Journal of Memory and Language 79–80 (2015) 76–96

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data

Bernhard Angele [a,*], Elizabeth R. Schotter [b], Timothy J. Slattery [c], Tara L. Tenenbaum [b], Klinton Bicknell [d], Keith Rayner [b]

[a] Psychology Research Centre, Faculty of Science and Technology, Bournemouth University, United Kingdom
[b] Department of Psychology, University of California San Diego, United States
[c] Department of Psychology, University of South Alabama, United States
[d] Department of Linguistics, Northwestern University, United States

ARTICLE INFO

Article history:
Received 29 April 2014
Revision received 26 September 2014
Available online 9 January 2015

Keywords:
Eye movements
Reading
Parafoveal-on-foveal effects
Lexical processing
Corpus analysis

ABSTRACT

In the past, most research on eye movements during reading involved a limited number of subjects reading sentences with specific experimental manipulations on target words. Such experiments usually only analyzed eye-movements measures on and around the target word. Recently, some researchers have started collecting larger data sets involving large and diverse groups of subjects reading large numbers of sentences, enabling them to consider a larger number of influences and study larger and more representative subject groups. In such corpus studies, most of the words in a sentence are analyzed. The complexity of the design of corpus studies and the many potentially uncontrolled influences in such studies pose new issues concerning the analysis methods and interpretability of the data. In particular, several corpus studies of reading have found an effect of successor word ($n + 1$) frequency on current word (n) fixation times, while studies employing experimental manipulations tend not to. The general interpretation of corpus studies suggests that readers obtain parafoveal lexical information from the upcoming word before they have finished identifying the current word, while the experimental manipulations shed doubt on this claim. In the present study, we combined a corpus analysis approach with an experimental manipulation (i.e., a parafoveal modification of the moving mask technique, Rayner & Bertera, 1979), so that, either (a) word $n + 1$, (b) word $n + 2$, (c) both words, or (d) neither word was masked. We found that denying preview for either or both parafoveal words increased average fixation times. Furthermore, we found successor effects similar to those reported in the corpus studies. Importantly, these successor effects were found even when the parafoveal word was masked, suggesting that apparent successor frequency effects may be due to causes that are unrelated to lexical parafoveal preprocessing. We discuss the implications of this finding both for parallel and serial accounts of word identification and for the interpretability of large correlational studies of word identification in reading in general.

Introduction

One of the major debates in reading research concerns the extent to which upcoming words can be processed before they are fixated (i.e., what is the extent of parafoveal

* Corresponding author at: Bournemouth University, Psychology Research Centre, Faculty of Science and Technology, Talbot Campus, Fern Barrow, Poole BH12 5BB, United Kingdom.
E-mail address: bangele@bournemouth.ac.uk (B. Angele).

http://dx.doi.org/10.1016/j.jml.2014.11.003
0749-596X/

preprocessing?). This question closely aligns with the issue of how many words a reader can process in parallel. Since only one word can be fixated at a time, and since there is only limited evidence that readers keep processing words after they have moved their gaze away from them (Binder, Pollatsek, & Rayner, 1999; Rayner, Well, & Pollatsek, 1980), readers processing multiple words at once must be engaging in parafoveal processing. Thus far in the literature, there have been two general approaches to answering this question: experimental manipulations and large corpus correlational techniques.

Evidence that readers are able to process parafoveal words at all was shown by McConkie and Rayner (1975; for a recent review of their research see Rayner, 2014), and since then a number of studies have converged to estimate that the area from which readers can obtain useful visual information (the *perceptual span*) extends up to 14–15 letter spaces to the right of fixation (usually including the current and next word). A different paradigm, the gaze-contingent boundary paradigm, which was introduced by Rayner (1975), provides insight into which properties of an upcoming word can be pre-processed. In this paradigm, unbeknownst to the reader, an invisible boundary is placed to the left of a target word of interest, which remains masked before the boundary is crossed. After the boundary is crossed, the display changes to reveal the actual target word. Subjects are usually not aware of this experimental manipulation. By varying how similar the mask is to the target word, researchers can infer which properties of the target word can be processed parafoveally; previews that are more similar to the target lead to faster reading time once the target is fixated (i.e., they yield preview benefit; for reviews, see Schotter, 2013; Schotter, Angele, & Rayner, 2012). In contrast to these experimental approaches, parafoveal processing is assessed in corpus analyses by entering properties of the upcoming word into a statistical model; if properties of the upcoming word account for variance in first-pass reading time on the current word, researchers infer that the reader was processing the upcoming word before fixating it (i.e., in parallel with processing of the current word).

The different accounts of parafoveal and serial/parallel processing are best summarized in the context of current computational models of eye movement control in reading. These models can be divided in two groups: Serial attention shift (SAS) models assume that attention can only be allocated to one word at a time. Usually, this means that attention (i.e., lexical processing) is initially allocated to the currently fixated word and then shifted to upcoming parafoveal words while the language processing system is waiting for the oculomotor system to plan and execute a saccade. The most prominent representative of SAS models is the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006; Reichle, Warren, & McConnell, 2009). In contrast to SAS models, processing gradient (PG) models assume that, during normal reading, attention can be spread over multiple words in a sentence, with processing speed being determined by the distance of each letter from the center of fixation (i.e., by its eccentricity). As a consequence, PG models predict that readers should be frequently engaging in lexical parafoveal processing of several upcoming words (although recent models have placed some limitations on which words can be processed in a given situation, e.g. Schad & Engbert, 2012). Prominent examples of PG models are the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Schad & Engbert, 2012) and the Glenmore model (Reilly & Radach, 2006).

While these two classes of models are similar in many respects (and consequently make similar predictions for most of the benchmark effects in reading), the detailed implementations of both the serial and parallel accounts of word identification in reading have stimulated a great deal of research aimed at testing their divergent predictions (for thorough reviews of this research, see Rayner, 1998, 2009a; Schotter et al., 2012). Here, we will focus on research addressing the most important difference between the models' predictions: parafoveal-on-foveal (PoF) effects. PoF effects are defined as effects of the linguistic properties (e.g., word frequency) of the upcoming word (word $n + 1$) on the ongoing processing of the currently fixated word (word n) as reflected by eye movement measures such as fixation time and, to a lesser extent, fixation probability. PoF effects are similar to but theoretically distinct from successor effects (although they have often been discussed similarly; see further discussion below). Since PG models assume that parafoveal words are constantly being processed (until completion), while SAS models predict parafoveal processing only after saccade programming away from word n has already begun, PoF effects are considered to be more compatible with PG models than with SAS models. Importantly, despite this generally accepted dichotomy, neither SWIFT nor E-Z Reader currently implement a mechanism that would allow parafoveal input to have an influence on the duration of the ongoing fixation.[1] Still, it could be argued that such a mechanism would be easier to implement in SWIFT than in E-Z Reader. We now turn to an important caveat regarding the debate surrounding parafoveal processing of words: the experimental methods and statistical approaches used to test for its presence.

The difference between PoF and successor effects

The difference between correlational and the experimental approaches can be described as follows: in the experimental approach, the variables of interest are controlled or manipulated a priori in the experimental design (e.g. by holding word length constant or varying it across conditions) whereas in correlational approaches the variable of interest is investigated post hoc by entering the word's property into the statistical analysis (e.g., by entering word length as a predictor variable in a regression model). In practice, one could argue most studies of reading include some degree of both approaches by manipulating some variables a priori while entering others into the

[1] SWIFT may allow an effect of parafoveal processing on refixation probability and thereby gaze duration. Additionally, first fixation duration and single fixation duration may be influenced to some degree by changes in the saccade-target selection (Risse, Engbert, & Kliegl, 2008; Schad & Engbert, 2012).

analysis, post hoc. The motivation for the present article concerns the most important differences between correlational and experimental studies: (1) the interpretability of the results, an issue closely connected to the degree of experimental control and (2) the ability to detect subtle relationships between covariates, which is closely tied to statistical control achieved by increasing model complexity.

In terms of PoF effects, in predominantly experimental studies, the sentence context preceding the target word is usually held constant, meaning that a difference between the fixation times associated with a variable of interest (e.g., word frequency of word $n+1$) can be directly interpreted as the consequence of the parafoveal information being processed – a PoF effect. However, such a study might miss PoF effects that only occur in in circumstances that are not included in the experimental design (e.g. a PoF effect that only occurs on short words when the experimental stimuli only contain long target words). In this case, the study will fail to detect PoF effects simply because it was not designed to test for them. In contrast, a predominantly correlational study may have more opportunity to observe such effects (due to inclusion of a broader range of linguistic stimuli), but statistical control poses its own challenges. In the case of PoF effects, it might not be clear whether an apparent difference between fixation times on words preceding high frequency and low frequency parafoveal words is actually due to parafoveal processing or due to another covariate (e.g., subtle differences between the sentence contexts preceding high vs. low frequency words; see Rayner, Pollatsek, Drieghe, Slattery, & Reichle, 2007; cf. Kliegl, 2007). As a consequence, $n+1$ frequency effects found in correlational corpus studies should not be called PoF effects (which implies that they are caused by parafoveal processing) but rather *successor effects* (Kliegl, Nuthmann, & Engbert, 2006), a more descriptive and less interpretative term that does not suggest a cause. In other words, successor effects may well be PoF effects, but they should not be interpreted as such if there are other explanations for the observed correlations (e.g., differences in the preceding context).

As we discuss the evidence for lexical parafoveal processing, it is important to keep the distinction between PoF effects and successor effects in mind. There are not many experimental studies showing lexical PoF effects (for a review, see Schotter et al., 2012). The strongest experimental evidence for PoF effects comes from studies that used a task other than natural sentence reading (like the ones described at the beginning of the introduction), calling into question whether the results would generalize to natural reading. Specifically, studies that report evidence for PoF effects have presented single words with parafoveal flankers (Abad, Noguera, & Ortells, 2003; Bradshaw, 1974; Ortells, Abad, Noguera, & Lupiáñez, 2001; Ortells & Tudela, 1996) and had subjects perform a lexical decision, naming, or identification task. However, when eye movements were monitored, and when only cases in which subjects did not fixate the flanker words were included in the analyses, no such effects were observed (Inhoff, 1982; Inhoff & Rayner, 1980). A similar experimental approach requires subjects to search for words from a certain category (e.g., items of clothing) within a string of unrelated words. In this task lexical PoF effects were also found (Kennedy, 1998; Kennedy, Pynte, & Ducrot, 2002; Schroyens, Vitu, Brysbaert, & d' Ydewalle, 1999). Overall, the evidence for the existence of PoF effects during sentence reading is quite mixed (see below) and depends on the level of word representation being considered.

Orthographic PoF effects

It seems clear that the presence of unusual parafoveal information such as infrequent letter combinations (e.g. *dw* or *xy* in English) can have an effect on fixation times in reading (Pynte, Kennedy, & Ducrot, 2004; White, 2008). Such *orthographic PoF* effects are also frequently found on the pre-boundary word in gaze-contingent boundary experiments (described above, Rayner, 1975; for more recent examples see e.g. Angele & Rayner, 2011; Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Cui et al., 2013) when a non-word preview is used. Usually, the presence of an unusual parafoveal letter string leads to longer fixation times on the currently fixated word (Blanchard, Pollatsek, & Rayner, 1989; Inhoff, Starr, & Schindler, 2000; cf. Pynte et al., 2004). Recently, Angele, Tran, and Rayner (2013) found that a parafoveal word that is the same as or visually similar to the current foveal word (e.g. *news* or *nuws* in the parafovea when the foveal word is *news*) can facilitate the processing of the foveal word and leads to shorter fixation times.[2] Even though such orthographic PoF effects are clear evidence for the impact of parafoveal processing, they occur on a sublexical level and therefore are not necessarily informative about whether word identification occurs for two words in parallel. In the context of the models, while SWIFT explicitly allows all letters within the perceptual span to be processed at the same time, E-Z Reader's *V* stage is thought to involve a scan of all the letters in the perceptual span at the beginning of each fixation before word identification starts. Even though this function of the *V* stage is not currently implemented, it would not be difficult to add this mechanism to the model. Thus, because visual processing is "pre-attentive," PoF effects of an orthographic nature are neither disputed, nor can they distinguish between the two classes of models of attention allocation during reading.

Lexical PoF effects

While there is mostly consensus about the existence of orthographic PoF effects, the evidence for lexical PoF effect is much less clear (again, see Schotter et al., 2012, for a detailed review). To understand the interplay between foveal and parafoveal word identification, we must consider PoF effects at the lexical and post-lexical level. The most investigated word property at the lexical level is word frequency, and these studies attempted to test whether there was an effect of word frequency of the upcoming word (word $n+1$ frequency) on fixation times on the currently fixated word (word n). As described above, studies aimed at investigating PoF effects such as the $n+1$ frequency effect

[2] A standard preview benefit effect was found on the post-boundary word.

generally use one of two methodologies: (1) a correlational approach relating $n+1$ frequency and word n fixation times in an eye-movement data corpus, or (2) a fully experimental approach with a systematic manipulation of the frequency of word $n+1$ (often achieved through the gaze-contingent boundary paradigm). Both methods have specific advantages and disadvantages. The main advantage of the correlational approach is the use of a large data set, encompassing subjects from diverse populations (e.g., college students, non-college bound young adults, developing readers, and older readers) thereby increasing the validity and generalizability of the results. The large amount of data also leads to higher statistical power, enabling correlational studies to find evidence for more subtle successor effects than experimental studies. The main advantages of the experimental approach are the high degree of control over the stimulus material and the straightforward data analysis and interpretation facilitated by the factorial design.

Experiments using sentence reading have largely found no evidence for lexical PoF effects (Angele & Rayner, 2011; Angele et al., 2008; Carpenter & Just, 1983; Henderson & Ferreira, 1993; Inhoff et al., 2000; Perea & Acha, 2009; Rayner, Fischer, & Pollatsek, 1998; White, 2008; Winskel & Perea, 2014; for a summary, see Drieghe, 2011). Hyönä and Bertram (2004) found apparent lexical PoF effects in one reading experiment, but failed to replicate this effect in another. In German, Risse and Kliegl (2012; see also Kliegl, Risse, & Laubrock, 2007) manipulated the frequency of the second word to the right of fixation (word $n+2$). While the frequency of parafoveal word $n+2$ did not have an immediate effect on the current fixation duration on word n, it did have an effect on the fixation duration on the subsequent word $n+1$. They interpreted this result as a *delayed* lexical PoF effect, although the size of this effect was very small (4 ms).

There is much more evidence for successor effects from correlational studies, but as noted above, it is unclear whether these are the same as PoF effects. The best-studied lexical-level successor effects are successor frequency effects (i.e., longer fixation times on words followed by a low frequency word compared to a high frequency word). In German, Kliegl et al. (2006; Kliegl, 2007) analyzed data from 222 subjects from very diverse populations (high school age, college age, and older readers) reading 144 sentences and consistently observed a negative correlation between successor frequency and fixation time consistently across all populations tested. More recently, Wotschack and Kliegl (2013) replicated this effect for both young and older readers and found that the successor effect was stronger for the older readers when the sentences were followed by difficult comprehension questions than when they were followed by easy comprehension questions. Kennedy and Pynte (2005) found successor frequency effects in an eye movement corpus consisting of data from English and French subjects reading long (2600–2800 words) newspaper articles, but this effect was only present when the parafoveal word was short. A recent study by Li, Bicknell, Liu, Wei, and Rayner (2014) that examined an eye movement corpus of reading in Chinese found that words were more likely to be fixated when their successor was higher frequency.

There is not much data on whether predictability of the parafoveal word from the sentence context influences fixation times on the currently fixated word. It is debatable whether predictability is a property of the word itself (i.e., lexical) or a property of the context that generates the expectation for the word (more properly called constraint); because there are opinions on both sides, we will discuss predictability effects in this section to give full consideration to the theoretical debate. In an experimental task, Kennedy, Murray, and Biossiere (2004; see also Murray, 1998; Murray & Rowan, 1998) found a PoF effect of word plausibility given the sentence context (*The savages smacked the child* vs. *The uranium smacked the child*), although their task was not a natural sentence reading task (see Rayner, White, Kambe, Miller, & Liversedge, 2003 for an unsuccessful attempt to replicate this result in the context of reading). Another experimental study by Rayner, Warren, Juhasz, and Liversedge (2004) found a similar effect, although they attributed their finding to mislocated fixations (Drieghe, Rayner, & Pollatsek, 2008). In a corpus analysis, Kliegl et al. (2006) found a successor effect of predictability such that highly predictable successor words were associated with longer fixation times on the currently fixated word. The recent corpus analysis by Li et al. (2014) also found such a successor predictability effect in Chinese. Specifically, more predictable upcoming words $n+1$ were associated with shorter gaze durations and higher skip rates on the preceding word n.

In summary, there is little evidence for lexical PoF effects from experimental studies. However, correlational studies seem to show fairly consistent evidence of successor effects. In the present study, we aimed to test whether successor effects can be observed in the absence of lexical parafoveal information. If this were the case, the successor effects observed in corpus studies may be caused by a factor other than parafoveal preprocessing. Specifically, we used a variation of the moving mask technique (developed by Rayner and Bertera (1979) where the fixated letter(s) were masked) to manipulate the information that was available during each fixation; in our modified paradigm, parafoveal information about the upcoming word $n+1$ and the subsequent word $n+2$ was either made available or unavailable through a moving *parafoveal* mask. As such, our approach is a hybrid between the experimental and the correlational approach: we collected a large data set (128 subjects reading 192 sentences), enabling us to analyze fixation times on every word in every sentence (with the exception of sentence-initial and final words), similar to the correlational approach. In addition, we experimentally controlled the parafoveal information available to readers during each fixation and crossed this factor with a manipulation of the frequency of a critical word in the sentence.

We then performed two sets of analyses on the data. First, we performed a corpus analysis on the full set of data (almost every word in the sentence) with a focus on potential successor frequency effects and investigated whether such apparent successor effects are dependent on the parafoveal information available; that is, if they are only found when no mask was present. Second, in an analysis more akin to the pure experimental approach, we report the results of a factorial analysis (i.e. the equivalent of an

ANOVA) testing whether the effects of an experimental frequency manipulation on a target word had any effect on fixation times on the word preceding the target word, and whether this was affected by the presence or absence of the parafoveal mask. This approach allowed us to investigate PoF effects of word frequency when all other variables were experimentally, not statistically, controlled.

Our predictions were straightforward: In principle we should replicate the successor effects found in previous correlational studies when parafoveal information was available. In the condition in which parafoveal information was masked, there could be two possible outcomes. If successor effects are indeed caused by parafoveal lexical processing, we should not find any evidence for successor effects when no parafoveal lexical information was available (i.e., in the parafoveal mask condition). On the other hand, if successor effects are independent of parafoveal lexical processing and rather linked to the prior sentence context (which was available in all conditions), we should expect to find the same successor effects regardless of the parafoveal mask condition. These predictions hold for both the experimental and the correlational analysis approach. However, it is possible that the correlational analysis can detect more subtle and more complex effects than the factorial analysis.

Method

Subjects

A total of 128 subjects participated in the experiment for the corpus analysis. For the analysis of specific target words (see below), we analyzed the data from a subset of 56 subjects who read target words in all conditions across the full factorial design. All subjects were recruited from the University of California San Diego community and were compensated for their time with either extra course credit or $10 per hour for the experiment session. All subjects were native English speakers with normal or corrected-to-normal vision and were naïve about the purpose of the experiment.

Apparatus

Eye-movements and gaze position were sampled every millisecond using an SR Research Eyelink 1000 eye-tracker. Eye movement data were only collected from the right eye, though viewing was binocular. Sentence stimuli were displayed on a computer monitor with a refresh rate of 150 Hz.[3] Viewing distance was 55 cm, with 3.2 characters equaling approximately 1° of visual angle. A video-game controller was used by subjects to end each trial and respond to comprehension questions (average accuracy 92%).

Materials and procedure

Subjects read 192 sentences silently for comprehension. One third (65) of the sentences were followed by a comprehension question. The sentence stimuli had between 7 and 17 words (average: 12). For a subset of 56 subjects, we added an experimental frequency manipulation in each sentence. Specifically, each of the sentences contained a target word which could either be high or low frequency.[4] For each word, we obtained word frequency (unigram probability) norms from an Americanized version of the British National Corpus (BNC). We also obtained the conditional word trigram predictability for each word (conditional probability of the target region following the two preceding words) estimated from Kneser–Ney-smoothed trigram models from the BNC data as an indicator of predictability from the sentence context. Table 1 shows word length, word frequency, and predictability measures for the sentences used.

Subjects started the experiment by completing a calibration procedure and reading ten practice trials. Before each trial, a drift check was performed to ensure that the calibration was still accurate. If this was the case, subjects then started the trial by fixating a gaze target on the left side of the screen for 250 ms. If the drift check deviated too much from the previous calibration, the subject was re-calibrated. Another calibration was always performed before the start of the experimental trials.

We used a variation of the moving mask paradigm (Rayner & Bertera, 1979) as subjects were reading the sentences to manipulate the parafoveal information available about the upcoming words. This parafoveal mask was gaze-contingent, with the display being updated every time subjects made a saccade between words. The masks were updated both on forward saccades and on regressions. We used four different parafoveal mask conditions, which were counterbalanced within subjects and items in a latin-square design[5]: (1) a control condition with no mask (i.e., the parafoveal words were always visible), (2) a condition in which both of the upcoming parafoveal words (*i.e.,* $n + 1$ and $n + 2$) were masked with Xs, (3) a one-word mask condition in which only the first upcoming parafoveal word $n + 1$ was masked, and (4) a one-word mask condition in which only the second upcoming parafoveal word $n + 2$ was masked. The first word and the last word in every sentence were always visible. The different parafoveal mask conditions were presented un-blocked and in random order. Fig. 1 shows the display change procedure for the $n + 1/n + 2$ mask condition.

[3] At a refresh rate of 150 Hz, the display changes used to update the mask took an average of 3 ms and a maximum of 6.7 ms to be completed after they were initiated. Occasionally, the display change completed more than 10 ms after the saccade that had triggered it completed. In this case, we removed the data for the word in question from the analysis. We did this as a precaution since Slattery, Angele, and Rayner (2011) found that subjects are more likely to be aware of slow display changes, which may affect their reading behavior.

[4] Both subject subsets saw the same sentence frames which had a high frequency target word and a low frequency target word version, although we did not include the frequency manipulation for 72 of the subjects. Instead, these subjects saw the same versions of the sentences (half of the sentences in the low frequency version and half of the sentences in the high frequency version). Subjects in both groups saw each sentence frame exactly once. The target words were either subjects or adjectives (see Appendix B for a list of all sentence frames and target words).

[5] For those 56 subjects in the frequency manipulation group, this design extended to a 4 (parafoveal mask) by 2 (target frequency) latin-square design, which was fully counterbalanced within subjects and sentence frames.

Table 1
Properties of the experimental stimuli.

Measure	Mean	Median	*SD*	Minimum	Maximum
A: Properties of the words used in the sentence stimuli, excluding the first word in a sentence, the two last words in a sentence, and target words					
Word length	5.43	6	2.52	1	10
Word frequency	9763	67.24	17222	0.19	42825
Predictability (conditional trigram probability)	0.03	0	0.05	0	0.21
B: Properties of the target words					
Word length	6.5	6	1.51	5	10
Word frequency (high frequency condition)	152.47	103.77	151.49	13.77	1034
Word frequency (low frequency condition)	2.86	1.78	3.59	0.07	21.65
Predictability (high frequency condition)	<0.01	0	0.03	0	0.43
Predictability (low frequency condition)	<0.01	0	<0.01	0	0.02

Variable	Mean	*SD*
C: Properties of the dependent and the continuous independent variables in the corpus analysis for SFDs		
Single fixation duration	218.35	67.96
$n-1$ Frequency[a]	−3.20	1.49
n Frequency[a]	−3.56	1.37
$n+1$ Frequency[a]	−3.51	1.32
$n-1$ Predictability[a]	−2.60	1.62
n Predictability[a]	−2.97	1.55
$n+1$ Predictability[a]	−2.95	1.57
$n-1$ Length[b]	0.30	0.19
n Length[b]	0.24	0.13
$n+1$ Length[b]	0.30	0.22
Incoming saccade length[c]	7.91	7.13
Outgoing saccade length[c]	5.14	14.60
Fixation position[d]	0.45	0.29

[a] As log 10(Probability of occurrence).
[b] In characters, inverse.
[c] In characters.
[d] In proportion of word length, .5 = word center.

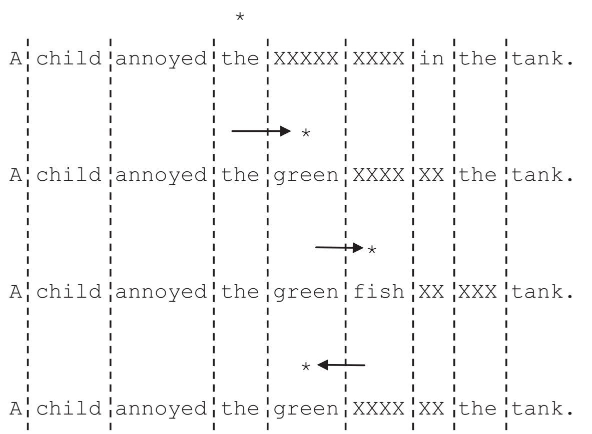

Fig. 1. Moving parafoveal mask procedure for the $n+1/n+2$ mask condition. Asterisks * denote fixation positions and arrows denote saccades. Dashed lines mark word boundaries triggering display changes, which update the mask.

Results and discussion

As described above, we first performed a corpus analysis on the full set of data using most words in each sentence. This was followed by a second analysis of only the target word data from the subset of subjects for which the frequency of the target words was manipulated within sentences. For both analyses, we computed single fixation durations and gaze durations for each subject and fixated word. *Single fixation duration* (SFD) applies to words that are fixated one time only during first pass (that is, before a reader makes any regressions back to the word, and excluding trials in which there was a refixation on it). *Gaze duration* (GD) is the sum of all first-pass fixations on a word (that is, the first fixation and all refixations that occur before a reader's gaze leaves the word). Before computing these dependent measures, very short fixations (<80 ms) were either merged with the previous or subsequent fixation if that fixation occurred within 11 pixels of the short fixation or excluded if this was not the case. Long fixations (>800 ms) were excluded as well (less than 0.3% of SFDs and GDs). From the computed SFDs and GDs (fixations on sentence-initial and the two sentence-final words were not included), we removed outliers that were more than three standard deviations away from the mean for each subject as well as those that were recorded on a word on which a blink had been observed during first-pass reading. Additionally, we excluded SFDs and GDs on any words that contained upper case letters (e.g. proper names), included an apostrophe (e.g. *Jack's*), or ended in a punctuation mark. Despite the various exclusion criteria, there were 60,640 data points available for the analysis of SFD and 64,210 data points available for the analysis of GD in the corpus analysis. Assuming that most of the difference between SFD data points and GD data points is due to refixations, this corresponds to a 5.5% refixation rate. In the analyses using the experimental approach, there were 3922 data points available for analysis of SFD on the pre-target word, 6137 data-points for analysis of SFD on the target word, 4114 data points for the analysis of GD on the pre-target word, and 6630 data points for the analysis of GD on the target word.

For all analyses, we used the lmer function from the lme4 package (version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014) within the R Environment for Statistical Computing (R Core Team, 2014) to fit the LMMs. For each predictor, we report regression coefficients (b), standard errors, and t-values. It is not clear how to determine the degrees of freedom for the t-statistics estimated by the LMMs, making it difficult to estimate p-values ((Baayen, Davidson, & Bates, 2008). However, since our analyses contained a large number of subjects and items, and not many fixed and random effects were estimated, we can operate under the assumption that the distribution of the t-values estimated by the LMMs approximates the normal distribution. We therefore used the two-tailed criterion $|t| \geqslant 1.96$ which corresponds to a significance test at the .05 α-level. The z-values from the generalized LMMs can be interpreted similarly. In addition to this, the models contained random intercepts and random slopes over subjects and word tokens (that is, each word in its specific sentence context) for a selection of fixed effects (Baayen et al., 2008). See below for information on the exact random effect structure.

Corpus analysis (correlational approach)

Model 1: Effects of the mask condition

Before adding any continuous effects to the LMM, we first examined the effect of the $n+1$ and $n+2$ mask conditions shown in Table 2 by fitting two linear mixed models (LMM) with single fixation duration or gaze duration (untransformed) as dependent variables and $n+1$ mask (masked vs. unmasked) and $n+2$ mask and their interaction as fixed effects (using sum contrasts). There were clear main effects of both the $n+1$ (SFD: $b = -16.25$, $SE = .75$, $t = -21.59$; GD: $b = -17.19$, $SE = .74$, $t = -23.29$) and the $n+2$ masks (SFD: $b = -4.25$, $SE = .33$, $t = -12.83$; GD: $b = -4.21$, $SE = .35$, $t = -12.02$), with shorter fixation times being observed when there was no parafoveal mask. In addition, there was a significant interaction between the $n+1$ and $n+2$ masks (SFD: $b = 2.59$, $SE = .27$, $t = 9.58$; GD: $b = 3.00$, $SE = .29$, $t = 10.26$) indicating that the effect of $n+1$ and $n+2$ masks together was less than additive – when $n+1$ was already masked, the $n+2$ mask did not make much of a difference. It is important to note that, even if these effects are interpreted as true PoF effects (as opposed to preview effects from the previous fixation), they would fall into the category of orthographic PoF effects. In order to find evidence for lexical processing, we need to examine the effect of the word frequency of the upcoming word on fixation times on the currently fixated word (i.e., successor frequency and predictability effects).

Table 2
Means for the $n+1/n+2$ mask conditions across the entire corpus. Standard deviations are in parentheses.

$n+1$	$n+2$	SFD	GD
Unmasked	Unmasked	201 (58.8)	204 (65.3)
Unmasked	Masked	203 (56.7)	206 (62)
Masked	Unmasked	229 (70.9)	233 (76.9)
Masked	Masked	243 (72.6)	248 (79.1)

As one of our goals was to compare the experimental and the correlational approach, we present two sets of analyses. The first set follows the correlational approach, including a large number of continuous predictors similar to the analyses presented by Kliegl et al. (2006) and Li et al. (2014). The second set follows the experimental approach, only including factors that were experimentally manipulated (that is, $n+1$ mask, $n+2$ mask, and target word frequency) and limiting the analyses to the pre-target and target words.

Model 2: Effects of successor lexical variables

Fixed effect structure. For this second set of models, we used log-transformed SFD and GD as dependent variables. In order to determine whether successor effects were present in our data, we used a model with similar variables to one used by Kliegl et al. (2006; for the LMM version of this model, see Kliegl, 2007), who found successor effects but didn't use a mask manipulation. Due to this difference, and due to the fact that our data set includes different subjects and items (particularly sentences in English as opposed to German) than Kliegl et al.'s study, it would be inappropriate to expect us to exactly replicate their results. However, we anticipated replicating the general finding of successor frequency effects that they observed, and therefore we based our model on theirs, selecting the same set of predictors as Kliegl et al. and adding a few interactions that we assumed to be theoretically important: word length (as 1/length, analogous to Kliegl et al.), word frequency, and predictability (conditional trigram probability) for the preceding word $n-1$, the currently fixated word n, and the successor word $n+1$. Like Kliegl et al., we included interactions between word n frequency and word n length, between word $n+1$ frequency and word n length, between word $n+1$ predictability and word n length, and, finally, between word n frequency and word $n+1$ frequency. Unlike Kliegl et al., we also included the interactions between word $n-1$ frequency and word $n-1$ length and between word $n+1$ frequency and word $n+1$ length, as spillover and successor frequency effects might be modulated by the length of the preceding or following word. Furthermore, we also included a factor representing the mask or parafoveal preview condition. Since the effect of $n+2$ preview availability was quite small compared to the effect of $n+1$ preview availability and since our primary goal was to investigate the effect of the availability of parafoveal information about $n+1$, we collapsed over the $n+2$ preview conditions and only included $n+1$ preview as a factor in the model. We used treatment contrasts for the $n+1$ preview factor, with the $n+1$ masked condition being the baseline (i.e., $n+1$ masked was coded as 0 and $n+1$ unmasked was coded as 1). As a consequence, the main effects in the model specify the effects observed when $n+1$ was masked, while the difference between the magnitude of effects in the masked vs. unmasked condition are indicated by the interaction terms.[6] Furthermore, the model included interaction terms between mask condition and word n frequency, word $n + 1$ frequency, word n length, word $n + 1$ length, word $n + 1$ predictability, and the three-way interactions between mask condition, word n frequency, and word n length as well as between mask condition, word $n + 1$ frequency, and word $n + 1$ length.

Finally, and again analogously to Kliegl et al., our model incorporated four measures reflecting fixation position on the currently fixated word. These predictors included a linear and a quadratic trend for fixation position (coded between 0 and 1, with .5 being the center of a word), the length of the incoming saccade in characters, and the length of the outgoing saccade in characters. The predictors reflecting the linear and quadratic trends for fixation position were centered and scaled to ensure that they were uncorrelated, as highly correlated predictors can make it harder to fit linear mixed models.

Random effect structure. In our data, there are two sources of random variance (apart from the residual error): subjects and words. As Barr, Levy, Scheepers, and Tily (2013) have recently demonstrated, selecting appropriate random effects for a linear mixed model is critically important. Unfortunately, our fixed effect structure is so complex that implementing Barr et al.'s main recommendation, allowing random slopes for all fixed effects (i.e., maximal random effects structure), was not feasible. However, Barr et al. also suggested that failing to include random slopes for some effects does not affect the interpretation of other critical effects, given that random slopes for the critical effects are present. Following this recommendation, we included random slopes for subjects and words only for the following effects which we deemed critically important: (1) the effect of $n + 1$ frequency; (2) the effect of $n + 1$ predictability; (3) the interaction between $n + 1$ frequency and $n + 1$ word length (necessarily also including the main effect of $n + 1$ word length); and (4) the interaction between $n + 1$ frequency and n frequency (necessarily also including the main effect of n frequency). As we still had convergence problems even with these simplified models, we removed the correlations between the random effects. Removing random correlations reduces model complexity (and potentially, power), but does not lead to the significance tests being anticonservative. Even with this simplification, we could not fit a model for SFD that included both random slopes for the $n + 1$ frequency by $n + 1$ length and the $n + 1$ frequency by n frequency interaction. In this case, we had to fit two separate models, one containing only $n + 1$ frequency by $n + 1$ length interaction and one containing only the $n + 1$ frequency by n frequency interaction. When evaluating the effect of the $n + 1$ frequency by n frequency interaction, we will report results from the latter model. Finally, in order to make model fitting easier, we transformed all continuous predictors (that is, all predictors except $n + 1$ mask), both for fixed and for random effects, into z-values. Table 1 shows means and standard deviations of the predictors before the transformation. Tables 3 and 4 present LMM results for single fixation duration and gaze duration for the predictors discussed above (see Figs. A1 and A2 in the Appendix for plots of the partial interaction effects as predicted by the models); in the interest of clarity and readability, we will discuss only variables that are relevant to the question of parafoveal preprocessing in this section. The remaining variables will be discussed in the appendix.

[6] While, normally, one would choose the baseline condition to be the condition that is most similar to natural reading – in this case, that would be the $n + 1$ unmasked condition – using the $n + 1$ masked condition as the baseline for our treatment contrasts makes it much easier to evaluate the impact of the novel $n + 1$ masked condition. If predictors such as frequency, length, and predictability have a significant effect in the $n + 1$ masked condition, the treatment contrasts specification will make this appear as main effects, while the interactions of frequency with the $n + 1$ mask factor shows how the effects differ in the $n + 1$ unmasked condition. Overall, this makes the interpretation of the rather complex model tables much easier.

Effects of the mask manipulation

Just as in the analysis including only the mask manipulation, the full model showed a significant effect of the $n + 1$ mask on SFD ($b = -0.1$, $SE = 0.004$, $t = -25.57$) and GD ($b = -0.1$, $SE = 0.0041$, $t = -24.92$), with fixation times being longer in the $n + 1$ masked condition (mean SFD = 230 ms, mean GD = 235 ms) than in the $n + 1$ unmasked condition (mean SFD = 208 ms, mean GD = 211 ms).

Successor effects

In total, we found three successor effects that could be considered PoF effects. First, we observed a significant effect of $n + 1$ predictability on SFD ($b = -0.019$, $SE = 0.0047$, $t = -4.02$) and GD ($b = -0.018$, $SE = 0.0047$, $t = -3.86$), with more predictable words $n + 1$ being associated with shorter fixation times on the currently fixated word. The interaction of $n + 1$ predictability and mask condition, however, did not reach significance (SFD: $b = 0.0036$, $SE = 0.0047$, $t = 0.76$; GD: $b = 0.0034$, $SE = 0.0048$, $t = 0.70$), suggesting that the $n + 1$ predictability effect was present both when $n + 1$ was masked and when it was unmasked. Predictability involves more than just lexical information (syntactic and semantic properties of both the context and the word also influence the predictability of the word). Still, an $n + 1$ predictability effect like the one we found would commonly be interpreted as a high-level PoF effect. This $n + 1$ predictability effect showed a significant interaction with word n length in SFD ($b = -0.0095$, $SE = 0.0038$, $t = -2.53$), while the same interaction term was only marginally significant in GD ($b = -0.0073$, $SE = 0.0037$, $t = -1.96$), indicating that the $n + 1$ predictability effect was stronger when the currently fixated word n was long than when it was short. The direction of the $n + 1$ predictability effect (higher $n + 1$ predictability leads to shorter fixation times on n) was opposite to the direction of the effects observed by Kliegl et al. (2006), but in the same direction as the effect observed by Li et al. (2014). This may be due to the difference in language between these studies (German in Kliegl et al., Chinese in Li et al.), but it is also worth mentioning that

Table 3
LMM results for single fixation duration in the corpus analysis. Only fixed effects are shown. Significant effects are represented in boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.40188**	**0.01189**	**454.397**
Predictors relevant to parafoveal processing			
Preview (unmasked)	**−0.10299**	**0.00403**	**−25.568**
Frequency *n*	**−0.01153**	**0.00546**	**−2.114**
1/Length *n*	−0.00898	0.00590	−1.522
Predictability *n*	**−0.02434**	**0.00324**	**−7.512**
Frequency *n* + 1	0.00761	0.00562	1.355
1/Length *n* + 1	**0.01548**	**0.00475**	**3.262**
Predictability *n* + 1	**−0.01884**	**0.00468**	**−4.024**
Frequency *n* + 1/Length *n* + 1	**−0.01658**	**0.00485**	**−3.417**
Frequency *n* * Frequency *n* + 1	**0.00882**	**0.00364**	**2.426**
Frequency *n* + 1/Length *n*	0.00145	0.00451	0.322
Predictability *n* + 1/Length *n*	**−0.00951**	**0.00377**	**−2.526**
Preview (unmasked) * Frequency *n*	0.00132	0.00346	0.382
Preview (unmasked) * Frequency *n* + 1	−0.00152	0.00564	−0.270
Preview (unmasked)/Length *n* + 1	0.00059	0.00437	0.136
Preview (unmasked) * Predictability *n* + 1	0.00359	0.00470	0.764
Preview (unmasked) * Frequency *n* + 1/Length *n*	−0.00215	0.00305	−0.704
Preview (unmasked) * Frequency *n* * Frequency *n* + 1	−0.00258	0.00293	−0.878
Preview (unmasked) * Frequency *n* + 1/Length *n* + 1	0.00031	0.00455	0.067
Other predictors			
Frequency *n* − 1	**0.02685**	**0.00364**	**7.383**
1/Length *n* − 1	**0.01093**	**0.00318**	**3.436**
Predictability *n* − 1	**−0.02827**	**0.00298**	**−9.473**
Incoming saccade length	**0.01191**	**0.00118**	**10.064**
Outgoing saccade length	**−0.00236**	**0.00114**	**−2.077**
Fixation position (linear trend)	**0.01205**	**0.00113**	**10.700**
Fixation position (quadratic trend)	**−0.01914**	**0.00118**	**−16.251**
Frequency *n* − 1/Length *n* − 1	**0.00905**	**0.00302**	**3.000**
Preview (unmasked)/Length *n*	**0.02164**	**0.00467**	**4.635**
Frequency *n*/Length *n*	0.00229	0.00501	0.456
Preview (unmasked) * Frequency *n*/Length *n*	**−0.00735**	**0.00352**	**−2.087**

our predictability measure (conditional trigram probability) is slightly different from the cloze predictability measures used by the other studies. As *n* + 1 conditional trigram probability was strongly correlated with *n* + 1 word frequency, it is possible that the collinearity made it difficult to distinguish predictability from frequency effects. Indeed, in a model without conditional trigram probability predictors, *n* + 1 frequency had a robust main effect in the *n* + 1 masked condition (SFD: $b = -0.011$, $SE = 0.0044$, $t = -2.40$; GD: $b = -0.011$, $SE = 0.0045$, $t = -2.56$).

When *n* + 1 predictability was included in the model, *n* + 1 frequency did not have an effect either when *n* + 1 was masked (SFD: $b = 0.0076$, $SE = 0.0056$, $t = 1.35$; GD: $b = 0.0061$, $SE = 0.0057$, $t = 1.08$) or when it was unmasked (SFD: $b = -0.0015$, $SE = 0.0056$, $t = -0.27$; GD: $b = -0.00088$, $SE = 0.0058$, $t = -0.15$). Also, there was a significant main effect of *n* + 1 length in the *n* + 1 masked condition (SFD: $b = 0.015$, $SE = 0.0047$, $t = 3.26$; GD: $b = 0.016$, $SE = 0.0048$, $t = 3.40$), indicating that reader spent more time fixating the current word when word *n* + 1 was short than when it was long. This effect did not seem to be modulated by the *n* + 1 mask condition, as indicated by the non-significant interaction term between *n* + 1 length and preview (SFD: $b = 0.00059$, $SE = 0.0044$, $t = 0.14$; GD: $b = -0.0016$, $SE = 0.0045$, $t = -0.35$). This is not surprising as valid length information for words was available whether the words were masked or not. However, there was a significant interaction between *n* + 1 frequency and *n* + 1 length (SFD: $b = -0.017$, $SE = 0.0049$, $t = -3.42$; GD: $b = -0.017$, $SE = 0.0049$, $t = -3.41$), indicating that, in the *n* + 1 masked condition, there was an *n* + 1 successor frequency effect that differed for *n* + 1 words of different lengths. Fig. 2 depicts this interaction on SFD (note that the GD pattern was very similar; see also Fig. A1 in the Appendix for the partial effects as predicted by the LMM). Subjects appeared to spend more time fixating the current word when the upcoming word *n* + 1 was very short (1 or 2 letters) and of high frequency than when it was very short and of low frequency.

This effect may be caused by skipping. Kliegl and Engbert (2005) found that subjects tend to make shorter fixations before skipping short and high frequency words than before fixating them. Of course, there are not many low frequency words that are this short, which makes this finding quite specific. There was no frequency effect for 3-, 4-, and 5-letter *n* + 1 words. However, readers seemed to be sensitive to the frequency of *n* + 1 words with 6 or more letters, spending more time on the current word when *n* + 1 was long and of low frequency than when it was long and of high frequency. This finding replicates the effect found by Kliegl et al. (2006), although, in their study, the frequency effect was present independent of *n* + 1 word length, while we only observed the effect for long (6 letters or more) *n* + 1 words. This may be due to differences between English and German, which has more long words and fewer short words than English.

Table 4
LMM results for gaze duration in the corpus analysis. Only fixed effects are shown. Significant effects are represented in boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.417940**	**0.011999**	**451.536**
Predictors relevant to parafoveal processing			
Preview (unmasked)	**−0.103161**	**0.004139**	**−24.924**
Frequency *n*	−0.004821	0.006141	−0.785
1/Length *n*	**−0.012029**	**0.005869**	**−2.050**
Predictability *n*	**−0.026219**	**0.003298**	**−7.951**
Frequency *n* + 1	0.006086	0.005651	1.077
1/Length *n* + 1	**0.016499**	**0.004846**	**3.405**
Predictability *n* + 1	**−0.018139**	**0.004694**	**−3.865**
Frequency *n* + 1/Length *n* + 1	**−0.016596**	**0.004870**	**−3.408**
Frequency *n* * Frequency *n* + 1	**0.008605**	**0.003723**	**2.311**
Frequency *n* + 1/Length *n*	0.001552	0.004593	0.338
Predictability *n* + 1/Length *n*	−0.007253	0.003694	−1.964
Preview (unmasked) * Frequency *n*	0.000315	0.003559	0.088
Preview (unmasked) * Frequency *n* + 1	−0.000875	0.005807	−0.151
Preview (unmasked)/Length *n* + 1	−0.001562	0.004499	−0.347
Preview (unmasked) * Predictability *n* + 1	0.003404	0.004842	0.703
Preview (unmasked) * Frequency *n* + 1/Length *n*	−0.003111	0.003142	−0.990
Preview (unmasked) * Frequency *n* * Frequency *n* + 1	−0.002526	0.003028	−0.834
Preview (unmasked) * Frequency *n* + 1/Length *n* + 1	−0.000052	0.004677	−0.011
Other predictors			
Frequency *n* − 1	**0.025556**	**0.003727**	**6.856**
1/Length *n* − 1	**0.011991**	**0.003272**	**3.665**
Predictability *n* − 1	**−0.027644**	**0.003067**	**−9.012**
Incoming saccade length	**0.011494**	**0.001219**	**9.428**
Outgoing saccade length	**−0.001230**	**0.001171**	**−1.051**
Fixation position (linear trend)	**0.013097**	**0.001166**	**11.231**
Fixation position (quadratic trend)	**−0.016788**	**0.001213**	**−13.844**
Frequency *n* − 1/Length *n* − 1	**0.006616**	**0.003097**	**2.136**
Preview (unmasked)/Length *n*	**0.021190**	**0.004818**	**4.398**
Frequency *n*/Length *n*	0.006041	0.004951	1.220
Preview (unmasked) * Frequency *n*/Length *n*	**−0.009347**	**0.003604**	**−2.593**

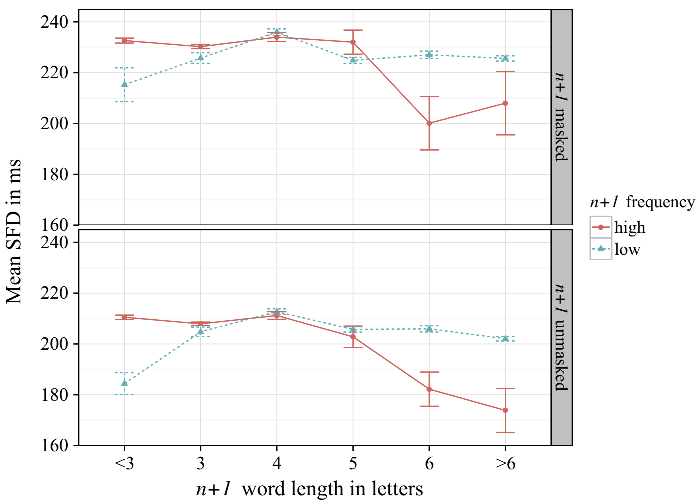

Fig. 2. Effects of $n + 1$ word length, $n + 1$ frequency, and $n + 1$ mask condition on single fixation duration (error bars show standard error over the entire data set).

Additionally, $n + 1$ word frequency significantly interacted with the frequency of the current word n, such that the $n + 1$ frequency effect was stronger when word n was of lower frequency (SFD: $b = 0.0088$, $SE = 0.0036$, $t = 2.43$; GD: $b = 0.0086$, $SE = 0.0037$, $t = 2.31$). It is possible that whatever process drives the apparent $n + 1$ frequency effect only has a chance to influence fixation times on the current word if that word is relatively difficult to process and is therefore fixated longer.

Critically, there was no three-way interaction between the mask condition and the interaction between $n + 1$ length and $n + 1$ frequency (SFD: $b = 0.00031$, $SE = 0.0045$,

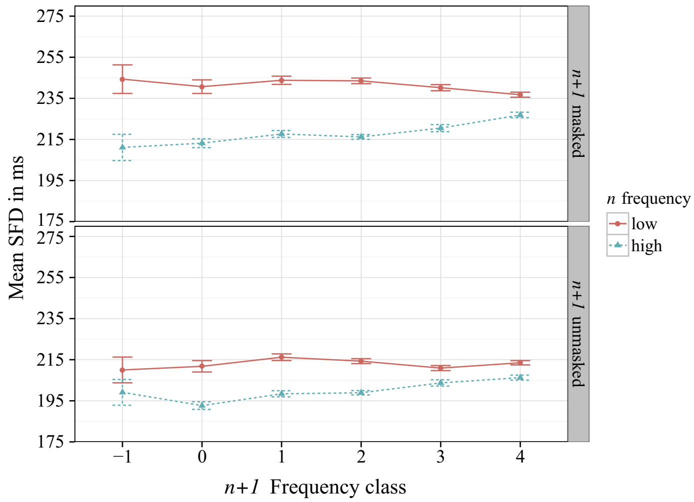

Fig. 3. Effects of n frequency, $n+1$ frequency, and $n+1$ mask condition on single fixation duration (error bars show standard error over the entire data set). $n+1$ frequency class is determined by taking the log10 of $n+1$ frequency per million, rounding to the nearest integer.

$t = 0.07$; GD: $b = -5.2e{-}05$, $SE = 0.0047$, $t = -0.01$), suggesting that the $n+1$ frequency by $n+1$ length interaction was present both when $n+1$ was masked and when $n+1$ was unmasked. The same was true for the three-way interaction between the mask condition and the interaction between $n+1$ frequency and n frequency (SFD: $b = -0.0019$, $SE = 0.0029$, $t = -0.64$; GD: $b = -0.0025$, $SE = 0.003$, $t = -0.83$), suggesting that the $n+1$ frequency by n frequency interaction was also present both when $n+1$ was masked and when $n+1$ was unmasked. At first glance, the three effects we found would seem to replicate previous findings of successor frequency effects. However, these apparent successor effects were present both when $n+1$ was unmasked and when $n+1$ was masked (see Figs. 2 and 3; Figs. A1 and A2 in the Appendix illustrate the corresponding partial effects predicted by the LMMs). This is quite surprising as it suggests that the $n+1$ frequency successor effects (as well as the $n+1$ predictability successor effect) we observed are caused by a process that is not dependent on lexical parafoveal processing.

Factorial analysis (experimental approach)

As described above, the experimental approach is characterized by evaluating the effects of experimentally introduced manipulations, attempting to control for any other influences by keeping them constant across conditions. For our analysis corresponding to the experimental approach, we only used the data from a subset of 56 subjects, for whom the frequency for one target word in each sentence was systematically manipulated (see Table 5 for pre-target and target word means). We fit LMMs for four dependent measures: SFD and GD on the pre-target word, and SFD and GD on the target word. As fixed effects, we included $n+1$ mask (coded with $n+1$ masked as the baseline) and target word frequency (that is, $n+1$ frequency in the pre-target word analyses and n frequency in the target word analyses, with a contrast coding low frequency as -1 and high frequency as 1). As with the corpus analysis approach, we only included theoretically critical random effects. The reported models have random slopes for $n+1$ mask, target frequency, and the interaction of $n+1$ mask by target frequency for items (sentences) and subjects. Tables 6 and 7 show the LMM results for SFD and GD on the pre-target word, while Tables 8 and 9 show the LMM results for SFD and GD on the target word, respectively.

Pre-target word

As observed in the corpus analysis, there was a significant effect of the $n+1$ mask on SFD ($b = -0.16$, $SE = 0.0082$, $t = -19.46$) and GD ($b = -0.16$, $SE = 0.0085$, $t = -19.31$), with SFD and GD on the pre-target word being shorter when $n+1$ was unmasked than when it was masked (SFD: $n+1$ masked: 227 ms, $n+1$ unmasked: 193 ms; GZD: $n+1$ masked: 229 ms, $n+1$ unmasked: 194 ms). Of course this effect can also be viewed as a standard preview benefit effect due to the fact that the the $n+1$ mask, apart from preventing concurrent processing of each upcoming word (which may affect the *ongoing* fixation), also prevented parafoveal pre-processing of each upcoming word, affecting the *subsequent* fixation. These two effects are perfectly confounded since the mask was present during all fixations. The mask effect may therefore be due to lack of parafoveal preview on the preceding fixation rather than the impact of the concurrent parafoveal input.

In the $n+1$ masked condition, there was no significant effect of the target word (i.e., $n+1$) frequency manipulation on either SFD ($b = -0.0076$, $SE = 0.0057$, $t = -1.34$) or GD ($b = -0.0096$, $SE = 0.0058$, $t = -1.65$) on the pre-target word. The lack of interactions between target frequency and the mask conditions (all $|t| < 1.5$) means that we did not find an $n+1$ frequency effect in the $n+1$ unmasked condition either. It is important to note that, since the length of the target word was controlled, we could not evaluate the effect of word length or its interaction with frequency in this analysis. However, the majority of the target words had 6 or more letters, which is within the range for which we observed

Table 5
Mean SFD and GD on the pre-target word and the target word as a function of mask condition and target word frequency in the factorial analysis. Standard deviations are in parentheses.

Word *n + 1* frequency	Word *n* + 1 mask	Pretarget SFD	Pretarget GD	Target SFD	Target GD
High	Masked	225 (69.7)	227 (73.8)	254 (64.1)	256 (72.4)
High	Unmasked	194 (53.9)	195 (56.4)	205 (53.9)	206 (57.5)
Low	Masked	228 (71.9)	231 (77.9)	271 (73.2)	274 (81.2)
Low	Unmasked	193 (55.4)	194 (59.1)	225 (65)	227 (71.8)

Table 6
LMM results for single fixation duration on the pre-target word in the factorial analysis. Only fixed effects are shown. Significant effects are represented by boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.375**	**0.017**	**320.119**
***n* + 1 Preview (masked vs. unmasked)**	**−0.160**	**0.008**	**−19.463**
Frequency *n* + 1 (low vs. high)	−0.008	0.006	−1.343
n + 1 Preview (masked vs. unmasked) * Frequency *n* + 1 (low vs. high)	0.013	0.008	1.636

Table 7
LMM results for gaze duration on the pre-target word in the factorial analysis. Only fixed effects are shown. Significant effects are represented by boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.381**	**0.016**	**327.711**
***n* + 1 Preview (masked vs. unmasked)**	**−0.163**	**0.008**	**−19.309**
Frequency *n* + 1 (low vs. high)	−0.010	0.006	−1.649
n + 1 Preview (masked vs. unmasked) * Frequency *n* + 1 (low vs. high)	0.015	0.008	1.775

Table 8
LMM results for single fixation duration on the target word in the factorial analysis. Only fixed effects are shown. Significant effects are represented by boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.539**	**0.015**	**370.093**
***n* + 1 Preview (masked vs. unmasked)**	**−0.202**	**0.006**	**−31.882**
Frequency *n* (low vs. high)	**−0.032**	**0.006**	**−5.821**
***n* + 1 Preview (masked vs. unmasked) * Frequency *n* (low vs. high)**	**−0.015**	**0.006**	**−2.427**

Table 9
LMM results for gaze duration on the target word in the factorial analysis. Only fixed effects are shown. Significant effects are represented by boldface.

Predictor	Estimate	Standard error	*t* value
(Intercept)	**5.541**	**0.015**	**377.316**
***n* + 1 Preview (masked vs. unmasked)**	**−0.202**	**0.007**	**−30.620**
Frequency *n* (low vs. high)	**−0.032**	**0.006**	**−5.579**
n + 1 Preview (masked vs. unmasked) * Frequency *n* (low vs. high)	−0.013	0.007	−1.953

$n + 1$ frequency effects in the corpus analysis. In summary, our analyses following the experimental approach did not find any evidence for lexical PoF effects.

Target word

An important test of the validity and statistical power of these analyses following the experimental approach is whether we can find the (immediacy) effect of target word frequency on the target word itself. We therefore also report analyses of SFD and GD on the target word.

We found the same effects of the $n + 1$ mask (SFD: $b = -0.2$, $SE = 0.0063$, $t = -31.88$; GD: $b = -0.2$, $SE = 0.0066$, $t = -30.62$) as in the pre-target analysis (SFD: $n + 1$ masked: 262 ms, $n + 1$ unmasked: 214 ms; GZD: $n + 1$ masked: 265 ms, $n + 1$ unmasked: 194 ms). More importantly, there was a clear effect of target word frequency on SFD ($b = -0.032$, $SE = 0.0056$, $t = -5.82$) and GD ($b = -0.032$, $SE = 0.0058$, $t = -5.58$). Specifically, SFDs and GDs were shorter in the high frequency condition (SFD: 227 ms, GD: 230 ms) than in the low frequency condition (SFD: 246 ms, GD: 248 ms). The frequency effect was modulated by the $n + 1$ mask condition, as the significant interaction shows (SFD: $b = -0.015$, $SE = 0.0063$, $t = -2.43$; GD, marginal: $b = -0.013$, $SE = 0.0066$, $t = -1.95$): when $n + 1$ was masked, the frequency effect was slightly smaller (SFD, high frequency: 254, low frequency: 271; GD, high frequency: 256, low frequency: 274) than when $n + 1$ was unmasked (SFD, high frequency: 205, low frequency: 225; GD, high frequency: 206, low frequency: 227).

In summary, our analyses were able to find a robust (immediacy) frequency effect on the target word, one of the benchmark effects in reading (Rayner, 1998), but we found no evidence for lexical PoF effects. This suggests that our analysis had at least enough power to detect the immediacy frequency effect. Of course, it is possible that there was an $n+1$ frequency effect with such a small effect size that the power of our experimental analysis approach was insufficient to detect it. Also, as described above, if there was an interaction between the $n+1$ frequency effect and another word property (as our corpus analysis suggests) it is possible that the target words were selected in such a way that they happened to be associated with a very weak $n+1$ frequency effect (e.g. by being medium-length words rather than covering the entire word length spectrum). The results from our experimental approach are much easier to interpret than the results from our corpus analysis, but they might not be as representative of the full spectra of subjects and language as the latter analysis was. Combined, the two classes of analyses represent the range of typical approaches described in the Introduction, and converge to describe a similar picture of parafoveal processing of the upcoming word during reading (see the general discussion below).

General discussion

In the present study, we manipulated the amount of parafoveal information available to readers during every fixation. This enabled us to test whether the lexical successor effects found in previous studies were actually based on parafoveal lexical processing or not. Our results from the correlational analysis involving most words in a sentence suggest that lexical successor effects may not be caused by parafoveal lexical processing: we found successor frequency and predictability effects both when parafoveal information about the upcoming word was available (i.e. when it was unmasked) and when it was masked by our experimental manipulation. This strongly suggests that the successor frequency and predictability effects we observed cannot be caused by parafoveal processing. Furthermore, when we restricted the analysis to one target word per sentence whose frequency was experimentally manipulated, there was no effect of $n+1$ frequency on the word preceding the target word.

The first part of this General discussion will focus on the implications of this finding for parallel and serial accounts of word identification during reading. In the second part, we will address potential causes of the successor effects that are not related to parafoveal processing. Finally, we will discuss how our finding that the predictability of the upcoming word affected fixation duration on the current word fits in with current accounts of how prediction of upcoming words influences word identification.

Implications for parallel and serial accounts of word identification

As described in the introduction, neither the E-Z Reader model (which is based on the assumption of serial word identification) nor the SWIFT model (which assumes that word identification is performed on all visible words in parallel) currently have a mechanism by which the content of the parafovea could influence fixation times (at least as far as first and single fixation durations are concerned). However, one can argue that it would be easier to adjust the SWIFT model to allow concurrent parafoveal lexical processing than it would be to adjust the E-Z Reader model (given its assumption of serial lexical processing, such an adjustment might even be impossible). As a consequence, if there was clear evidence for parafoveal lexical processing that is concurrent with foveal lexical processing, this would be very problematic for the serial word identification account. Up to now, lexical successor effects constituted some of the most reliable potential pieces of evidence for concurrent parafoveal lexical processing. However, our data show that these effects can arise even without access to information about the parafoveal word (i.e., when it is masked), suggesting that conclusions along these lines are premature.

Importantly, our findings are not evidence against concurrent parafoveal lexical processing per se. However, they demonstrate that lexical successor effects, while indeed reliable, are not necessarily evidence for parafoveal lexical processing. Instead, they may be evidence for the existence of processes that are not based on parafoveal input but whose effect is correlated with the frequency of the parafoveal word. Regardless of whatever causes these effects, if they are reliable, computational models of eye-movements will have to be adapted to explain them. An account that attributes successor effects to parafoveal lexical processing (which has been ruled out by our data) may be more difficult to implement for E-Z Reader than for SWIFT; however, formal simulations must be conducted before any such conclusions can be made (Rayner, 2009b; Schotter, Reichle, & Rayner, 2014). In contrast, an account that attributes successor effects to processes other than lexical pre-processing (as suggested by our data) may pose challenges that are similar for both models (for an attempt to do this with E-Z Reader, see Reichle & Drieghe, in press). In the following section, we will discuss some processes that might be candidates for non-parafoveal causes for the apparent successor effects we found.

Potential causes of successor effects that are independent of parafoveal preprocessing

Several potential causes (that are not due to parafoveal pre-processing) for apparent lexical PoF effects have been discussed (Drieghe, 2011), namely calibration errors, binocular disparity, and mislocated fixations (see Drieghe et al., 2008; Rayner, Juhasz, & Brown, 2007; Rayner, Pollatsek, et al., 2007 for discussion). All these explanations follow the same general principle: they propose that a reader's gaze is recorded in a location that does not correspond to either the actual gaze location or the reader's actual attentional focus. In the case of calibration error, the explanation is simply that some fixations are erroneously recorded farther to the left than they actually are (for readers of English). This results in a fixation that is measured on one word but actually reflects processing of the upcoming word, resulting in an apparent PoF effect. Reichle and Drieghe (in press) show that E-Z Reader actually predicts such effects if a certain amount of calibration error is assumed.[7] The binocular disparity explanation proposes that, as readers' eyes are slightly crossed much of the time, studies recording only the position of the right eye might find apparent PoF effects in situations where the left and the right eye are fixating different words (see Kirkby, Webster, Blythe, & Liversedge, 2008, for a detailed discussion of binocular disparity in reading). While readers might actually be processing the word fixated by the left eye, a study using monocular recording would use the position of the right eye to determine which word was fixated, resulting in the fixation being coded as being located on the preceding word. Finally, the mislocated fixations explanation states that readers attempting to skip a word (which in the E-Z Reader model necessarily means that they have shifted their attention to the subsequent word and are processing it) occasionally undershoot their target and accidentally fixate the word they intended to skip. In this case, the fixation will reflect the properties of the word that is actually being processed (i.e., the successor) instead of the currently fixated word.

None of the phenomena described above can explain our results, as they all require the upcoming word to be available for processing. If, in the $n + 1$ mask condition, a fixation is erroneously coded as being on the word preceding the word that is actually being processed, be it due to calibration error, due to binocular disparity, or due to mislocated fixations, the $n + 1$ mask will deny readers any information about that word. This may result in a disruption of the reading process as readers try to process the mask, but it cannot possibly lead to successor frequency effects since the frequency of the successor is unknown when it is masked.

A more plausible group of explanations concerns readers forming expectations about the upcoming words. For example, readers may be able to anticipate that the next word is likely to be difficult (either low frequency or low predictability), given the sentence context and, in anticipation of this, slow down in preparation for them. If readers' predictions about the difficulty of upcoming words are usually correct, the predicted processing difficulty will be correlated with the actual frequency and predictability of those words, leading to apparent PoF effects of upcoming words. Such predictions need not necessarily be for specific words: even if readers formed expectations for part-of-speech, those would be correlated with frequency and predictability and could lead to the effects we observed. The difficulty of producing such an expectation, or perhaps, the expectation of upcoming difficulty, might be correlated with the frequency of the upcoming word and cause the successor effects we observed. On first glance, this idea might seem problematic: since the predictability of the upcoming word was included as a predictor in the model (and was significant), in theory it should account for the expectation effects. However, it is important to keep in mind that the actual probability of a word given its context (surprisal) is not the same as the reader's expectation for a word's probability given the context (entropy). Our predictor (that is, conditional trigram probability) might not capture all the variance caused by expectation as its effect might be too subtle and non-linear. Furthermore, readers might be generating expectations that are more general than can be tested with these trigram models: for example, readers might anticipate a particular part of speech, which could be correlated with word frequency, rather than a particular word form.

One such process, along with its influences on eye movements, has been described by Altmann and Kamide (1999). In a visual world experiment where subjects were viewing a simple scene accompanied by a spoken sentence, subjects were able to use information from a verb (e.g. *eat* when there was only one edible object visible in the scene) to restrict their attention (indexed by their eye movements) to the appropriate object even before it was named. Such cues for predictive inferencing are not restricted to verb-object combinations: Grondelaers, Speelman, Drieghe, Brysbaert, and Geeraerts (2009) reported that, in Dutch, speakers (and writers) tend to produce the optional word *er* (*there*) in situations where the continuation of a sentence might be unexpected by the reader or listener and to omit it in situations where the continuation of a sentence is expected and not surprising (cf. Ferreira & Schotter, 2013, for a lack of speakers adaptation to listeners' expectations). Readers might be quite sensitive to similarly subtle cues in English and might adjust their fixation duration accordingly. Of course, our design cannot control either experimentally or statistically for the effect of such cues, which may well be correlated to some degree with $n + 1$ frequency.

In the present study, we have presented a comparison between the correlational and experimental approaches in an attempt to study parafoveal processing of words during reading. We have summarized the advantages and disadvantages of each approach and performed both types of analyses on the same data set, providing a clear demonstration of the limitations of making conclusions based on the results of only one approach. Specifically, the results of the experimentally controlled factorial analysis did not replicate the results of the statistically controlled correlational analysis. Additionally, the correlational approach still suggested successor effects (which, in other studies have been interpreted as PoF effects), even when no identification of the upcoming word was possible (i.e., when the upcoming words were masked).

Our results cast doubt on whether findings of successor frequency in corpus analyses can be easily interpreted as evidence of parallel lexical processing of foveal and parafoveal words. The most obvious interpretation of the successor effect as evidence of such parafoveal lexical processing is contradicted by the fact that we observed effects of successor length and frequency both when preview of the upcoming word $n + 1$ was available (i.e., when $n + 1$ was unmasked) throughout sentence reading and when $n + 1$ was masked, denying parafoveal preview. On a more general level, our results raise some questions concerning the interpretability of very subtle, but statistically significant results with small effect sizes that typically only reach significance in large corpus analyses. Even though improved study designs and statistical techniques such as

[7] However, mislocated fixations and calibration error cannot explain the combination of a negative $n + 1$-frequency and positive $n + 1$-predictability effect reported in Kliegl et al. (2006).

LMMs provide us the power necessary to detect such small effects, it may be hard to pinpoint exactly what causes them. Furthermore, we must emphasize the practical impact of detecting such subtle relationships. While it is certainly important to understand even small influences on the duration and probability of fixations during the reading process, it is more important to single out those factors that have strong influences on reading, such as visual and lexical properties of the currently fixated word.

In conclusion, the purpose of this study is not to declare a "winner" in the contest between experimental and correlational approaches to the study of reading. On the contrary, we believe that in order to get a clear picture of what is happening during the reading process, it is imperative to use information from both of these methods. However, when interpreting their results, the individual strengths and weaknesses of each approach need to be considered. The best evidence for the influence of a variable on the reading process would be to observe it with both correlational and experimental approaches. This means that efforts should be made to reconcile apparent contradictions between the findings in experimental and corpus studies in the literature. The present study is a first attempt at such a reconciliation, with the eventual goal to conclusively map the relationship between sentence context, current word properties, and parafoveal input.

Acknowledgments

This research was supported by Grant HD065829 from the National Institute of Child Health and Human Development. We thank Sally Andrews, Denis Drieghe, and Reinhold Kliegl for their comments on an earlier version of the article. Open access publication of this article was supported by the Bournemouth University Open Access Publication Fund.

Appendix A. Other effects included in the LMMs for the correlational approach

In this appendix, we report all those effects observed in the correlational approach LMM analyses that were not directly relevant to parafoveal preprocessing and therefore were not reported in the main text.

Effects of fixation position. We observed very strong effects of fixation position on SFD and GD. The linear trend was positive and significant in both SFD ($b = 0.012$, $SE = 0.0011$, $t = 10.70$) and GD ($b = 0.013$, $SE = 0.0012$, $t = 11.23$). There also was a significant quadratic trend in SFD ($b = -0.019$, $SE = 0.0012$, $t = -16.25$) and GD ($b = -0.017$, $SE = 0.0012$, $t = -13.84$). The length of the incoming saccade showed a significant effect on both SFD ($b = 0.012$, $SE = 0.0012$, $t = 10.06$) and GD ($b = 0.011$, $SE = 0.0012$, $t = 9.43$), with longer fixation times when the incoming saccade was large, while there was no such effect for the length of the outgoing saccade (all $t < 1.96$).

Lag effects: effects of lexical properties of the word preceding the currently fixated word (word $n - 1$). We observed significant lag effects of frequency in both SFD ($b = 0.027$, $SE = 0.0036$, $t = 7.38$) and GD ($b = 0.026$, $SE = 0.0037$, $t = 6.86$), with words receiving longer fixations when they had been preceded by low-frequency words than when they had been preceded by high frequency words. This effect was modulated by the length of the preceding word, such that words preceded by short low-frequency words $n - 1$ were fixated longer than words preceded by long low-frequency words $n - 1$ (SFD: $b = 0.009$, $SE = 0.003$, $t = 3.00$; GD $b = 0.0066$, $SE = 0.0031$, $t = 2.14$. Finally, we observed a lag effect of predictability, with words showing lower SFD ($b = -0.028$, $SE = 0.003$, $t = -9.47$) and GD ($b = -0.028$, $SE = 0.0031$, $t = -9.01$) when the preceding word had been highly predictable from the context than when it had not been predictable.

Immediacy effects: effects of the properties of the currently fixated word (word n). There was a significant frequency effect on SFD ($b = -0.012$, $SE = 0.0055$, $t = -2.11$) but not GD ($b = -0.0048$, $SE = 0.0061$, $t = -0.78$), indicating that words received shorter fixations when they were of high frequency.

There also was a significant effect of predictability, with words showing shorter SFD ($b = -0.024$, $SE = 0.0032$, $t = -7.51$) and GD ($b = -0.026$, $SE = 0.0033$, $t = -7.95$) when a word was predictable from the sentence context than when it was not.

When $n + 1$ was masked, there was a significant main effect of word length on GD ($b = -0.012$, $SE = 0.0059$, $t = -2.05$), but not SFD ($b = -0.009$, $SE = 0.0059$, $t = -1.52$), and there was no significant interaction between word frequency and length (all $t < 1.96$). However, there was a significant interaction between mask condition and word length, indicating that word length did have an effect when $n + 1$ was unmasked (SFD: $b = 0.022$, $SE = 0.0047$, $t = 4.63$; GD: $b = 0.021$, $SE = 0.0048$, $t = 4.40$). There was a significant three-way interaction between mask condition, frequency, and word length (SFD: $b = -0.0073$, $SE = 0.0035$, $t = -2.09$; GD: $b = -0.0093$, $SE = 0.0036$, $t = -2.59$). This interaction seems to indicate that the reverse effect of word length in the unmasked condition was strongest when word n was of low frequency, while it was weaker for higher frequency words n and even reversed for extremely high frequency words n. It may seem a bit surprising that there was no strong main effect of word n length in the $n + 1$ masked condition in SFD, and that the effect of n length goes in the opposite direction (with shorter words receiving longer fixations) in the unmasked preview condition. However, it is important to keep in mind that the predictors in LMMs describe partial effects, that is, the effect of the independent variable in question when all other variables in the model are statistically controlled (see Figs. A1 and A2 for examples of partial effects with regard to word $n + 1$). In our raw data, longer words were associated with longer fixations, just as expected. Word n length was also highly correlated with word n frequency and predictability, though, with shorter words tending to be higher frequency and more predictable. With n frequency and predictability in the model (and therefore being statistically controlled), there either was no remaining partial effect of word length (in the masked condition) or that the remaining partial effect of word length was reversed (in the unmasked condition). This situation is known as a suppressor effect.

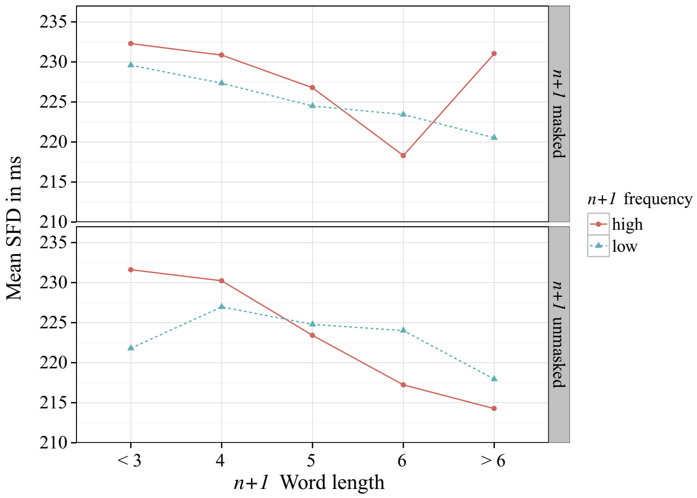

Fig. A1. Predicted partial effect of $n + 1$ word length, $n + 1$ frequency, and $n + 1$ mask condition on single fixation duration(from the fitted LMMs using the *remef* function by Hohenstein & Kliegl, 2014).

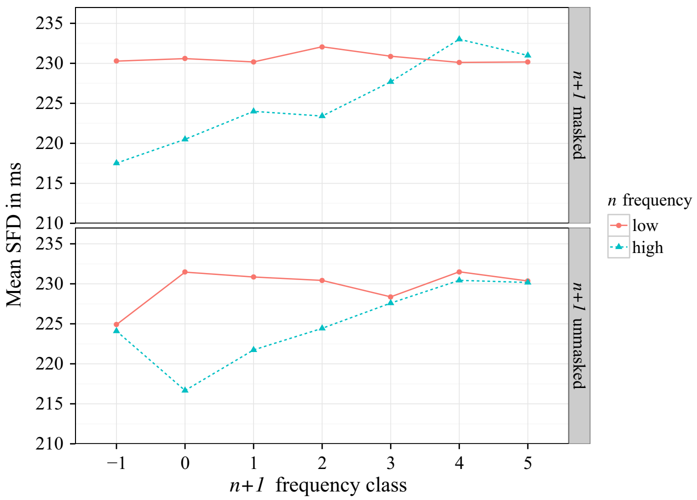

Fig. A2. Predicted partial effect of $n + 1$ frequency, n frequency, and $n + 1$ mask on single fixation duration (from the fitted LMMs using the *remef* function by Hohenstein & Kliegl, 2014). $n + 1$ frequency class is determined by taking the log 10 of $n + 1$ frequency per million, rounding to the nearest integer.

Appendix B: Sentence stimuli

Target words are in parentheses. For each sentence frame, the 72 subjects in the group without the frequency manipulation always saw the target word in *italics*.

1. The child pestered the (green/*timid*) fish that was hiding behind the pondweed.
2. The kindergarten students thought the (*green*/timid) bird at the zoo was beautiful.
3. After the attack, the formerly (peaceful/*tranquil*) island was in chaos.
4. She designed the (*peaceful*/tranquil) garden behind her house herself.
5. The decorators worked on the (royal/*dingy*) home until it looked wonderful.
6. They had difficulty selling the (*royal*/dingy) house in the current market.
7. Lynn succeeded in her new (business/*vocation*) and made a good salary.
8. My father loved his (*business*/vocation) and always enjoyed working.
9. Cloe noticed that her (white/*beige*) blouse had a red stain on the left sleeve.
10. Thomas purchased the (*white*/beige) truck from his neighbor for next to nothing.
11. The cameraman quietly followed the (animal/*gorilla*) from a safe distance.
12. Nobody expected that the (*animal*/gorilla) would be so friendly and docile.
13. Julie met the (young/*suave*) artist at the gallery and quickly fell in love.

14. Everyone thought the (*young*/suave) star of the action movie was arrogant.
15. Maggie went shopping for a (dress/*scarf*) at the new department store downtown.
16. Lynn borrowed the yellow (*dress*/scarf) from her older sister Patty.
17. Phil managed to read the (short/*pithy*) book within a few days.
18. Dr. Leon was nervous about giving the (*short*/pithy) speech at the fund raising banquet.
19. The carpenter examined the (middle/*shoddy*) door that had to be mended.
20. The workers replaced the (*middle*/shoddy) window on the second floor of the house.
21. Sheep grazed on the (quiet/*rival*) farm across the road.
22. Ray noticed that the (*quiet*/rival) coach had been ten minutes late to the game.
23. Take your money out of the (account/*satchel*) and pay the debt.
24. Christine put the money in her (*account*/satchel) for safe keeping.
25. I told Betty about the long (argument/*splinter*) that I had last week.
26. Tom said that the painful (*argument*/splinter) caused him much distress.
27. Deb did not read the (chapter/*tabloid*) when I told her what was in it.
28. Rick opened up the (*chapter*/tabloid) and read it out loud to his wife.
29. Fiona stopped to pick up the (children/*broccoli*) after work.
30. Betty says that she hates (*children*/broccoli) but it is not true.
31. Mark went to the (department/*pharmacist*) for help with his problem.
32. Mary asked the (*department*/pharmacist) to recommend a possible solution.
33. My mom said that a degree in (history/*zoology*) would be very helpful.
34. Beth wanted to study (*history*/zoology) at a college in Canada.
35. Pete climbed to the top of the large (machine/*trellis*) and cleaned it.
36. John was able to repair the broken (*machine*/trellis) very quickly.
37. Being a good (teacher/*educator*) is a noble profession for anyone.
38. Tony was a great (*teacher*/educator) and took pride in helping people.
39. Tara realized she had left her (picture/*mascara*) in her other handbag.
40. Pam picked up the discarded (*picture*/mascara) off the dirty floor.
41. Tina witnessed the clumsy (president/*ballerina*) fall off of the stage.
42. Kathy disliked the snobby (*president*/ballerina) and refused to say hello.
43. Clive hates studying (science/*algebra*) because he finds it very hard.
44. Mr. Jones taught (*science*/algebra) because he loved the subject.
45. Sue said that the (service/*cuisine*) is bad at that restaurant.
46. I enjoyed the great (*service*/cuisine) at the local Indian restaurant.
47. Chris was sad about his (situation/*deformity*) and refused to see guests.
48. Michael's odd (*situation*/deformity) was the topic of many conversations.
49. The neighbor's loud (telephone/*accordion*) really bothered Barbara.
50. The sudden sound of her friend's (*telephone*/accordion) woke Valerie up.
51. My grandma's (trouble/*amnesia*) started right after her sixtieth birthday.
52. The physician said that Dad's (*trouble*/amnesia) was only temporary.
53. Ralph rested in the (village/*hammock*) before he started on his trip.
54. Sam sat down in the small (*village*/hammock) since he was really tired.
55. Ruth admired the ring's (quality/*emerald*) and asked how much it cost.
56. We knew from the (*quality*/emerald) that the necklace was expensive.
57. I had too much (success/*tequila*) very quickly and could not handle it.
58. It was nice of Robert to share his (*success*/tequila) with his brother.
59. The solicitor carefully studied the (public/*morbid*) case although it was very tedious.
60. The politician was troubled by the (*public*/morbid) affair and resigned.
61. His favorite place to read was the (brown/*plush*) room overlooking the garden.
62. The couple wanted a (*brown*/plush) carpet to put in their livingroom.
63. She was captivated by the (first/*stark*) view across the cliff tops.
64. Calvin's grandfather owned the (*first*/stark) ranch in Southern Wyoming.
65. Joe blew out the candle and made a (quick/*hasty*) wish that he kept secret.
66. Margaret was stressed and her (*quick*/hasty) typing resulted in many typos.
67. After the battle the (whole/*irate*) army wanted to go home.
68. When the issue was discovered, the (*whole*/irate) jury threatened mistrial.
69. The small town was far away from the (modern/*hectic*) city where the decisions were made.
70. Many scientists believe the (*modern*/hectic) lives we lead can be harmful to our health.
71. The supporters cheered when the (local/*inept*) team finally won the match.
72. Michael was enraged when the (*local*/inept) cops pulled him over for no reason.
73. The hikers were concerned that the (mixed/*scant*) fruit wouldn't last very long.

74. The statistician analyzed the (*mixed*/scant) data before he consulted his manager.
75. According to the report, it should be (clear/*humid*) today and chilly tomorrow.
76. The long-range forecast was for a (*clear*/humid) week followed by a month of rain.
77. Jennifer had to meet with her (usual/*weird*) clients later this afternoon.
78. The first few days of Bill's (*usual*/weird) week were very busy.
79. The kindly old man suffered a (recent/*lethal*) fall when he was shopping.
80. The officer was investigating the (*recent*/lethal) attack of a businessman.
81. Karen knew that the (basic/*taboo*) fact would soon be discovered.
82. They were debating whether the (*basic*/taboo) actions of the officer warranted a reprimand.
83. The nurse tended to the (large/*pasty*) feet of the grumpy old woman.
84. Nobody wanted to look at the (*large*/pasty) belly of the man at the beach.
85. After Thomas finished eating his (daily/*bland*) cereal he had a cup of coffee.
86. The cook ordered the (*daily*/bland) food from the local market.
87. In the field, the (happy/*agile*) puppy was playing catch with a frisbee.
88. Outside the school the (*happy*/agile) girl skipped around the other children.
89. At the salon the stylist cut the (thick/*sleek*) hair for the regular customer.
90. The witness clearly remembered the (*thick*/sleek) beard of the murder suspect.
91. The experts concluded that the (small/*frail*) statue had been made in Rome.
92. Henry held his mother's (*small*/frail) hand as she walked down the steps.
93. The midwife admired the (famous/*serene*) baby after the difficult birth.
94. As they listened to the (*famous*/serene) song the elderly couple held hands.
95. The old man reminisced about his (great/*regal*) life while his family listened.
96. Heather read a book about a (*great*/regal) navy that ruled the high seas.
97. The artist criticized the (black/*askew*) line that was the focus of the painting.
98. He worked hard to create the (*black*/askew) grid for the art class project.
99. Tim never missed an episode of the (legal/*witty*) drama until it was moved to a new time.
100. The radio program broadcast (*legal*/witty) news each morning before midday.
101. Yesterday, Nichole made a (brief/*terse*) comment about her boss that was scandalous.
102. Sam remembered to write a (*brief*/terse) note before he left the house.
103. Many in the audience didn't understand the (final/*focal*) point of his argument.
104. Jack hated the essay and knew that the (*final*/focal) part would have to be rewritten.
105. Dave regretted his (sorry/*jaded*) past despite his recent successes.
106. They could tell by the (*sorry*/jaded) look on his face that the meeting hadn't gone well.
107. Luciel said she enjoyed the (fresh/*manic*) music but not everyone thought it was good.
108. The team devised a (*fresh*/manic) plan with a strict schedule.
109. We tried to get the celebrity's (attention/*autograph*) but we could not.
110. Sandy got the star's (*attention*/autograph) after the movie was over.
111. Mary loved her little (brother/*terrier*) so much that she spoiled him.
112. Kim took care of her friend Bob's (*brother*/terrier) when he went away.
113. I read a book about a useless (character/*scoundrel*) who no one liked.
114. My uncle is a strange (*character*/scoundrel) and is not very trustworthy.
115. The man was brought to the (council/*dungeon*) after committing a crime.
116. Tim was scared to see the (*council*/dungeon) so he decided not to go.
117. Mark received his (education/*doctorate*) from a prestigious university.
118. Chris wanted to finish his (*education*/doctorate) soon and get a job.
119. Ron discussed the painful (experience/*amputation*) that he recently had.
120. Bob had a horrible (*experience*/amputation) that left him very weak.
121. We were unable to repair the (marriage/*ligament*) even though we tried.
122. I was sad that my brother's (*marriage*/ligament) could not be fixed.
123. Val needed some (material/*scissors*) before she could start the project.
124. Please bring me the (*material*/scissors) and a needle right away.
125. Dan said that the extensive (practice/*tutorial*) helped him on the exam.
126. Melanie attended the lengthy (*practice*/tutorial) yesterday afternoon.
127. The boy could not solve the (problem/*anagram*) so he asked for help.
128. The teacher gave a difficult (*problem*/anagram) as the final question.
129. We saw the entire (process/*autopsy*) being performed by the expert.
130. Paul asked whether the (*process*/autopsy) would take a long time.
131. I learned a lot from the wise (professor/*astronaut*) who spoke in class.
132. I would like to be a respected (*professor*/astronaut) when I get older.
133. The scientist's (research/*abstract*) was submitted to a conference.

134. Jane found the (*research*/abstract) interesting and wanted to read more.
135. The burglar shot an innocent (secretary/*bystander*) during the robbery.
136. The journalist interviewed an attractive (*secretary*/bystander) for the story.
137. Jane yelled as the (student/*sparrow*) fell out of a tree in the park.
138. Duncan thought that the (*student*/sparrow) was small for his age.
139. I heard that the (company/*brewery*) made a large profit this past year.
140. We toured a local (*company*/brewery) and wrote a report for our course.
141. Sara rushed her (husband/*toddler*) to the hospital after he hurt himself.
142. Liz showed us a photo of her (*husband*/toddler) during the lunch.
143. Mark asked for some (support/*aspirin*) when he was not feeling well.
144. I need some (*support*/aspirin) before I can finish this long project.
145. I took a tour of a famous (building/*catacomb*) while I was on holiday.
146. The police closed off the dangerous (*building*/catacomb) yesterday.
147. Rebecca soaked the (surface/*platter*) with soap to get the grease off.
148. Please clean the filthy (*surface*/platter) before you put food on it.
149. The scientists created a new (technology/*camouflage*) for the military.
150. Soldiers are now safer due to better (*technology*/camouflage) and weapons.
151. Jennifer thought her roommate's outfit was (ideal/*lousy*) for a first date.
152. Frank told all his friends about the (*ideal*/lousy) beer he had at the irish pub.
153. He assumed the victim had been struck with a (sharp/*blunt*) object just behind his ear.
154. The mechanic used the (*sharp*/blunt) tool to complete the difficult repair.
155. Pamela was dating a (doctor/*banker*) who drove a convertible BMW.
156. Gary was afraid that the (*doctor*/banker) would give him bad news.
157. The college student was reading about (common/*bizarre*) medical disorders for her class.
158. The veterinarian specialized in treating (*common*/bizarre) pets and farm animals.
159. Crystal found the (ancient/*jagged*) arrowhead during the archeological dig.
160. Kevin knew he had to be very careful with the (*ancient*/jagged) knife from Egypt.
161. The wealthy executive fired the (worker/*butler*) who had been late three days in a row.
162. Nobody was surprised when the friendly (*worker*/butler) received a big raise.
163. Bob was late because he missed the early (train/*ferry*) and had to take the next one.
164. Lisa decided to take the (*train*/ferry) so she could avoid the traffic on the bridge.
165. John proudly displayed the (beautiful/*majestic*) statue in his living room.
166. The portrait of the (*beautiful*/majestic) eagle was purchased at the auction by the collector.
167. Henry found the red (chair/*futon*) at a yard sale last weekend.
168. Nichole thought that Bob's new (*chair*/futon) was more comfortable than his old one.
169. James just built a (stone/*brick*) patio in his backyard.
170. Lucy followed the long (*stone*/brick) path around the side of the lodge.
171. Theodore eventually found the (right/*bleak*) road late at night.
172. Susan thought the (*right*/bleak) painting had to be in the gallery.
173. The residents of the (little/*ornate*) town voted for a new mayor.
174. My mother purchased the (*little*/ornate) cottage for a summer home.
175. Behind the apartments there is a (level/*rowdy*) area where teenagers play football.
176. Near the University there is a (*level*/rowdy) field where students practice rugby.
177. The teacher waited in the (bright/*barren*) hall before the school assembly.
178. The couple considered the (*bright*/barren) yard before placing an offer on the house.
179. The technician compared the (broken/*dilute*) half with the remaining substance.
180. The investigator examined the (*broken*/dilute) portion for any signs of tampering.
181. Mike found the (rapid/*brisk*) work tiresome but rewarding.
182. The squad didn't enjoyed the (*rapid*/brisk) task but the captain insisted it be done.
183. Liz hung a portrait on her (empty/*mauve*) wall to add some decoration to the room.
184. Ginger was careful with the (*empty*/mauve) bowl after she learned how old it was.
185. The waitress was a (friendly/*sociable*) woman and immediately took their orders.
186. For the position, only competent and (*friendly*/sociable) candidates will be considered.
187. He adamantly refused to accept (different/*divergent*) opinions from anyone, even his aides.
188. She instructed the teams to test (*different*/divergent) strategies and compare the results.
189. According to Keith, any short (straight/*uncoiled*) cable can easily serve as an antenna.
190. They used the (*straight*/uncoiled) cord to connect their phone.
191. The thief is reported to be a moderately (heavy/*obese*) person with a bald head.
192. With great effort he managed to carry his (*heavy*/obese) brother upstairs.

Appendix C. Random effects in the corpus analysis

This table describes the random effects for the LMM with GD as the dependent variable. Note that correlations between the random effects were set to 0 to reduce model complexity.

Group	Random effect	*SD*
Word	(Intercept)	5.41×10^{-2}
	$n+1$ length	2.89×10^{-2}
	$n+1$ frequency	2.22×10^{-2}
	n frequency	7.18×10^{-3}
	$n+1$ predictability	2.10×10^{-2}
	$n+1$ length by $n+1$ frequency	1.52×10^{-2}
	n frequency by $n+1$ frequency	8.50×10^{-5}
Subject	(Intercept)	1.24×10^{-1}
	$n+1$ length	4.56×10^{-3}
	$n+1$ frequency	6.77×10^{-3}
	n frequency	3.08×10^{-2}
	$n+1$ predictability	7.76×10^{-3}
	$n+1$ length by $n+1$ frequency	0.00
	n frequency by $n+1$ frequency	0.00
Residual		2.76×10^{-1}

References

Abad, M. J. F., Noguera, C., & Ortells, J. J. (2003). Influence of prime-target relationship on semantic priming effects from words in a lexical-decision task. *Acta Psychologica, 113*, 283–295.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1.

Angele, B., & Rayner, K. (2011). Parafoveal processing of word $n+2$ during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1210–1220. http://dx.doi.org/10.1037/a0023096.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating $n+1$ and $n+2$ previews simultaneously. *Visual Cognition, 16*, 697–707. http://dx.doi.org/10.1080/13506280802009704.

Angele, B., Tran, R., & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 526–538. http://dx.doi.org/10.1037/a0029492.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2014). lme4: Linear mixed-effect models using S4 classes (R package Ver. 1.1-7). *Vienna, Austria: R Foundation for Statistical Computing*.

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1162–1172. http://dx.doi.org/10.1037/0096-1523.25.4.1162.

Blanchard, H. E., Pollatsek, A., & Rayner, K. (1989). The acquisition of parafoveal word information in reading. *Perception & Psychophysics, 46*, 85–94.

Bradshaw, J. L. (1974). Peripherally presented and unreported words may bias the perceived meaning of a centrally fixated homograph. *Journal of Experimental Psychology, 103*, 1200–1202.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). Hillsdale, NJ: Academic Press Inc..

Cui, L., Drieghe, D., Yan, G., Bai, X., Chi, H., & Liversedge, S. P. (2013). Parafoveal processing across different lexical constituents in Chinese reading. *The Quarterly Journal of Experimental Psychology, 66*, 403–416. http://dx.doi.org/10.1080/17470218.2012.720265.

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *Oxford Handbook on Eye Movements* (pp. 839–855.). Oxford University Press.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology, 61*, 1239–1249. http://dx.doi.org/10.1080/17470210701467953.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636. http://dx.doi.org/10.1016/S0042-6989(01)00301-7.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813. http://dx.doi.org/10.1037/0033-295X.112.4.777.

Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *Quarterly Journal of Experimental Psychology, 66*, 1548–1571.

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch *er* "there" as a higher-level expectancy monitor. *Acta Psychologica, 130*, 153–160.

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology, 47*, 201–221.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 166–190. http://dx.doi.org/10.1037/a0033670.

Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology, 16*, 104–127.

Inhoff, A. W. (1982). Parafoveal word perception: A further case against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 137–145.

Inhoff, A. W., & Rayner, K. (1980). Parafoveal word perception: A case against semantic preprocessing. *Perception & Psychophysics, 27*, 457–464.

Inhoff, A. W., Starr, M., & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception & Psychophysics, 62*, 1474–1484.

Kennedy, A., Murray, W., & Boissiere, C. (2004). Parafoveal pragmatics revisited. *European Journal of Cognitive Psychology, 16*, 128–153.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45*, 153–168.

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology, 55A*, 1307–1337.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 149–179). Oxford, UK: Elsevier.

Kirkby, J. A., Webster, L. A. D., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin, 134*, 742–763. http://dx.doi.org/10.1037/a0012979.

Kliegl, R. (2007). Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General, 136*, 530–537.

Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review, 12*, 132–138.

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

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word $n+2$. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250–1255.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General, 143*, 895–913. http://dx.doi.org/10.1037/a0033580.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578–587.

Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research, 27*, 1–22.

Murray, W. S. (1998). Parafoveal pragmatics. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 181–199). Oxford, UK: North-Holland/Elsevier.

Ortells, J. J., Abad, M. J. F., Noguera, C., & Lupiáñez, J. (2001). Influence of prime-probe stimulus onset asynchrony and prime precuing manipulations on semantic priming effects with words in a lexical-decision task. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 75–91.

Ortells, J. J., & Tudela, P. (1996). Positive and negative semantic priming of attended and unattended parafoveal words in a lexical decision task. *Acta Psychologica, 94*, 209–226.

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research, 49*, 1994–2000.

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology, 16*, 178–202.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*, 65–81. http://dx.doi.org/10.1016/0010-0285(75)90005-5.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372.

Rayner, K. (2009a). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*, 1457–1506. http://dx.doi.org/10.1080/17470210902816461.

Rayner, K. (2009b). Eye movements in reading: Models and data. *Journal of Eye Movement Research, 2*, 1–10.

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition, 22*, 242–258. http://dx.doi.org/10.1080/13506285.2013.879084.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*, 468–469.

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research, 38*, 1129–1144.

Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word $n + 2$? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 230–245. http://dx.doi.org/10.1037/0096-1523.33.1.230.

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., & Reichle, E. D. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General, 136*, 520–529.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1290–1301. http://dx.doi.org/10.1037/0278-7393.30.6.1290.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics, 27*, 537–544.

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

Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 213–234). Oxford, UK: Elsevier.

Reichle, E. D. & Drieghe, D. (in press). Using E-Z Reader to examine the consequences of fixation–location measurement error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/a0037090.

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. http://dx.doi.org/10.1037/0033-295X.105.1.125.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4–22. http://dx.doi.org/10.1016/j.cogsys.2005.07.002.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1–21. http://dx.doi.org/10.3758/PBR.16.1.1.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*, 34–55.

Risse, S., Engbert, R., & Kliegl, R. (2008). Eye-movement control in reading: Experimental and corpus-analytic challenges for a computational model. In K. Rayner, D. Shen, X. Bai, & G. Yan (Eds.), *Cognitive and cultural influences on eye movements* (pp. 65–91). Tianjin, China: Tianjin People's Publishing House.

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word $n + 2$ in reading. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1026–1042. http://dx.doi.org/10.1037/a0027735.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*, 391–421. http://dx.doi.org/10.1080/13506285.2012.670143.

Schotter, E. R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language, 69*, 619–633. http://dx.doi.org/10.1016/j.jml.2013.09.002.

Schotter, E., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*, 5–35. http://dx.doi.org/10.3758/s13414-011-0219-2.

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

Schroyens, W., Vitu, F., Brysbaert, M., & d' Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *Quarterly Journal of Experimental Psychology, 52A*, 1021–1046.

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1924–1938. http://dx.doi.org/10.1037/a0024322.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 205–223. http://dx.doi.org/10.1037/0096-1523.34.1.205.

Winskel, H., & Perea, M. (2014). Can parafoveal-on-foveal effects be obtained when reading an unspaced alphasyllabic script (Thai)? *Writing Systems Research, 6*, 94–104. http://dx.doi.org/10.1080/17586801.2013.843440.

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology, 66*, 548–562. http://dx.doi.org/10.1080/17470218.2011.625094.