency	33.92	18.12	1.87
	Predictability	53.62	18.68	**2.87**
	Freq x Pred	-9.77	37.18	-0.26
Regression Probability	Frequency	0.21	0.25	0.84
	Predictability	0.85	0.28	**3.07**
	Freq x Pred	-0.01	0.48	-0.02
Invalid Preview				
Skipping Probability	Frequency	-0.20	0.16	-1.30
	Predictability	-0.25	0.22	-1.10
	Freq x Pred	0.12	0.41	0.29
First Fixation Duration	Frequency	10.68	4.27	**2.50**
	Predictability	-0.35	4.27	-0.08
	Freq x Pred	-2.17	9.05	-0.24
Gaze Duration	Frequency	31.18	6.15	**5.07**
	Predictability	3.02	5.65	0.53
	Freq x Pred	5.85	11.52	0.51
Go-Past Time	Frequency	28.81	14.94	1.93
	Predictability	26.62	18.89	1.41
	Freq x Pred	-17.35	36.82	-0.47
Regression Probability	Frequency	-0.13	0.16	-0.81
	Predictability	0.19	0.17	1.14
	Freq x Pred	-0.54	0.32	-1.68

Table 4. Experiment 1 statistical results from separate mixed-effects models for valid and invalid preview conditions. Model specification is described in the text.

Finally, to directly assess the evidence for the critical null effects, we supplemented our analysis of target word reading times with a Bayes Factor analysis. The Bayes Factor is an odds ratio, indicating the relative marginal likelihood of the data under a model that does include a given effect (in this case, an effect of predictability) and under one that does not. The larger model is penalized for the flexibility that it gains through the inclusion of an additional parameter. Thus, this method can be used to quantify the extent to which the data favor the smaller model. We compute Bayes Factors using the *lmBF()* function from the BayesFactor package for the R environment (Morey, Rouder, & Jamil, 2015). In all analyses, we assumed the default Cauchy prior for effect size; see Abbott and Staub (2015) for discussion.

For the valid and invalid preview conditions separately, Bayes Factors were calculated to compare a model that included subject and item intercepts and an effect of frequency to a null model, and to compare a model that included subject and item intercepts and both frequency and predictability effects to the null model. The ratio of these Bayes Factors provides a Bayes Factor for the predictability effect. For the invalid preview conditions, this analysis favored the frequency-only models of first fixation duration and gaze duration by factors of 8.26 and 7.78, respectively. By contrast, for the valid conditions, the model that included both main effects was favored over the frequency-only model for first fixation and gaze duration by factors of 7.09 and 35.49, respectively. Thus, this analysis delivers evidence in favor of a predictability effect on these measures with valid preview and against such an effect with invalid preview. As a guide to interpretation, Jeffreys (1961) proposed the rule-of-thumb that a Bayes Factor between 3.2 and 10 should be regarded as "substantial" evidence in favor of the null, and a Bayes Factor greater than 10 should be regarded as "strong" evidence in favor of the null.

Discussion

The present experiment provides clear confirmation, in a within-subject design, of the dissociation that has previously appeared across multiple experiments: The predictability effect on early reading time measures is eliminated by invalid preview, but the frequency effect is not. With invalid preview the effect of frequency was significant for first fixation and gaze duration. Indeed, though these effects were numerically smaller with invalid preview than with valid preview, the frequency-by-preview interaction was not significant. The present results closely mimic those of Reingold et al. (2012), in that the effect of frequency on first fixation duration was numerically smaller with invalid preview than with valid preview (11 ms vs 19 ms), but this difference did not increase at all in the gaze duration measure (31 ms vs 38 ms). Thus, to the extent that invalid preview diminishes the frequency effect, it seems to do so only for the first fixation on the target word. On the other hand, while the effect of predictability with valid preview was substantial (16 ms and 29 ms for first fixation duration and gaze duration, respectively), this effect was essentially nonexistent with invalid preview (0 ms and 3 ms for first fixation and gaze, respectively). Bayes Factors favored a model of the invalid preview data that did not include a predictability effect, for both measures.

The pattern in later measures is less clear. Go-past time patterned like the earlier reading time measures in the full model (i.e., three main effects, and a significant predictability-by-preview interaction). However, in the model of the valid preview conditions, only the predictability effect was present, while in the model of the invalid preview conditions, no effect reached significance. Inspection of the condition means, however, suggests that both predictability and frequency have some effect with valid preview, and that these effects are reduced but not eliminated with invalid preview. We return to the interpretation of the go-past data in the General Discussion.

Experiment 2

Experiment 2 was designed to determine if the lack of predictability effect with invalid preview depends on the lexical status of the preview string. The experiment investigated whether a random letter string preview, like an invalid word preview, eliminates the predictability effect. As discussed in the Introduction, the use of a random letter string preview was motivated by the possibility that the elimination of the predictability effect by invalid preview is due to *lexical suppression*: When an (invalid) preview word is presented parafoveally, the activation of this word may result in rapid suppression of the reader's lexical expectations, prior to direct fixation on the target word. This account predicts that a random letter string preview should not eliminate the predictability effect on first fixation and gaze duration.

Method

Subjects. Participants were 61 students from the same pool as Experiment 1, who did not participate in that experiment. One subject was excluded due to below-chance performance on comprehension questions. An additional 13 subjects were excluded based on a criterion of losing more than 33% of experimental trials to track loss or blink on first pass reading of the critical word, or to incorrectly timed display change. These exclusions leave 47 subjects in the analysis.

Materials. In each of 90 critical items, a single target word was rotated through six conditions. The six conditions resulted from a 2 (high vs. low target word predictability) x 3 (valid vs unrelated word vs random lettter string preview) design. Each subject therefore read 30 target words with valid preview, and 60 with invalid preview. An example item with the target word *voice* is in (3), with the three preview strings in parentheses; (3a) is the predictable version, and (3b) the unpredictable version.

3. (a) She could tell he was mad by the tone of his (voice/color/wmlmn) voice and his rotten demeanor.

   (b) He can't believe what is happening to his (voice/color/wmlmn) voice even though the doctor warned him.

The predictable conditions were selected from items developed and normed by Block and Baldwin (2010). The mean cloze probability of the target word for these 90 items, in the Block and Baldwin norms, was .93 (sd = .026). Post-target sentence completions were constructed for the present study. The corresponding unpredictable versions were newly constructed for this study. These new contexts, up to but not including the target word, were presented to 20 Amazon Mechanical Turk workers, who were asked to provide the word that seemed most likely to come next. For all but eight of the items, the target word was never provided, while for 8 of the items it was provided by one subject; the mean cloze probability of the target words was .004[3]. The target word's mean position, in number of words from the start of the sentence, was nearly identical in the predictable and unpredictable frames (predictable mean = 9.16, sd = 1.75; unpredictable mean =

[3] For two items, replacement unpredictable contexts were constructed after norming. We assume zero cloze probability for the target words *deaf* and *name* in the contexts *The energetic pit bull in the park was* and *The diligent waiter saw his*, respectively.

9.17, sd = 1.93). The word that immediately preceded the target was identical in the two frames.

The target words averaged 4.2 characters in length, ranging from three to six characters. They were generally of moderate to high frequency, with a mean Zipf frequency based on the Subtlex corpus of 4.74 (sd = .61). The preview words were selected to match the target words in length and in the position of ascending and descending letters. These words were unrelated to the target word in meaning, and anomalous in their context. The random letter string previews were created by a program that randomly substituted a different letter for each letter of the target, with the constraint that positions of ascenders and descenders were preserved. (Thus, the strings are more correctly referred to as pseudo-random.)

These items were arranged into six experimental lists. Each participant read 15 sentences in each of the six experimental conditions, and each item was read by an approximately equal number of participants in each of the six conditions. The 90 critical sentences were randomly intermixed with 48 other sentences from an unrelated experiment on subject-verb agreement processing. These 48 sentences were all followed by two-alternative comprehension questions. The 138 sentences were preceded by eight practice sentences.

Procedure. The procedure was identical to Experiment 1. The experiment lasted approximately 40 minutes.

As noted above, one subject was excluded due to below chance performance on comprehension questions; all others achieved at least 71% correct, with a mean of 85.6%. Also as noted above, 13 subjects lost more than 33% of trials based on blink or track loss on first pass reading of the critical word, or incorrectly times display change, and these subjects were excluded from subsequent analysis. For the remaining 47 participants, blinks, track loss or display change error resulted in deletion of 19.2% of trials, leaving 3419 trials for inclusion in the analysis.

In the post-experiment debriefing, twenty of the 47 included subjects (43%) did not report noticing any display changes. Of the 27 subjects who reported some awareness of a change or flicker, the median estimate of the percentage of trials on which it occurred was 20%; the actual percentage of trials in the experiment that contained a display change was 44%. Many of the subjects who did notice changes commented on the random letter strings, which have been shown, in an explicit display change detection task (Angele, Slattery, & Rayner, 2016), to be highly detectable compared to word-like previews.

Results

We report the same five eye movement measures as for Experiment 1. Because the preceding word was held constant across conditions in this experiment, unlike in Experiment 1, we report these measures for this pre-target word as well as for the target word. This permits an assessment of potential parafoveal-on-foveal effects, i.e., effects of the manipulations of target word predictability and preview type that occur before the target is directly fixated. We expected that the orthographically illegal letter string previews might increase reading times on the preceding word (see Schotter et al., 2012, for a review of relevant findings), but we did not expect any other parafoveal-on-foveal effects (Brothers, Hoversten, & Traxler, 2017).

Condition means for the pre-target word are shown in Figure 2, and for the target word in Figure 3. We describe the qualitative patterns before discussing the statistical models. Reading times on the pre-target word appear to be slightly inflated – particularly go-past time – in the random letter string preview conditions. It also appears that regressions from the pre-target word may have been more common when the target word was unpredictable. Skipping of the target word was less frequent in the invalid word preview condition than in the valid preview condition, and less frequent still in the random letter string preview condition; this is expected on the assumption that word skips result from relatively complete lexical processing (Gordon, Plummer, & Choi, 2013). Moreover, it appears that predictability influenced target word skipping, but only with valid preview. There was clearly a preview validity effect on reading times on the target word. In addition, it appears that the predictability effect on first fixation and gaze duration was restricted to the valid preview conditions. Critically for the present study, there is no hint that the predictability effect survives in these measures with random letter string preview; it is essentially nonexistent in both invalid preview conditions. On the other hand, it appears that a predictability effect on go-past time may be present in all preview conditions. Finally, there was a clear effect of predictability on regressions from the target word, as well as an effect of preview validity. Regressions are most common with random letter string preview.

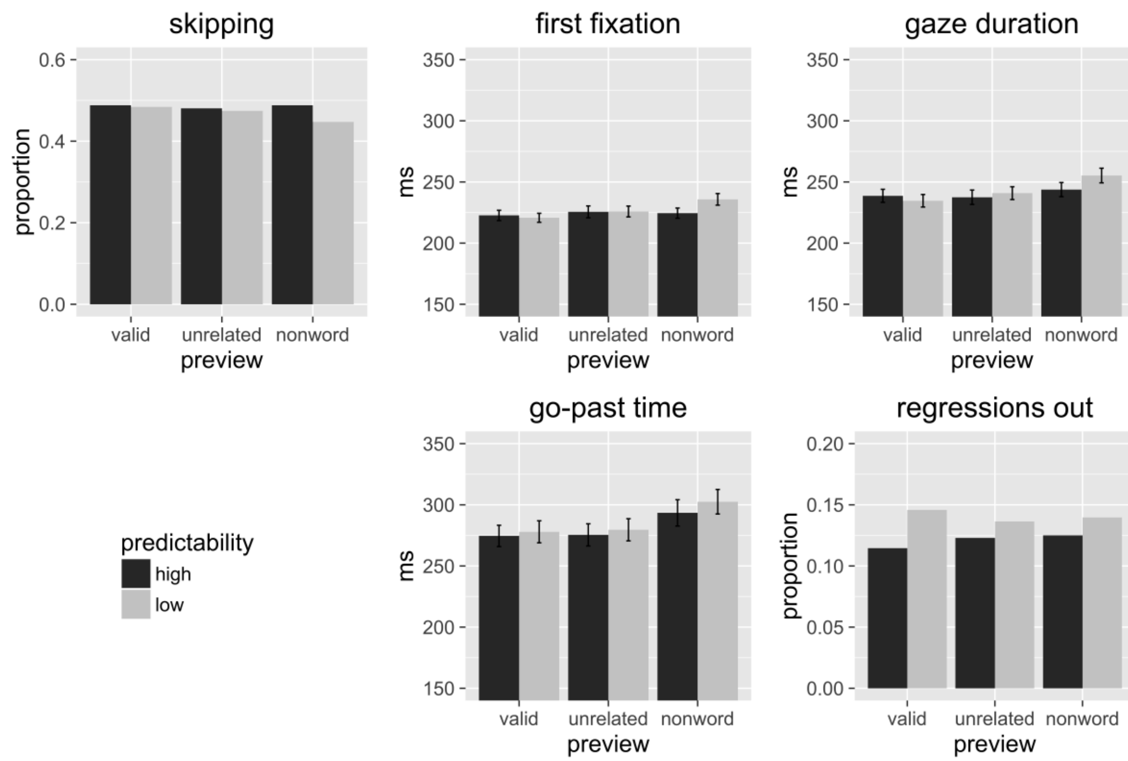

Figure 2. Experiment 2 pre-target word mean first fixation duration, gaze duration, and go-past time, with standard error; skipping proportion, and regression proportion.

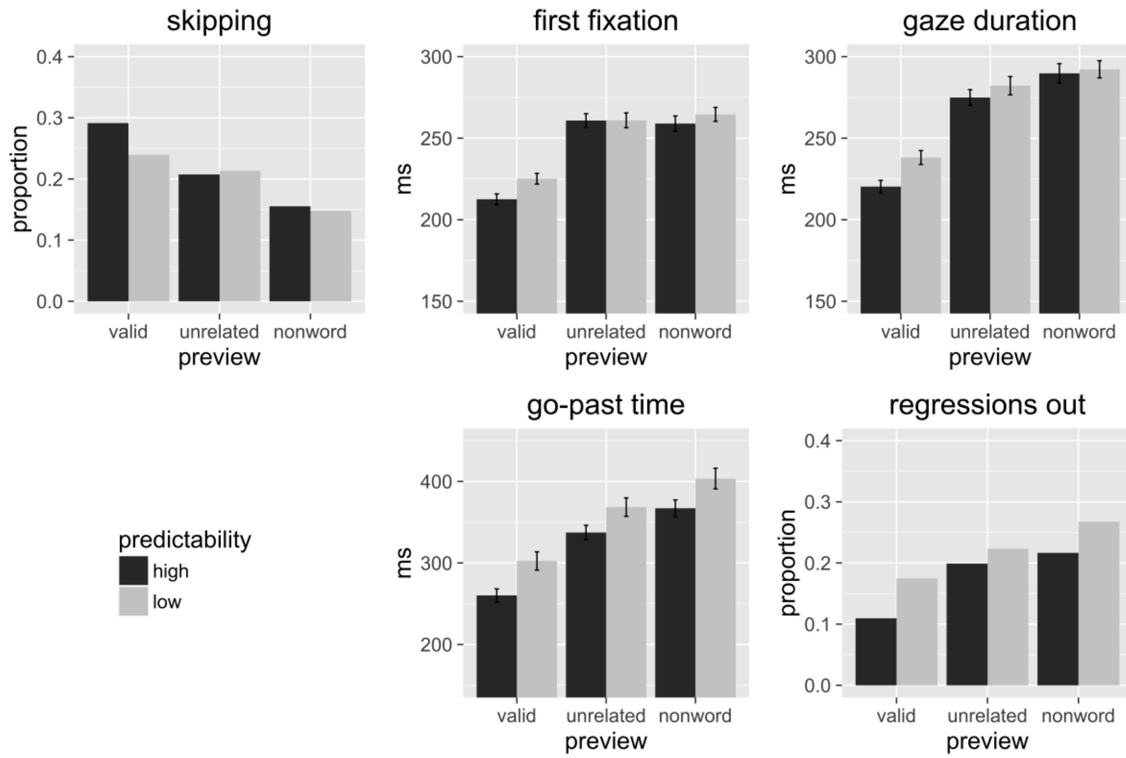

Figure 3. Experiment 2 target word mean first fixation duration, gaze duration, and go-past time, with standard error; skipping proportion, and regression proportion.

Initial analyses for both the pre-target word (Table 5) and target word (Table 6) were carried out by means of linear or logistic mixed-effects models in which the fixed effects were predictability (centered, with the predictable condition coded as -.5, and the unpredictable condition as .5) and two orthogonal contrasts representing the manipulation of preview type. The first contrast, which assess the effect of preview validity, compared the valid preview condition (coded as -.5) to both invalid preview conditions (coded as .5). The second contrast directly compared the invalid word preview condition (coded as -.5) to the random letter string preview condition (coded as .5). For many of the models it was necessary to simplify the random effects structure to obtain convergence, so we report models with this simplified structure in all cases. These models include random intercepts for subjects and items, and random slopes for the predictability factor for the both subjects and items. Random effect correlation parameters have been removed.

Measure	Effect	Estimate	SE	*t/z*-value
Skipping Probability	Predictability	-0.07	-.10	-0.69
	Preview Validity	-0.06	-.08	-0.75
	Invalid Preview Type	-0.01	0.10	-0.07
	Pred x Prev Validity	-0.08	0.16	-0.50
	Pred x Invalid Prev Type	-0.07	0.19	-0.38
First Fixation Duration	Predictability	2.03	3.59	0.56
	Preview Validity	7.29	3.57	**2.04**
	Invalid Preview Type	4.12	4.11	1.00
	Pred x Prev Validity	9.73	7.12	1.37
	Pred x Invalid Prev Type	12.94	8.22	1.57
Gaze Duration	Predictability	1.57	4.59	0.34
	Preview Validity	9.11	4.51	**2.02**
	Invalid Preview Type	10.29	5.20	1.98
	Pred x Prev Validity	10.29	9.01	1.51
	Pred x Invalid Prev Type	11.18	10.39	1.08
Go-Past Time	Predictability	4.06	10.04	0.41
	Preview Validity	14.86	7.66	1.94
	Invalid Preview Type	17.99	8.84	**2.04**
	Pred x Prev Validity	7.38	15.28	0.48
	Pred x Invalid Prev Type	11.91	17.64	0.68
Regression Probability	Predictability	0.22	0.19	1.13
	Preview Validity	0.04	0.16	0.28
	Invalid Preview Type	0.01	0.19	0.06
	Pred x Prev Validity	-0.14	0.32	-0.43
	Pred x Invalid Prev Type	0.10	0.37	0.27

Table 5. Experiment 2 statistical results from mixed-effects models for the pre-target word, with all factors included. Model specification is described in the text.

Measure	Effect	Estimate	SE	*t/z*-value
Skipping Probability	Predictability	-0.15	-.09	-1.69
	Preview Validity	-0.52	-.09	**-5.87**
	Invalid Preview Type	-0.43	0.11	**-3.81**
	Pred x Prev Validity	0.27	0.18	1.50
	Pred x Invalid Prev Type	-0.10	0.22	-0.44
First Fixation Duration	Predictability	8.74	3.34	**2.62**
	Preview Validity	41.94	3.34	**12.56**
	Invalid Preview Type	2.15	3.73	0.58
	Pred x Prev Validity	-10.66	6.67	-1.60
	Pred x Invalid Prev Type	4.43	7.47	0.59
Gaze Duration	Predictability	12.22	4.61	**2.65**
	Preview Validity	55.14	4.08	**13.52**
	Invalid Preview Type	14.27	4.56	**3.13**
	Pred x Prev Validity	-13.85	8.15	-1.70
	Pred x Invalid Prev Type	-6.23	9.12	-0.68
Go-Past Time	Predictability	38.24	11.03	**3.47**
	Preview Validity	86.81	8.83	**9.83**
	Invalid Preview Type	32.26	9.87	**3.27**
	Pred x Prev Validity	-10.54	17.66	-0.60
	Pred x Invalid Prev Type	1.65	19.75	0.08
Regression Probability	Predictability	0.38	0.13	**2.89**
	Preview Validity	0.64	0.12	**5.33**
	Invalid Preview Type	0.17	0.12	1.41
	Pred x Prev Validity	-0.33	0.24	-1.36
	Pred x Invalid Prev Type	0.18	0.24	0.77

Table 6. Experiment 2 statistical results from mixed-effects models for the target word, with all factors included. Model specification is described in the text.

For the pre-target word, there was a significant effect of preview validity on first fixation and gaze duration, and a significant effect of invalid preview type on go-past time. The apparent effect of predictability on regressions from the pre-target word does not approach significance.

In the statistical model of skipping of the target word, all three main effects were significant, but the interaction that is apparent in Figure 3 did not reach significance. The three reading time measures and the regressions out measure also show significant effects of both predictability and preview validity, and again the interactions between these factors did not reach significance. The effect of invalid preview type was significant in gaze duration and go-past time.

As for Experiment 1, the predictions that motivated this experiment are not directly tested by the tests of interaction effects. An effect of predictability should be present in the valid preview conditions and absent in the invalid preview conditions, and this null effect of predictability in the invalid preview conditions should not depend on invalid preview type. To directly test these predictions, we computed separate models for the valid preview and invalid preview conditions, shown in Table 7. Fixed effects in these models were centered, with the factor levels coded as -.5 and .5. The random effects structure in these models was maximal for the linear models (i.e., random intercepts for subjects and items, as well as random subject and item slopes for each fixed effect and their interaction). For the logistic models of the invalid preview condition, the random slopes for the interaction effect had to be removed to obtain convergence.

Measure	Effect	Estimate	SE	*t/z*-value
Valid Preview Conditions				
Skipping Probability	Predictability	-0.39	0.17	**-2.34**
First Fixation Duration	Predictability	12.12	5.13	**2.36**
Gaze Duration	Predictability	16.94	6.00	**2.83**
Go-Past Time	Predictability	40.02	16.32	**2.45**
Regression Probability	Predictability	0.20	0.39	0.52
Invalid Preview Conditions				
Skipping Probability	Predictability	0.03	0.14	0.24
	Preview Type	-0.50	0.14	**-3.63**
	Pred x Preview	-0.01	0.23	-0.04
First Fixation Duration	Predictability	3.80	4.31	0.88
	Preview Type	2.66	4.29	0.62
	Pred x Preview	4.72	8.02	0.59
Gaze Duration	Predictability	5.61	5.73	0.98
	Preview Type	15.01	5.42	**2.77**
	Pred x Preview	-6.02	10.55	-0.57
Go-Past Time	Predictability	33.72	12.78	**2.64**
	Preview Type	32.57	15.03	**2.17**
	Pred x Preview	1.82	23.83	0.08
Regression Probability	Predictability	0.17	0.15	1.12
	Preview Type	0.16	0.16	1.03
	Pred x Preview	0.22	0.25	0.85

Table 7. Experiment 2 statistical results from separate mixed-effects models for the target word for the valid and invalid preview conditions. Model specification is described in the text.

As expected, predictability of the target word in the valid preview conditions affected the probability that the target was skipped, as well as all three reading time measures. The trend in the regressions measure was toward more regressions from an unpredictable word, but this effect did not reach significance. With invalid preview, predictability did not affect skipping, confirming that predictability-based skipping requires parafoveal preview of the predictable word (Balota et al., 1985), rather than simply resulting from a guess as to the next word's identity in a constraining context. Consistent with previous results, skipping of a random letter string preview was less likely than skipping of an unexpected word preview (e.g., Choi & Gordon, 2013; Drieghe, Rayner, & Pollatsek, 2005). In first fixation duration, there was no evidence of either a predictability effect or an effect of invalid preview type. In gaze duration there was still no predictability effect, but there was a preview type effect; gaze duration was significantly longer with random letter string preview that with unrelated word preview. Both effects were significant in go-past time; while there was little evidence of a predictability effect on first fixation and gaze duration with invalid preview, predictability had a significant 34 ms effect on go-past time with invalid preview. While neither manipulation significantly affected the probability of a regression, the trends were toward more regressions when the target was unpredictable, and when there was a random letter string preview. The interaction effect did not approach significance for any measure.

Finally, we performed Bayes Factor analyses as follows. First, we computed the Bayes Factor for a model of first fixation and gaze duration in the invalid preview conditions, including only random subject and item intercepts; this provides the relative marginal likelihood of the data under this model compared to a null model that does not assume even subject and item variability. We then computed the Bayes Factor for a model that includes both random subject and item intercepts and a fixed effect of predictability, compared to the null model. The critical value is the ratio of these two Bayes Factors: It is itself a Bayes Factor comparing the model with an effect of predictability and subject and item intercepts, to a model with only subject and item intercepts. This critical Bayes Factor was 12.46 in favor of a null effect of predictability on first fixation, and 10.17 in favor of a null effect of predictability on gaze duration. As expected, the same Bayes Factor analysis of the valid preview conditions favored the model that includes the predictability effect over the model that does not, by a factor of 3.02 for first fixation duration and 8.09 for gaze duration.

Discussion

This experiment again replicated the standard predictability effects that occur in normal reading, i.e., with valid preview. A predictable target was more likely to be skipped, and all three measures of first pass reading time were shorter; the effect of predictability was about 12 ms in first fixation duration, 17 ms in gaze duration, and 40 ms in go-past time.

The critical questions that motivated this experiment were about the invalid preview conditions. The experiment replicated the lack of predictability effect on both first fixation and gaze duration with invalid preview. Linear mixed effects models did not find evidence for such an effect, and a Bayes Factor analysis found evidence for the lack of such an effect. However, there was indeed an effect of predictability on go-past time in the invalid preview conditions.

While the predictability-by-preview interaction did not reach significance in the full model, there is clear statistical support for predictability effects on both first fixation and gaze duration with valid preview, and clear evidence against such effects with invalid preview. We assume that the failure of the interaction to reach significance is simply an issue of power. Note that while Experiment 1 had 72 subjects, Experiment 2 had only 47. In any event, the interactive trend in the same direction as in previous studies (i.e., previous studies in the literature, and Experiment 1 of this study) should increase confidence in this effect (e.g., Francis, 2012).

There was no indication that invalid preview type (unrelated word vs. random letter string) modulated the effect of predictability on any measure. However, preview type itself did have an effect on both gaze duration and go-past time, with longer reading times following random letter string previews. This result is consistent with findings (e.g., Risse & Kliegl, 2014) suggesting that the difficulty of processing the preview string itself may appear at a delay, in reading time measures on the target word.

In sum, the present experiment demonstrates that the null effect of predictability on first fixation and gaze duration with invalid preview does not depend on the lexical status of the invalid preview. This result provides evidence against a lexical suppression hypothesis, according to which the lack of predictability effect is due to the parafoveal activation of an unexpected word. Some other explanation of the lack of predictability effect with invalid preview is required.

General Discussion

The goal of these experiments was to directly investigate a pattern that has emerged incidentally across several previous studies, whereby the predictability effect on early reading time measures is eliminated with invalid parafoveal preview, but the frequency effect is not. Experiment 1 replicated this dissociation in a single experiment. The dissociation between the two variables argues against a hypothesis holding that invalid preview eliminates lexical influences on early reading time measures due to the operation of a 'time out' mechanism (e.g., Henderson & Ferreira, 1990). This hypothesis would suggest that because invalid preview dramatically lengthens initial fixations on a target word, these fixations are not terminated in the usual way, i.e., by completion of a stage of lexical processing. The fact that there are frequency effects on these measures even with invalid preview shows that such an explanation does not suffice. Experiment 2 revealed that the lack of predictability effect with invalid preview does not depend on the lexical status of the preview, as the predictability effect on first fixation and gaze duration was absent with both unrelated word previews and with random letter string previews. This argues against a lexical suppression hypothesis holding that the predictability effect is eliminated only when an unexpected word is activated parafoveally, suppressing readers' contextually based expectations for the next word.

The present data cannot entirely rule out a hypothesis emphasizing suppression of expectations by invalid parafoveal preview, however. It is possible that the reader's contextually-based expectations are suppressed by invalid preview regardless of whether this preview is word-like in its orthography; perhaps the reader abandons her expectations when discrepant parafoveal evidence is encountered, even if this parafoveal input does not activate any particular alternative word. This account would predict that when parafoveal preview is simply absent, as opposed to invalid, predictability effects on reading times on the target word should re-appear. A recent study by Parker et al. (2017) has tested this idea. This study compared the predictability effect when the target word appears at the start of a second line of text, in which case there is no preview during the previous fixation at the end of the first line, and when it appears mid-line. Parker et al. found a 43 ms effect of predictability on gaze duration when the target word appears at the beginning of the second line. They suggested that the lack of predictability effect with invalid preview, in previous studies, should therefore be understood as reflecting the invalid preview's suppression of current expectations.

Though the manipulation in the Parker et al. (2017) study is interesting and the result is suggestive, further research is needed. In the control conditions, in which the target word was presented in the center of a line, with fully valid preview, the predictability effect on gaze duration was 1 ms, i.e., essentially absent altogether. We do not know of other studies that have failed to replicate this standard predictability effect, so the failure of this experiment to do so suggests caution in interpreting its other results. Moreover, Parker et al. (2017) did not find an effect of predictability on first fixation duration (reported in their supplementary materials) in the beginning-of-line condition, making the overall pattern more difficult to ascertain.

Here we offer an explanation of the critical patterns by means of a modification to a Bayesian account of the influence of both predictability and frequency on visual word recognition (Norris, 2006; Smith & Levy, 2013). We attempt to explain why denying parafoveal preview eliminates the effect of predictability on early reading time measures, but does not eliminate the effect of word frequency. This account endorses the idea that the two variables influence lexical processing, and eye movements in reading, by somewhat different means, contrary to the assumptions of the E-Z Reader (Reichle et al., 2003) model. We note, at the outset, that our account is tentative; we regard it as the best current explanation of a puzzling and intriguing empirical pattern.

Our proposal is in the spirit of Norris' (2006) Bayesian Reader model, though departs from it in some details. Norris' model offers a principled answer to the question of why word frequency influences word recognition time. The model's Bayesian conception of the process of word recognition assumes that recognition occurs when the posterior probability that the visual input corresponds to a particular word (i.e., $p(word|visual\ input)$) reaches a criterion level. This posterior probability is a function of two things: the word's prior probability, and the likelihood. Word frequency provides a prior probability for each word in the reader's lexicon. The likelihood is the probability of the perceptual evidence given that the letter string is, in fact, a particular hypothesized word. Because of the influence of a frequency-based prior, the likelihood need not be as high in order for a high frequency word to be recognized (i.e., for the posterior to reach the criterion level) as would be required for a low frequency word. This means that recognition will typically take place faster for a high frequency word.

While Norris' (2006) Bayesian Reader is motivated by the need to explain frequency effects, it is explicit that the model is also intended to provide a natural explanation of predictability effects (see also Smith & Levy, 2013). Like word frequency, contextually-based expectations may be thought of as providing a prior probability distribution over upcoming words.

We propose that the phenomena presently under discussion can be accounted for by means of two modifications to this model. The first involves emphasizing a feature that is already present in the Bayesian calculation: The influence of the prior will be strong when perceptual evidence is weak, and weak when perceptual evidence is strong. Bayesian calculation tells us that when evidence is unequivocal (i.e., $p(visual\ input|word_{target})$ is very high, and $p(visual\ input|word_{other})$ is very low), the prior probability of the hypothesis ($p(word_{target})$) will have little influence on the posterior probability. However, when perceptual evidence is equivocal, the prior may have a substantial influence. Norris (2006) does indeed emphasize that the prior will only have an influence when there is some ambiguity either in the stimulus itself, or because the participant must respond rapidly, before all available information can be processed.

Early orthographic processing of a word is typically carried out during parafoveal viewing, before the eyes have directly fixated the word (e.g., Balota et al., 1985; Drieghe et al., 2005; White, Johnson, Liversedge, & Rayner, 2008). When parafoveal preview is denied, however, early orthographic processing of the target word must be carried out in foveal vision. When a word is viewed in foveal vision, where acuity is highest, the perceptual evidence will be relatively unequivocal, compared to when a word is viewed in parafoveal vision, where acuity is lower. Thus, the influence of the prior may be much weaker. We propose that when early orthographic processing is carried out on a word that is already located in foveal vision, the perceptual evidence is simply too strong for the prior distribution to have much effect.

It may be argued that there is a sense in which, in the invalid preview conditions, perceptual evidence as to the identity of the target word is not unequivocal. As noted above, there is evidence that invalid preview does not simply delay processing, as properties of a preview word itself influence reading time on the target (e.g., Risse & Kliegl, 2014; Veldre & Andrews, 2016, 2017, 2018). Even in the present Experiment 2, we see effects of preview type on target word reading times, with longer reading times when the preview was a random letter string. Thus, one may regard evidence about the identity of the target word as coming from a combination of the preview and target strings, and therefore, as inherently ambiguous in conditions in which these two strings are different. However, what is critical for the present account is the assumption that the visual input from the *actual* target string – i.e., the string whose predictability has been manipulated - is clearer when this string is present in foveal vision than it is when this string is in the parafovea. This assumption is built into E-Z Reader (Reichle et al., 2003), for example, in the form of a function that discounts the rate of early lexical processing based on the degree of eccentricity at which the word is viewed.

On our preferred Bayesian account, why does the word frequency effect survive with invalid preview? This question brings us to our second proposed modification to Norris' (2006) Bayesian model. We assume that, unlike the effect of predictability, the effect of word frequency may be due only in part to the influence of a Bayesian prior on early stages of orthographic processing. Frequency may not only affect retrieval of an orthographic word form (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), but may also influence later stages of lexical access, such as retrieval of the word's meaning. One way of making this distinction is proposed by the E-Z Reader model (Reichle et al., 2003), which distinguishes a lexical 'familiarity check' (the L1 stage in the model) from full lexical access (the L2 stage). At present, E-Z Reader assumes that both predictability and frequency influence both the L1 and L2 stages, but it is consistent with the theoretical description of these stages that predictability might influence only the former, while frequency influences both.

Critically, any effect of word frequency on a late stage of lexical processing would not depend on whether early processing is carried out parafoveally or foveally. Indeed, while E-Z Reader proposes that, as just noted, the duration of the L1 (familiarity check) stage is modulated by visual acuity, the model proposes that the duration of the L2 (access) stage is not; Reichle (2011) remarks that this stage "is not affected by visual acuity because the information being accessed is semantic, not visual" (p. 774).

It is consistent with this account that there should be some attenuation of the frequency effect by invalid preview, as the frequency effect is also, to some extent, due to the influence of a Bayesian prior. It is also predicted that this attenuation should occur primarily in the earliest measures. Though frequency clearly has an effect on gaze duration that is larger than its effect on first fixation duration, the present study and Reingold et al. (2012) converge on the conclusion that this additional, later effect of frequency is not at all attenuated by invalid preview.[4]

In sum, we propose that the predictability effect is eliminated by invalid preview because this effect requires the presence of ambiguity in the perceptual evidence that is available during early orthographic processing. When early processing of the target takes place foveally, the perceptual evidence is simply too clear for predictability to have a measurable effect. We propose that while the effect of word frequency may also arise partially by means of a prior probability distribution over orthographic word forms, this variable also affects later processing stages that are not dependent on visual acuity.

Additional support for a distinction between the mechanisms by which predictability and frequency influence early eye movement measures has come from

---

[4] It is clear from Table 2 that this pattern is not consistent across all studies. However, it is again worth noting the relative power of these studies. With 60 subjects x 60 observations per condition, each cell mean in the Reingold et al. (2012) study reflects 3600 observations (prior to any data loss). But in the Inhoff and Rayner (1986) study, for example, 24 subjects each read 20 trials with valid preview at each level of word frequency, and only 10 trials with invalid preview, for a total of 240 observations in each of the invalid preview conditions. The present Experiment 1 had 72 subjects, with 20 trials at each level of frequency and preview, for a total of 1440 observations in each condition, making it the second most powerful of the studies to have investigated the interaction of frequency and preview validity.

fitting of the ex-Gaussian distribution (Ratcliff, 1979) to individual subjects' fixation duration distributions. The effect of predictability on first fixation duration is distributionally similar to the effect of stimulus quality (i.e., visual contrast), with manipulations of both variables resulting in a shift in the location of the distribution (the μ parameter of the ex-Gaussian distribution), with little or no change in the weight of the distribution's right tail (the τ parameter; Sheridan & Reingold, 2012; Staub, 2011; Staub & Benatar, 2013; White & Staub, 2012). By contrast, the effect of frequency on first fixation duration is due to reliable effects on both parameters (Reingold et al., 2012; Staub et al., 2010). The fixation duration distribution is shifted to the right for low frequency words, but the right tail is also more pronounced for low frequency words than for high frequency words, i.e., there are more very long fixations. Notably, the same distributional effects of frequency are obtained in single word recognition tasks (Andrews & Heathcote, 2001; Balota & Speiler, 1999), while the shift-only effect of predictability patterns like an effect of semantic priming in single word tasks (Balota, Yap, Cortese, & Watson, 2008).

An intriguing question is whether these two differences between the effects of lexical predictability and frequency – only the frequency effect survives with invalid preview, and only the frequency effect is manifested in a specific effect on the right tail of the fixation duration distribution – are, in fact, related. If word frequency's effect on the tail of the distribution does reflect the operation of a distinct, later process, as tentatively suggested by Staub and Benatar (2013), it might be expected that it is specifically word frequency's effect on the right tail of the distribution that survives with invalid preview. The Reingold et al. (2012) study, which fit the ex-Gaussian distribution to each subject's first fixation duration data in each condition, provides a striking confirmation of this prediction. With valid preview, the frequency effect on first fixation duration was almost equally partitioned into effects on the μ (9 ms) and τ (11 ms) parameters. Both effects were significant. The overall reduction in the frequency effect that occurred with invalid preview was not equally distributed across the effects on the two distributional parameters, however. The τ effect actually increased in size, to 16 ms, while the μ effect was entirely eliminated; indeed, there was a non-significant reversal (-8 ms). Thus, the results of Reingold et al. suggest that while denying parafoveal preview decreases the overall size of the frequency effect on mean first fixation duration, this reduction is distributionally selective. With invalid preview, the first fixation duration distribution for low frequency words is substantially more skewed than for high frequency words, but it is no longer shifted to the right.

As we have seen, the predictability effect on mean first fixation duration is essentially eliminated with invalid parafoveal preview. One possibility, then, is that denying parafoveal preview eliminates the distributional shifting that is common to low frequency and low predictability words; this would eliminate the entire predictability effect, while preserving the effect on the right tail of the distribution that is unique to low frequency words. Testing this conjecture will be a goal of future research. We cannot test it with the present data, as we did not collect a sufficient number of observations in either experiment to obtain reliable ex-Gaussian fits (Heathcote, Brown, & Mewhort, 2002).

As we have noted, the empirical dissociation between the effects of predictability and frequency with invalid preview is not predicted by the E-Z Reader model (Reichle et al., 2003), nor does our account of it square with E-Z Reader's assumption that the two variables influence the same stages of lexical processing. The SWIFT model (Engbert, Nuthmann, Richter, & Kliegl, 2005) distinguishes between the effects of the two variables in a way that may be seen as more similar to the current proposal. In the SWIFT model, predictability influences lexical processing earlier than frequency does, by influencing the rate of processing both during a 'preprocessing' stage, much of which takes place parafoveally, as well as during a 'lexical completion' stage. Word frequency influences the duration of the lexical completion stage by influencing the difficulty of recognizing a word, i.e., the amount of activation that must be accrued.

However, we do not endorse SWIFT's architecture as a way of capturing predictability effects. This is because of a logical problem that has been pointed out by Slattery, Pollatsek, and Rayner (2007; see Reichle, Liversedge, Pollatsek, & Rayner, 2009, for related arguments). The problem relates to what is arguably the core difference between SWIFT and E-Z Reader: SWIFT assumes parallel lexical processing of multiple words, while E-Z Reader assumes that though visual processing may be carried out in parallel across multiple words, lexical processing is serial, with the currently fixated word being fully identified prior to the initiation of lexical access for the next, parafoveal, word. The problem is as follows. The empirically determined cloze probability for word *n* is based on the entire preceding sentence context, up through word *n*-1. Often, the identity of word *n*-1 will have a dramatic effect on the cloze probability of word *n*. Moreover, the effect of the cloze probability of word *n* on eye movements appears to be the same whether it is word *n*-1, or some earlier portion of the discourse, that is responsible for that cloze probability (Fitzsimmons & Drieghe, 2013). However, in the

SWIFT model, the cloze probability of word *n* affects processing of that word well before processing of word *n*-1 is complete; indeed, it affects processing of word *n* even while processing of word *n*-2 is still ongoing. Logically this cannot be correct, as a variable cannot influence lexical processing if the value of that variable is not set at the time of its putative influence.

In sum, we do not think that either E-Z Reader or SWIFT can account for the predictability-related phenomena under discussion. E-Z Reader does not predict that lack of parafoveal preview should eliminate the predictability effect, or for that matter, that there should be any dissociations at all between the effects of predictability and frequency. While SWIFT predicts that such dissociations are possible, and does propose that the effect of predictability should be especially pronounced during parafoveal processing, its parallel architecture creates a logical problem in explaining how predictability, as measured by cloze probability, has an effect when it does.

Finally, it is also worth considering why, in the late measures of regression probability and go-past time, some effect of predictability might remain even with invalid preview. An effect of predictability on go-past time was significant in the invalid preview conditions of Experiment 2, and there was a similar trend in the invalid preview conditions of Experiment 1. We assume that interword regressions, which contribute to go-past time, reflect truly post-lexical processes of syntactic and semantic integration of a word with its context (Reichle et al., 2009). The effect of predictability on regressions, which has appeared in multiple experiments (e.g., Frisson et al., 2017; Staub, 2011; Kretzschmar et al., 2015) may reflect the relative difficulty of integrating a low predictability target word with its sentence context. If so, there is no reason that invalid preview should inhibit such regressions.

Conclusion

Two experiments present evidence that while invalid preview eliminates the effect of predictability on early reading time measures on a target word, it does not eliminate the effect of frequency. A Bayesian account of this pattern suggests that predictability influences the prior probability that a reader assigns to an upcoming word, but the influence of this prior is minimal, or even nonexistent, when all processing is carried out on a foveal stimulus where the perceptual evidence is very clear. The dissociation between frequency and predictability suggests that frequency also influences late stages of lexical access, consistent with distributional evidence for distinct influences of the two variables. The present findings may lead to more finely articulated models of how these variables influence lexical processing in reading.

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.

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task- specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 514–544.

Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic*

Bulletin & Review, 23, 1241-1249.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364–388.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*, 32–55.

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59*, 495–523.

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

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*, 665-670.

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

Choi, W., & Gordon, P. C. (2013). Coordination of word recognition and oculomotor control during reading: the role of implicit lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 1032-1046.

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and aging, 32*, 232-242.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

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.

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

Fitzsimmons, G., & Drieghe, D. (2013). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1054-1063.

Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, *7*, 585-594.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200-214.

Gordon, P. C., Plummer, P., & Choi, W. (2013). See before you jump: full recognition of parafoveal words precedes skips during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 633-641.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review, 9*, 394–401.

Henderson, J.M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 417–429.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Perception & Psychophysics, 40*, 431–439.

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1560–1579.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 68–81.

Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from co-registration of eye movements and

EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1648-1662.

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

Morey, R. D., Rouder, J. N., & Jamil, X. (2015). BayesFactor: Computation of Bayes factors for common designs. Version 0.9.11. <http:// bayesfactorpcl.r-forge.r-project.org/>.

Norris, D. (2006). The Bayesian reader: Exploring word recognition as an optimal Bayesian decision process. *Psychological Review, 113*, 327– 357.

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

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446–461.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *16*, 65-81.

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

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model.

Journal of Experimental Psychology: Human Perception and Performance, 30, 720–732.

Reichle, E. D. (2011). Serial-attention models of reading. In: S.P. Liversedge, Iain D. Gilchrist and Stefan Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 767-786). Oxford, UK: Oxford University Press.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13*, 115-119.

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., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain sciences*, *26*, 445-476.

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

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*, 177–206.

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*, 653-688.

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

Schotter, E. R., Lee, M., Reiderman, M., & Rayner. K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language, 83*, 118–139.

Sereno, S. C., Hand, C. J., Shahid, A., Yao, B., & O'Donnell, P. J. (2018). Testing the limits of contextual constraint: Interactions with word frequency and parafoveal preview during fluent reading. *The Quarterly Journal of Experimental Psychology, 71*, 302-313.

Sheridan, H., & Reingold, E. M. (2012). The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations. *Visual Cognition*, *20*, 733–745.

Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, *35*, 1283-1292.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302-319.

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review, 18*, 371–376.

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

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review*, *20*, 1304-1311.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1280-1293.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*, 1176-1190.

Veldre, A., & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 939 - 962.

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

White, S. J., & Staub, A. (2012). The distribution of fixation durations during reading: Effects of stimulus quality. *Journal of Experimental Psychology: Human Perception & Performance*, *38*, 603-617.

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1261–1276.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading.

Psychonomic Bulletin & Review, 12, 466–471.

Author Note

Audiences at the 57th Annual Meeting of the Psychonomic Society and the UMass Psycholinguistics Workshop provided helpful comments on parts of this work, and Chuck Clifton provided valuable feedback on a draft of the manuscript. Wesley Albright, Atreyi Mukherji, Sinthema Roy, Audrey O'Neill, Laurel Whitfield, Sophia Dodge, Daniel Godwin, and Owen Jordan provided assistance with data collection. This work was supported by a grant from the National Science Foundation (BCS 1732008) to the first author.