What are the costs of degraded parafoveal previews during silent reading?

Martin R. Vasilev[1]

Timothy J. Slattery

Julie A. Kirkby

Bernhard Angele

Bournemouth University

Corresponding author[1] at:

Department of Psychology

Bournemouth University

P104, Poole House

Talbot Campus, Fern Barrow

Poole, Dorset, BH12 5BB

United Kingdom

Phone: +44 7835202606

Email: mvasilev@bournemouth.ac.uk

Acknowledgements

We thank Stefan Hawelka and Christina Marx for kindly providing us with the original scripts used for the visual degradation.

Abstract

It has been suggested that the preview benefit effect is actually a combination of preview benefit and preview costs. Marx et al. (2015) proposed that visually degrading the parafoveal preview reduces the costs associated with traditional parafoveal letter masks used in the boundary paradigm (Rayner,1975), thus leading to a more neutral baseline. We report two experiments of skilled adults reading silently. In Experiment 1, we found no compelling evidence that degraded previews reduced processing costs associated with traditional letter masks. Moreover, participants were highly sensitive to detecting degraded display changes. Experiment 2 utilized the boundary detection paradigm (Slattery, Angele, & Rayner, 2011) to explore whether participants were capable of detecting actual letter changes or if they were responding purely to changes in degradation. Half of the participants were instructed to respond to any noticed display changes; the other half were instructed to respond only to changes in letter identities. Participants were highly sensitive to degraded changes. In fact, these changes were so apparent that they reduced the sensitivity to letter masks. In the context of the model proposed by Angele, Slattery, and Rayner (2016), we suggest that degraded previews interfere with the attentional stage, as evidenced by the general lack of foveal load effects. In summary, we found that increasingly degrading parafoveal letter masks does not reduce their processing costs in adults, but that both degraded valid and invalid previews introduce additional costs in terms of greater display change awareness.

Keywords: reading, parafoveal processing, eye-movements, display change detection

During reading, information is obtained not only from the currently fixated word, but also from the word to the right of fixation (Rayner, 1998). When the word to the right is previewed in parafoveal vision, this leads to shorter fixation durations once that word is fixated in foveal vision (the so-called *preview benefit* effect). The preview benefit is usually calculated by subtracting fixation durations when parafoveal preview is accurate (i.e. participants have an identical preview of the word) from fixation durations when parafoveal preview is inaccurate (e.g. the target word was masked with a string of letters).

Although the preview benefit is a well-established finding, it has recently been argued that the size of this effect reflects not only the benefit of having a valid preview of the target word, but also processing costs associated with the type of parafoveal mask. Kliegl, Hohenstein, Yan, and McDonald (2013) showed that the traditional preview benefit is a complex mixture of preview benefits and preview costs, and that both can vary with the length of the gaze duration on the pre-boundary word and the amplitude of the saccade to the target word. In addition to this, Hutzler et al. (2013) used fixation-related brain potentials to demonstrate that parafoveal X masks interfere with foveal word recognition. Therefore, these findings raise the question of what is an adequate baseline for considering the preview benefit effect.

In one study, a novel parafoveal preview manipulation was introduced where the letters are visually degraded by replacing some of the black pixels (Gagl, Hawelka, Richlan, Schuster, & Hutzler, 2014). Following this, further research has utilized the administration of visual degradation to study parafoveal processing in developing readers (Marx, Hawelka, Schuster, & Hutzler, 2015; Marx, Hutzler, Schuster, & Hawelka, 2016). For example, Marx et al. (2015) compared visually degraded parafoveal masks to traditional parafoveal masks such as a string of letters or a string of Xs. Their results corroborated the findings of Kliegl et al. (2013) and Hutzler et al. (2013) that these traditional masks are associated with preview costs. Furthermore, their results suggested that, by using an incremental boundary technique where the valid parafoveal preview is increasingly degraded (cf. Jacobs, Grainger, & Ferrand, 1995), it is possible to achieve an accurate estimate of the preview benefit.

Based on their results, Marx et al (2015, p. 559) argued that "Future studies on the emergence of parafoveal preprocessing in beginning readers may consider a salience manipulation of valid previews as an alternative to using parafoveal masks." This was based on the finding that degraded valid previews resulted in longer gaze durations on the target word compared to valid previews- a finding that they interpret as indicating reduced parafoveal preview of this word. However, what we find far more interesting about the results of Marx et al.'s experiment is that they imply degrading invalid masks actually reduces their costs, as evidenced by decreased gaze durations.

In the current study, we explored whether the administration of visual degradation to invalid preview masks can reduce their costs in adults and whether visually degrading any type of parafoveal preview (valid or not) may introduce additional costs to sentence processing. In this respect, the present research was not an attempt to compensate for any potential shortcomings of parafoveal masks. Rather, we aimed to investigate whether the preview costs of parafoveal masks demonstrated in Marx et al.'s study can be reliably replicated in adults during silent reading, and whether the general procedure of administering visual degradation may introduce any costs of its own.

Foveal Load and the Preview Benefit

One important effect related to preview benefit in display change studies is that of foveal load (Henderson & Ferreira, 1990). When the processing demands of foveal information are high (i.e. the fixated word is low frequency), less information is obtained from parafoveal processing. The E-Z Reader model of eye-movement control (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006; Reichle, Warren, & McConnell, 2009) simulates this effect by decoupling the movement of attention needed for lexical processing from the oculomotor planning of saccades. The model includes two stages of lexical processing (L1 and L2) and two stages of oculomotor processing (M1 and M2). The time needed to complete each of the lexical stages is a function of lexical frequency and word predictability, but the duration of the oculomotor stages are not affected by these two variables. The L1 stage corresponds to a familiarity check on the currently processed word, which begins once attention is allocated to that word. When stage L1 is completed, it prompts the start of both the L2 (lexical completion) and M1 (labile saccade planning) stages. For valid previews, the size of the preview benefit in this model is determined by the race between L2 and the sum of M1 and M2. When stage L2 is completed quickly relative to M1 and M2, covert attention is shifted to the next word before the arrival of the eyes, thus leading to greater preprocessing of this next word (i.e. larger preview benefits). However, foveal load also predicts that any costs associated with invalid previews will be larger when processing demands at fixation are lower (i.e. foveal load is low).

Preview benefit effects in reading can also be explained by models that assume parallel processing of words. In the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005), several words can be processed in parallel via a spatially-distributed activation field. In this model, the rate of lexical processing is determined by the distance of the word from the current point of fixation. Therefore, processing of the upcoming word in parafoveal vision starts as soon as this word falls within the activation field. A more recent version (Schad & Engbert, 2012) has implemented a zoom-lens-of-attention mechanism in which highly activated foveal words can make the processing span narrower in size. This mechanism has been offered as a theoretical account of the foveal load effect (Henderson & Ferreira, 1990).

Display Change Detection and the Preview Benefit

In addition to foveal load and the properties of the parafoveal mask, another factor that may influence the size of the preview benefit effect is the awareness of the display change that happens on the screen during invalid preview conditions. Due to saccadic suppression (Matin, 1974), the display change is usually not perceived. This is also consistent with the finding that a change of the case of letters in a sentence is not perceived during a saccade (McConkie & Zola, 1979). However, as it is often reported in more recent boundary studies, there are usually at least some participants who are aware of the display changes. Therefore, one important question is whether and to what extent this awareness influences the preview costs.

Inhoff, Starr, Liu, and Wang (1998) were first to address this issue by systematically manipulating the refresh rate of the monitor (and thus the timing of the display change). Their findings indicated that the preview benefit effect is not compromised by the display change timing. In addition to this, White, Rayner, and Liversedge (2005) performed an analysis in which they separated participants into two groups: those who detected display changes and those who did not. Their findings showed that preview effects were larger in size for participants who noticed display changes, and that these participants did not show a foveal load effect.

Instead of classifying participants according to whether they reported noticing display changes after the experiment, Slattery, Angele, and Rayner (2011) implemented a boundary detection task in which participants are asked after every trial if they noticed a display change. Their key finding was that the closer participants' fixation was to the boundary before crossing it, the higher their detection rate. Additionally, detection increased with the delay in display change timing.

More recently, Angele, Slattery, and Rayner (2016) used the same paradigm to investigate both foveal load and properties of the parafoveal preview on detection sensitivity and fixation time measures. They manipulated foveal load by embedding either a high or low frequency word immediately prior to the boundary and manipulated the preview to be either word-like or nonword-like. They found that readers are more sensitive to detecting display changes from nonword-like compared to word-like previews. Moreover, display change detection was not influenced by their foveal load manipulation of word frequency. However, the foveal load manipulation still resulted in greater preview benefit for high compared to low frequency words. Additionally, they reported large parafoveal-on-foveal (PoF) effects on the pre-boundary word only for trials in which participants actually detected a display change. These PoF effects were found on gaze durations and reflected inhibition (i.e. slowing down) of foveal word processing of the pre-boundary word due to seeing letter masks in parafoveal vision. To account for these effects, they proposed a two-stage model of parafoveal processing. The first stage is pre-attentive and can proceed concurrently with lexical processing of the foveal word. This stage is sensitive to the raw visual parafoveal information and can give rise to PoF effects. The second stage is attention-dependent and is responsible for generating parafoveal preview benefits and costs.

Present Study

The administration of visual degradation to invalid previews has indicated that conventional parafoveal masks introduce processing costs and thus may not represent a neutral baseline for studying parafoveal preview effects (Marx et al, 2015). However, as Marx et al.'s study was done with children, it is currently not known if the same processing costs can also be observed in adult readers. This is an important question as the overwhelming majority of studies that have used the boundary paradigm (Rayner, 1975) to study parafoveal processing in reading have been conducted with adults. For this reason, establishing such potential costs is of major theoretical importance as there are dozens, if not hundreds, of studies that have used conventional masks to study parafoveal processing during reading (see Vasilev & Angele, 2016 for a recent review). Additionally, Marx et al.'s participants were reading out aloud, while most research on parafoveal processing in adults has been done in a silent reading paradigm. Therefore, another aim was to investigate whether these results can also be extended to adults reading silently.

One interesting aspect of administering visual degradation to valid or invalid parafoveal previews is that this procedure involves changing the stimulus quality of the text. Due to the fact that recent studies (Marx et al., 2015, 2016) have administered degradation not only to the target word, but also to the whole remaining sentence, it is currently not known if this degradation results in any costs of its own. For example, it has long been known that reducing the stimulus quality of the text results in costs to processing in both visual word recognition and reading paradigms (Borowsky & Besner, 1993; Norris, 1984; O'Malley, Reynolds, & Besner, 2007; Reingold & Rayner, 2006; White & Staub, 2012; Yap & Balota, 2007). Therefore, a second aim of the present research was to test whether visual degradation may introduce any costs to processing due to the fact that it affects the stimulus quality of the sentence.

Our first experiment was a conceptual replication of the study by Marx et al. (2015) and it aimed to investigate preview costs associated with using traditional parafoveal masks. More specifically, we attempted to replicate Marx et al.'s (2015) findings that visually degrading the parafoveal preview reduces the processing costs associated with the mask and gives a more neutral (i.e. smaller) preview benefit effect. To do this, we used two types of parafoveal preview: a same-shape different-letter mask (henceforth "letter mask" for simplicity) and the same letter mask that was visually degraded. We used two levels of visual degradation: 10 and 20%. These were the same levels used in Marx et al.'s study and we also adopted them here for consistency purposes.

In our second experiment, we sought to explore a different type of costs: namely, those related to participants' sensitivity to detect display changes. Would readers be more likely to detect display changes after degraded compared to non-degraded parafoveal previews and would any such detections be accompanied by large PoF effects? To investigate this, we employed a boundary detection paradigm where participants had to indicate if they noticed changes on the screen after each trial (Angele et al., 2016; Slattery et al., 2011). This experiment included three types of parafoveal preview: letter mask, letter mask degraded by 20% and identical preview degraded by 20%. In both experiments, we also manipulated the lexical frequency of the pre-target word (i.e. high or low).

Experiment 1

Method

Participants. Thirty-two[1] undergraduate students participated for course credit. Their mean age was 20.1 years (SD: 4.6; range: 18-44). Participants were native speakers of British English and reported normal or corrected-to-normal vision, as well as no prior diagnosis of reading disorders. Participants were naïve about the purpose of the experiment. Ethical approval for the study was obtained from Bournemouth University (ID #9160).

Materials. The stimuli consisted of 136 English sentences. There were 17 sentences per condition and the presentation of sentences and conditions was counterbalanced with a full Latin square design. In each sentence, there was a pre-target word (N-1) for which the lexical frequency was manipulated, and a target word (N) whose parafoveal preview was manipulated. The mean length of the pre-target word was 5.8 letters (SD= 1.5) and the mean length of the target word was 5.9 letters (SD= 2). High and low frequency pre-target words were matched on word length for each sentence. The position of the target region was varied, but there were always at least two words before and after it. Pre-target words were contrasted on the frequency count of the SUBTLEX database (van Heuven, Mandera, Keuleers, & Brysbaert, 2014); however, for completeness, frequencies from the British National Corpus (BNC) are also reported (see Table 1). The high frequency words had significantly higher log frequency than the low frequency words on the SUBTLEX database ($t(266) = 39.15$, $p < 0.001$), its Zipf scale ($t(267) = 39.34$, $p < 0.001$), and the BNC ($t(269) = 32.42$, $p < 0.001$).

(Insert Table 1 here)

Parafoveal preview of word N was manipulated with the boundary paradigm (Rayner, 1975). The invisible boundary was placed at the first pixel after the last letter of word N-1. There

[1] Two more participants were tested but they were replaced for the following reasons: one participant had low comprehension score (76%), and another one contributed less than 50% of trials to the data analysis.

were 4 types of word N preview: 1) identical, 2) a random letter mask that preserves the shape of letters (in terms of ascenders and descenders), 3) the same letter mask degraded by 10%, and 4) the same letter mask degraded by 20%. An example of the parafoveal preview conditions is shown in Figure 1. Degradation refers to the random exchange of black pixels with white pixels, which was done using the same script as in Marx et al. (2015). For the previews that contained degradation, the whole sentence after word N was also degraded. This was done to follow the procedure in Marx et al. (2015) as closely as possible, since the authors of this study also degraded the rest of the sentence.

(Insert Figure 1 here)

Apparatus. Participants' eye movements were recorded with an Eyelink 1000 eye-tracker. Viewing was binocular, but only the right eye was recorded. The sampling frequency was 1000 Hz (i.e. the position of the right eye was sampled every 1 ms). The resolution noise was < 0.01° and the velocity noise was < 0.5° on average. Participants rested their head on a chin-and-forehead rest in order to minimize the occurrence of head-movement artifacts.

The experiment was run using the EyeTrack.0.7.10h software (http://blogs.umass.edu/eyelab/). The stimuli were presented on an Iiyama Vision Master Pro 510 monitor with a screen resolution of 1024 x 768 pixels and a refresh rate of 150 Hz. The sentences were formatted in a bold, monospaced font that appeared as black text over white background. The sentences appeared on a single line in the middle of the screen. The number of pixels per letter was 11. The distance between participants and the monitor was 60 cm. The experiment was run on a PC with Window XP.

Procedure. Participants were tested individually in a session that lasted about 30-40 minutes. Before the actual testing, participants were informed about the experiment and were asked to sign an informed consent form. Prior to the experiment, a 3-point horizontal calibration was performed. Additionally, there was a drift check before each trial and participants were re-calibrated as necessary. The calibration error was kept at < 0.3°. Sentences were presented one by one on the screen. Each sentence fitted on a single line. Participants had to fixate on a gaze box on the left side of the screen, which triggered the presentation of the sentence. Participants indicated that they had finished reading a sentence by pressing a button on a joystick. There was a yes/no question following 33% of the sentences. After the experiment was finished, participants were asked if they noticed anything unusual on the screen while doing the experiment and were then debriefed.

Data analysis. The experiment had a within-subject design with two factors: word N-1 frequency (high vs. low) and parafoveal preview of word N (identical, random letter mask, random letter mask with 10% degradation, random letter mask with 20% degradation). Two measures of fixation durations were used as dependent variables: first fixation duration (FFD), and gaze duration (GD). FFD refers to the very first fixation on the word, while GD also includes all fixations before the eyes move to another word.

Statistical analysis of the data was done with Linear Mixed Models (LMMs) by using the “lme4” package v.1.1-12 (Bates, Mächler, Bolker, Walker, 2015) in R 3.3.0 (R Core Team, 2016) and RStudio (2015). Word N-1 frequency and parafoveal preview of word N were entered as fixed effects in the models. Random intercepts, as well as random slopes for word N-1 frequency and word N preview, were added for both participants and items (Baayen, Davidson, & Bates, 2008). This random-effects structure is consistent with the suggestion to specify a random slope for the predictors of theoretical interest (Barr, Levy, Scheepers, & Tily, 2013). Fixation durations were log-transformed in all models. This was done because LMMs assume that the model residuals are normally distributed (e.g. see Gelman & Hill, 2006), and log-transformation improved the distribution of residuals in our models.

Because the present study predicts differences in raw fixation durations, it can be argued that analyses on the raw data may be more theoretically appropriate. Recently, Lo and Andrews (2015) have suggested that using GLMMs with the raw data can satisfy this theoretical assumption. We repeated our analyses with inverse-Gaussian GLMMs, but our main conclusions remained unchanged. For this reason, we report the GLMM results in the Supplementary files and discuss there the few differences that were found. We also report in the Supplementary files post-hoc tests that investigated whether parafoveal preview effects are modulated by saccade launch site (cf. Kliegl et al., 2013).

The following contrasts were used. In the models that calculated the preview benefit effects, a treatment contrast was used where each invalid preview condition was compared to the valid preview condition. High frequency words were coded as -0.5 and low frequency words were coded as 0.5. In the models used to compare letter mask previews with degraded letter mask previews, letter masks previews were coded with 0.5 and degraded letter mask previews were coded as -0.5. This was motivated by the suggestion of Marx et al. (2015) that degraded previews introduce less processing costs compared to conventional masks such as a string of letters. If this proposition is true, the LMM slopes should be positive. Similarly, in the comparison between 10 and 20% degraded previews, the former were coded as 0.5 and the latter as -0.5. The reason was again that increasing the level of degradation should further decrease the processing costs. Results were considered statistically significant if the |t| values were ≥ 1.96. P values are also reported for completeness.

Results

All subjects had comprehension accuracy greater than 86% (mean= 93.4%). There were no significant differences in comprehension accuracy across the 8 conditions, $F_1(7, 217)= 0.57$, $MSE= 0.01$, $p= .78$, $\eta^2_G = 0.02$; $F_2(7, 308)= 0.82$, $MSE= 0.01$, $p= .57$, $\eta^2_G = 0.01$. A GLMM analysis of the accuracy data by specifying both participants and items as random intercepts (or slopes) was not possible due to convergence failure. Sixty-two percent[2] of participants reported seeing display changes after the experiment. Only one participant saw both the random letter preview and the degraded previews; all other participants reported seeing the degraded previews only. Participants estimated seeing between 10 and 75 % of trials with display changes (mean= 45.5; SD= 22.6). The actual number of trials with degraded previews was 50%.

Trials with blinks on word N-1 or word N during first-pass reading were excluded (12.6 %). Moreover, trials in which the display change was triggered prematurely or was completed after fixation onset of word N were also removed (14.1%). Further 0.8% of trials we discarded due to track losses. Fixations shorter than 80 ms that occurred within 1 letter of another fixation were combined with that fixation. Any remaining fixations shorter than 80 ms were discarded. Trials with fixation durations longer than 800 ms for FFD and 1600ms for GD were removed as outliers from all analyses. Because the average fixation duration in reading is about 225-250 ms (Rayner, 2009), the 800 ms cut-off is more than three times this number (the cut-off was doubled

[2] Only 34% of participants reported the display changes before they were debriefed about the experiment. However, when they were showed the stimuli, an additional 28% of participants reported having seen the degradation. Most of them did not report it initially because they thought “it was just my eyes” or “the monitor was buffering”. This suggests that the actual detection rate may be even higher if other participants noticed the degradation but did not report it during the debrief.

for GD to account for re-fixations). This is a standard cut-off that has been used in previous boundary experiments (e.g. Risse & Kliegl, 2014; Schotter, Lee, Reiderman, & Rayner, 2015). The data trimming resulted in the removal of 0.66% of observations in the target region.

Pre-target word (N-1). Descriptive statistics for word N-1 are presented in Table 2 and the results from LMMs are presented in Table 3. There were robust frequency effects measured with both FFD and GD. This shows that the frequency manipulation on word N-1 was effective. There were no statistically significant effects of the target word N preview on fixation durations on word N-1. In other words, there were no PoF effects associated with either the conventional letter mask preview or the degraded letter mask previews.

(Insert Table 2 here)

(Insert Table 3 here)

Target Word (N). Descriptive statistics for word N are presented in Table 4 and the results from LMMs are presented in Table 5. Preview effects were obtained with all three masks. These effects may represent preview benefits or a combination of preview benefits and preview costs (due to pre-processing of the masks). There were no statistically significant spillover effects of word N-1 frequency on fixation durations of word N. Additionally, there was no significant foveal load effect (i.e. an interaction between word N-1 frequency and preview benefit of word N; Henderson, & Ferreira, 1990).

(Insert Table 4 here)

(Insert Table 5 here)

Of greater interest in this experiment was how the degraded previews differ in comparison to the conventional non-degraded letter mask. The effect of preview saliency (i.e. the amount of degradation) is plotted in Figure 2 (cf. Marx et al., 2015, Figure 3). Compared to the letter mask previews, 10% degraded previews were shorter by 4 ms for FFD and 11 ms for GD, but the difference was not statistically significant (FFD: b=0.01, S.E= 0.02, t= 0.66, p= 0.52; GD: b=0.03, S.E= 0.02, t= 1.61, p= 0.12). Compared to the letter mask previews, 20% degraded previews were longer by 7 ms for FFD and shorter by 1 ms for GD, and the difference was again not statistically significant (FFD: b= -0.02, S.E= 0.02, t= -1.17, p= 0.25; GD: b=0.004, S.E= 0.02, t= 0.20, p= 0.84). However, this last comparison for GD interacted significantly with word N-1 frequency (b= 0.08, S.E= 0.03, t= 2.37, p= 0.02). In other words, GD was longer after letter mask previews compared to 20% degraded previews when word N-1 was low frequency, but shorter after letter mask previews compared to 20% degraded previews when word N-1 was high frequency. Follow-up tests on the simple effects indicated that the difference between letter mask previews and 20% degraded previews for GD was not statistically significant for either low frequency words N-1 (b= 0.04, S.E= 0.03, t= 1.5, p= 0.14) or high frequency words N-1 (b= -0.03, S.E= 0.02, t= -1.17, p= 0.25). However, it is possible that the t-tests used did not have enough statistical power to detect the simple effects.

(Insert Figure 2 here)

Finally, fixation durations for 20% degraded previews were compared to fixation durations for 10% degraded previews. Fixation durations were significantly longer after 20% degraded previews than after 10% degraded previews when measured with FFD (b=-0.04, S.E= 0.02, t= -2.43, p= 0.02), but not with GD (b=-0.03, S.E= 0.02, t= -1.51, p= 0.13). For all comparisons reported in this section, there were no other statistically significant interactions with word N-1 frequency (all $|t|$s ≤ 1.58).

Discussion

The results of Experiment 1 showed that using a visually-degraded parafoveal mask did not generally lead to a statistically significant change in fixation durations compared to the more conventional pseudo-random letter mask. There was an indication in the data that fixation durations after 10% degraded previews decrease for GD, but the same was not found for FFD. Additionally, previews degraded by 20% resulted in longer fixation durations compared to previews degraded by 10%. This latter result is contrary to the predictions of the incremental boundary technique, which would predict a decrease in fixation durations from 10 to 20% degradation. Therefore, the results do not corroborate Marx et al.'s (2015) finding that fixation durations decrease with increasing degradation of the parafoveal mask. In this sense, Experiment 1 did not find support for the proposition that the preview costs associated with traditional parafoveal masks can be decreased by administering increasing levels of visual degradation to such masks.

Two interesting results were that the degraded previews did not cause any significant PoF effects, and that we found no evidence of foveal load. One possible explanation for the lack of foveal load effect might be the awareness of display changes. Indeed, one noteworthy finding in Experiment 1 was that participants were aware of these changes to a much greater degree than what has been previously reported in the literature (e.g. Angele & Rayner, 2013; White et al., 2005). Also, the reports of participants were highly skewed towards noticing degraded previews. Therefore, it is possible that the present findings may have been influenced by the fact that the majority of participants noticed display changes and did so far more often in the degraded conditions. However, according to the two stage model proposed by Angele et al. (2016), the large extent of participants' awareness of display changes should have also resulted in large PoF effects.

In order to formally investigate the effect of display change sensitivity, Experiment 2 adopted the display change detection paradigm introduced by Slattery et al. (2011). Therefore, the critical difference from Experiment 1 was that, after every trial, participants also had to give a confidence rating on whether they saw changes on the screen. Experiment 2 had the same frequency manipulation on word N-1. Additionally, the parafoveal preview of word N was orthogonally manipulated by crossing two factors: preview type (identical vs letter mask) and degradation (non-degraded vs degraded by 20%). Only one degradation level was used in order to preserve statistical power after the addition of a new preview condition (identical degraded). Degradation level of 20% was chosen because Marx et al. (2015) concluded that, for identity previews, it "… prevents the extraction of (useful) parafoveal information." (p. 557). Therefore, 10% degradation level would not be suitable in this experiment because participants may still obtain a small amount of useful information from degraded identity previews.

One aim of Experiment 2 was to distinguish between two potential interpretations of the large number of reported display changes in Experiment 1: namely, are participants noticing anything degraded on the screen? Or are they capable of distinguishing a letter mask (degraded or not) from an identical but degraded preview of word N? Moreover, the addition of degraded identical preview in Experiment 2 allowed us to investigate any potential display change awareness costs associated with it. Due to that fact that valid previews represent the actual target word, there are no letters masking this word that may induce preview costs. However, it is still possible that the administration of visual degradation to valid previews may result in display change awareness costs. This question is theoretically important, given that recent studies (Marx et al., 2015, 2016) have suggested the use of visually-degraded valid previews as an alternative to using conventional parafoveal masks in developing readers.

To answer the question of what display changes are actually being detected, we manipulated the instruction to participants between subjects. More specifically, half of the participants were instructed to indicate if they saw anything change on the screen (Experiment 2a). The remaining half were instructed to indicate only if they saw any letters change (Experiment 2b). The general method is presented in Experiment 2a, and the small changes to the instructions and data analysis are presented in Experiment 2b.

Experiment 2a

Method

Participants. Sixteen undergraduate students participated for course credit. Their mean age was 19.1 years (SD= 0.68). None of them had participated in Experiment 1.

Apparatus. Same as in Experiment 1.

Materials and procedure. Eighty-eight sentences from Experiment 1 were used as stimuli (11 per condition). There were 3 within-subject factors. First, as in Experiment 1, the frequency of word N-1 was either high or low. Second, the parafoveal preview of word N was either identical or a letter mask. Third, the preview of word N was either degraded by 20% or not degraded. The presentation of sentences and conditions was counterbalanced with a full Latin square design. In Experiment 2a, participants were instructed to respond positively to display change prompts anytime they noticed anything change on the monitor when reading a sentence.

In addition to the experimental sentences, 44 filler sentences were also included in the experiment. The fillers were divided into four equal groups, and each group had one type of parafoveal preview that was used in the experimental sentences (identical, identical degraded, letter mask, degraded letter mask). The display change of these filler sentences was delayed for 15 ms so that participants would occasionally see an easy-to-notice display change. Furthermore, to make sure that participants are not primed to look for display changes in the middle of the sentence, the position of the boundary in the fillers was systematically varied: for one third of them it occurred in the beginning of the sentence, for another third in the middle, and for the remaining third in the end of the sentence.

The sentences were presented in the same way as in Experiment 1. However, after each sentence, participants were prompted to give a confidence rating on whether they think any display changes occurred (1= confident that the display did not change; 4= confident that the display changed). At the start of the experiment, there were eight practice sentences in which the display change was delayed by 20 ms (i.e. easy to notice). During the practice, participants were also provided with feedback on the correct answer (i.e. whether the display actually changed). The practice sentences contained all four preview types (two sentences per preview condition). In addition to this, participants were encouraged to ask any questions that they may have about the task. At the end of the practice, participants were verbally asked by the experimenter if they understood their task and any remaining questions were answered. After the practice session ended, no more feedback was provided. On 33% of the experimental and filler trials, participants answered a yes/no comprehension question. This happened immediately after the confidence rating screen.

Data analysis. The same random-effects structure as in Experiment 1 was used, except that we also added a random slope for word N degradation for both subjects and items. For all LMMs, the following contrasts were used. High frequency target words were coded as -0.5 and low frequency words were coded as 0.5. Identical preview conditions were coded as -0.5 and masked preview conditions were coded as 0.5. Finally, degraded previews were coded as -0.5 and non-degraded previews were coded as 0.5. This was done to be consistent with the analyses in Experiment 1. However, it should be noted that the direction of the effects should be in the opposite direction for degraded identity and degraded letter mask previews in Experiment 2. On one hand, the estimates for degraded letter masks should be positive if they introduce less processing costs (cf. Marx et al., 2015). On the other hand, the estimates for the degraded identity previews should be negative because Marx et al. reported longer fixation durations with increasing degradation for identical (i.e. valid) previews. Fixation durations were log-transformed in all analyses. Due to the smaller number of participants in Experiment 2a (16), the degrees of freedom of the t values were estimated with the "lmerTest" package in R. Therefore, in Experiment 2, p-values are also reported. P-values smaller than .05 were taken as a statistically significant result.

The calculation of d primes and confidence ratings as a measure of display change sensitivity followed the general procedure outlined in Slattery et al. (2011) and Angele et al. (2016). Hit rates and false alarm rate were calculated for each subject and condition. Trials with identical preview were used for calculating false alarm rate. The confidence rating data were analyzed with a Cumulative Link Mixed Model (CLMM; Knoblauch & Maloney, 2012) by entering participants and items as random intercepts. In addition to that, the present study also used Bayesian t-tests for analyzing the d prime data (Morey & Rouder, 2015; Mulder, &

Wagenmakers, 2016; Rouder, Speckman, Sun, Morey, & Iverson, 2009). The advantage of this method is that, unlike frequentist statistics, it can also quantify evidence in support of the null hypothesis (Dienes, 2016). The results are expressed as a Bayes factor, which is the posterior odds of the alternative and null hypothesis, given the data. Bayes factors bigger than 1 in the present paper favour the alternative hypothesis, while Bayes factors smaller than 1 favour the null hypothesis. For all analyses, the default prior (r= $\sqrt{2}/2$) of the BayesFactor package on the effect size was used. However, a sensitivity analysis with different priors showed that the conclusions from the analysis did not change (see the Supplementary files).

Results

Comprehension accuracy. The mean comprehension accuracy was 89% overall and 88% for trial sentences only. A detailed analysis of the comprehension accuracy data is presented in the Supplementary files. Correlations between question accuracy and d prime were calculated to check whether reading comprehension was related to the ability to detect display changes. The correlation between d prime for letter masks and comprehension accuracy was $r= -0.17$, $p= .53$ [95% CI: -0.61, 0.36]. The correlation between d prime for degraded letter masks and comprehension accuracy was $r= -0.18$, $p= .51$ [95% CI: -0.62, 0.35]. The generally weak correlations suggest that the ability to detect display changes was not associated with comprehension accuracy.

D prime and confidence rating. Descriptive statistics for the display change sensitivity measures are presented in Table 6. By fitting a CLMM on the confidence rating data, a comparison of the valid preview condition (i.e. identical non-degraded) to all three invalid preview conditions showed that participants gave higher ratings (indicating confidence in the occurrence of a change) to invalid preview conditions ($b=0.43$, $S.E= 0.02$, $z= 20.02$, $p< .001$). A simpler comparison revealed that participant were more confident detecting the two degraded previews compared to the traditional letter mask previews ($b= 0.79$, $S.E= 0.03$, $z= 25.79$, $p< .001$). Interestingly, there was no difference in confidence rating between degraded identical previews and degraded letter mask previews ($b= -0.03$, $S.E= 0.05$, $z= -0.63$, $p= 0.53$). However, this last comparison interacted significantly with word N-1 frequency ($b= -0.26$, $S.E= 0.10$, $z= -2.56$, $p= 0.01$). Follow-up tests of the interaction showed that participants were more confident at detecting degraded identical previews when word N-1 frequency was high ($b= -0.16$, $S.E= 0.07$, $z= -2.25$, $p= 0.02$), but not when word N-1 frequency was low ($b= 0.10$, $S.E= 0.07$, $z= 1.39$, $p= 0.16$).

The Bayes factor t tests on the d prime data corroborated the results from the CLMM. There was decisive evidence (cf. Jeffreys, 1961) that participants were better at detecting degraded identical previews (BF_{10}= 8763797) and degraded letter masks (BF_{10}= 8045876) compared to the traditional letter mask previews. The test between the degraded identical and degraded letter mask preview favored the null hypothesis of no difference (BF_{10}= 0.3). In other words, participants were equally good at detecting degraded identical and degraded letter mask previews when instructed to look for any display changes.

(Insert Table 6 here)

Fixation durations. The same data exclusion criteria as in Experiment 1 were applied. Overall, 19.6 % of data was excluded (9.8 % for blinks and track losses, 9.7 % for display changes completed after the onset of fixation, and 0.14 % as outliers). The descriptive statistics for word N-1 and word N are presented in Tables 7 and 8. The results from the LMMs are presented in Tables 9 and 10.

(Insert Table 7 here)

(Insert Table 8 here)

Similarly to Experiment 1, there was a robust frequency effect on word N-1 for both FFD and GD. Additionally, there were again no significant PoF effects on word N-1. Degradation resulted in significantly longer fixation durations for both FFD and GD on word N. The effect of degradation on word N also interacted significantly with parafoveal preview, thus confirming the intuitive expectation that the preview effect will be larger for degraded identical compared to non-degraded identical preview. This pattern is superficially similar to the one found by Marx et al. (2015) in that the difference between identical and letter mask previews is smaller in the degraded condition, but this is not because adding visual noise facilitates processing of the target word after receiving a letter mask preview that could be interpreted as a reduction in preview cost (like in Marx et al.'s study). Rather, adding visual noise to any of the preview conditions causes an increase in the subsequent fixation times on the target word.

If visual noise was added to the letter mask, the increase was a little smaller than if visual noise was added to the identical preview. The reason for this is that the degraded identical preview does not contain much useful information about the target word[3]. In other words, we did not observe a reduction in preview cost, but only a small difference in the preview benefit. Finally, frequency interacted significantly with parafoveal preview on word N for both FFD and

[3] This is consistent with Marx et al.'s (2015) conclusion that a valid preview degraded by 20% "…prevents the extraction of (useful) parafoveal information" (p. 557). However, what the authors consider to be "useful" information is not clear from their paper. Because the degradation is random, some letters may be more degraded than others. Therefore, it is still possible that readers may be able to extract some orthographical information from such previews. This is a question that needs to be addressed in future research.

GD, and with degradation for FFD. In other words, the effect of parafoveal preview measured with FFD and GD, and the effect of degradation measured with FFD was bigger for high compared to low frequency word N-1.

(Insert Table 9 here)

(Insert Table 10 here)

Discussion

The main finding from Experiment 2a is that participants were highly likely to detect a degraded preview, but they were much less likely to detect a letter mask preview without degradation. In fact, participants showed much higher detection rates and d-primes for the degraded previews compared to Angele et al. (2016; nearly 3 vs. 1.2), while the detection rates and d-primes for the non-degraded letter mask previews were four times lower than in Angele et al. (average d-prime 0.3 vs. 1.2). This suggests that the presence of the very salient degraded previews inhibited the detection of the letter mask previews. Additionally, either the degraded previews themselves or making the decision to respond to them appeared to interfere with foveal word recognition as these previews resulted in longer fixation durations on word N. Therefore, not only did visual degradation fail to reduce the processing costs associated with conventional masks, but it actually increased them.

Finally, we did not replicate the PoF effects associated with display change detection tasks that were observed by Angele et al. (2016) on the pre-target word. However, the effect of foveal load was statistically significant (we will return to this in the General Discussion). In summary, it seems that adding visual noise to a parafoveal preview changes reading behavior and also dramatically increases display change awareness. In Experiment 2b, we went on to explore if the explicit instruction to only detect letter masks could remedy this.

Experiment 2b

Method

Participants. Sixteen[4] undergraduate students participated for course credit. Their mean age was 19.3 years (SD= 0.70). None of them had participated in the previous experiments.

Materials, apparatus, and procedure. The design was the same as in Experiment 2a, except for the following differences. Unlike Experiment 2a, participants were instructed to respond positively only if they noticed letters change on the screen. The confidence rating screen was also modified to reflect the different instruction (1= confident that the letters did not change; 4= confident that the letters changed). Therefore, in this study, participants should respond to the degraded identical condition negatively, resulting in a "correct rejection". This is in contrast to Experiment 2a, where a positive response indicated a "hit". In calculating d prime, degraded identity previews served as false alarm rate for degraded letter mask trials as the former contain degradation but no letter changes. The d prime for non-degraded letter masks was calculated as in Experiment 2a.

Results

Comprehension accuracy. The mean comprehension accuracy was 86% overall and 83% for trial sentences only. Two participants had comprehension accuracies of 66% and 68% respectively. Thus, it is possible that this instruction condition represented a more difficult task to participants than the one in Experiment 2a. There were no significant differences in comprehension accuracy across the conditions: $F1(7, 105)= 1.24$, $MSE= 0.04$, $p= .29$, $\eta_G^2 = 0.07$; $F2(7, 196)= 1.73$, $MSE= 0.05$, $p= .1$, $\eta_G^2 = 0.04$.

[4] One more participant was tested but their data were lost due to equipment failure.

The correlation between d prime for letter masks and trial accuracy was $r= 0.18$, $p= .5$ [95% CI: -0.35, 0.62]. The correlation between d prime for degraded letter masks and trial accuracy was $r= -0.39$, $p= .14$ [95% CI: -0.74, 0.13]. This latter correlation also supports the conclusion that the display change detection task in this experimental condition was more difficult as participants' comprehension accuracy was negatively related to their ability to detect letter changes in degraded previews. Nevertheless, this correlation was still relatively weak and not statistically significant.

D prime and confidence ratings. A summary of participants' display change detection sensitivity is displayed in Table 11. The confidence rating data were again analyzed with a CLMM. Consistent with Experiment 2a, a comparison between the two degraded and the two non-degraded previews showed that participants were heavily biased towards responding positively to degraded previews ($b= 1.58$, $S.E= 0.07$, $z= 22.85$, $p< 0.001$). Therefore, even though degraded identity previews did not contain a letter change, participants still responded positively to those trials. On the other hand, the comparison between non-degraded identical and letter mask previews was not statistically significant ($b= 0.07$, $S.E= 0.05$, $z= 1.47$, $p= 0.14$). This shows that participants' ability to detect non-degraded letter masks was not significantly better than chance. Interestingly, however, the CLMM model showed that participants were slightly more confident when detecting degraded letter mask previews compared to the baseline of degraded identity previews ($b= 0.10$, $S.E= 0.04$, $z= 2.31$, $p= 0.02$). Therefore, there was a slight hint that participants may be able to perceive a change in letters even when the previews are degraded. However, the magnitude of the effect was quite small compared to the effect between degraded and non-degraded previews more generally.

(Insert Table 11 here)

One interesting question is whether this ability to distinguish between degraded letter mask and degraded identity previews is better than the ability to distinguish between non-degraded letter mask and non-degraded identity previews. A Bayes factor t test on the d prime data favored the null hypothesis of no difference (BF_{10}= 0.37). In other words, participants were no better at detecting the difference between identical and letter masks, than they were at detecting the difference between the degraded letter mask and degraded identical previews. An examination of the hit and false alarm rates for these conditions indicates that these participants were highly biased to say that a letter change occurred when the preview was degraded. This resulted in a high hit rate for the degraded mask condition and an almost equally high false alarm rate with degraded identical previews.

Fixation durations. During the pre-processing stage, 21.5 % of data was excluded (7.8 % for blinks and track losses, 13.4 % for display changes completed after the onset of fixation, and 0.28 % as outliers). The descriptive statistics for word N-1 and word N are presented in Tables 12 and 13. The results from the LMMs are presented in Tables 14 and 15. Similar to the previous two experiments, there was a frequency effect on word N-1 for GD. The frequency effect for FFD was in the same direction but did not reach significance. The effect of degradation from Experiment 2a was also replicated in Experiment 2b. Again, degraded previews resulted in longer fixation durations on word N compared to non-degraded ones. The effect of parafoveal preview was significant for GD, but not for FFD (although it was in the same direction). In this experiment, however, there were no significant interactions, even though the interaction between preview and degradation was in the expected direction.

(Insert Table 12 here)

(Insert Table 13 here)

(Insert Table 14 here)

(Insert Table 15 here)

Discussion

Overall, the effects observed in Experiment 2b were markedly different from 2a: participants did not seem to be able to reliably perform the letter mask detection task in the presence of degraded previews. This can be seen in Figure 3, which shows the Receiver Operating Characteristic (ROC) curves for both experiments. An ROC curve is a plot that displays participants' sensitivity to changes as a function of their false alarm rate (e.g., see Streiner & Cairney, 2007 for a review). If participants have no sensitivity to detect display changes, then their hit rate would be the same as their false alarm rate and the ROC curve would overlap with the diagonal line. However, if their sensitivity to detect display changes is above chance level, then their hit rate would be higher than their false alarm rate and the ROC curve would appear in the upper part of the graph.

(Insert Figure 3 here)

The ROC curves in Figure 3 clearly show that, in both experiments, participants did not detect letter masks, but rather degradation alone. There may be a tendency for participants to be slightly more cautious with their ratings in Experiment 2b. In this experiment, we also did not find any PoF effects on the pre-target word. Additionally, the effects of word N-1 frequency and word N preview were not statistically significant for FFD, although they were in the expected direction and were both present in GD. The reason for this could be that the display change detection task was more challenging, which was evidenced by participants' lower comprehension accuracy compared to Experiment 2a.

One of the most important findings of Experiment 2b was that, even when explicit instructions were given to ignore the visual degradation and to only respond to the letter masks, participants were unable to do so. This finding cannot be explained by assuming that participants in Experiment 2b did not understand their task because their ability to detect letter masks markedly improved in the filler trials where display changes were delayed and were thus easier to notice (compare Figures S3 and S4 in the Supplementary files). This result is indeed what one would expect if participants had understood their task- namely, they will get better at it as the task gets easier. In this sense, the present findings suggest that visual degradation is extremely salient and pervasive.

One interesting result in both Experiment 2a and 2b is that degraded letter masks resulted in longer fixation durations compared to non-degraded letter masks. This is in contrast to Experiment 1, where letter masks degraded by 20% did not result is longer gaze durations. There was a numerical trend for longer first fixations, but it was not statistically significant. It is possible that this discrepancy may be explained by the additional task of looking for display changes in Experiment 2. For example, the prolonged fixation durations may be due to participants' making a decision to respond to the display change. This would be consistent with the fact that sensitivity to degraded display changes was much higher and it may have thus affected the foveal processing of the target word to a greater extent.

General Discussion

To summarize, in three experiments we investigated whether administering visual degradation can show that traditional parafoveal masks introduce processing costs in adult readers, and whether applying visual degradation to any type of preview may introduce any costs of its own. In Experiment 1, we failed to extend Marx et al.'s (2015) results to adults reading silently. In other words, we could not find support for the prediction that administering increasing levels of visual degradation would reduce processing costs, as evidenced by increasingly shorter fixation durations. In Experiments 2a and 2b, we investigated costs related to display change awareness. The results showed that degraded valid and invalid previews are associated with very high sensitivity to display changes and that visual degradation is so salient that it suppresses the detection of traditional letter mask previews.

There are a number of differences between the present Experiment 1 and Marx et al.'s study (2015) that may potentially explain the different results. One reason could be that adults detect the degradation more easily than children do, which in turn may cause additional processing costs. A second difference is that the present study was done in English, while Marx et al.'s study was carried out with developing readers of German. Although parafoveal preview effects from the upcoming word have been established in a number of languages in adults, including English and German (Vasilev & Angele, 2016), it is possible that there may be greater cross-language differences in parafoveal processing in developing readers. This is a question that needs to be explored in future research.

A third difference is that the present study involved silent reading, while Marx et al.'s (2015) study utilized oral reading. Differences between these two modes of reading have received less attention in the literature, but there is evidence that parafoveal preview effects in skilled readers of alphabetical languages are larger in silent compared to oral reading (Ashby, Yang, Evans, & Rayner, 2012; Inhoff, & Radach, 2014; but see Pan, Laubrock, & Yan, 2016 for readers of Chinese). At present, it is not known how parafoveal processing differs in children as a function of reading mode. However, if we assume that developing readers of alphabetical languages also make better use of parafoveal processing in silent compared to oral reading, it can be argued that the effects found in Marx et al.'s study should be bigger in size in silent-reading experiments such as the present study. However, this is a speculation that remains to be tested in future research.

A final possibility is that adults may engage in parafoveal processing to a greater extent compared to children. Previous studies with the moving window paradigm (McConkie & Rayner, 1975) have shown that the perceptual span in children increases with reading instruction (Häikiö, Bertram, Hyönä, & Niemi, 2009; Rayner, 1986; Sperlich, Meixner, & Laubrock, 2016), although it already approximates the adult span by sixth grade (Häikiö et al., 2009; Rayner, 1986). Interestingly, Häikiö, Bertram, and Hyönä (2010) found that parafoveal preview effects in fourth- and sixth-graders measured with the boundary paradigm (Rayner, 1975) are numerically similar to those of skilled adult readers. In fact, Marx et al.'s (2015) participants were in the exact same grades. Additionally, there is evidence that beginning readers of German are able to extract phonological information from parafoveal vision (Tiffin-Richards & Schroeder, 2015), and that fourth-grade English readers extract information about letter position and letter identity parafoveally (Pagán, Blythe, & Liversedge, 2016). These findings suggest that Marx et al.'s child participants should have been able to engage in parafoveal processing to a degree that is similar to the one that is typically found in adults. Of course, it is still possible that there are ways in which parafoveal processing differs between fourth- and sixth-grade children, and the adult participants in our study. These in turn may help explain some of the discrepancies between the two studies. A better understanding of these differences will be possible after more is known about what type of information children process parafoveally.

One of the main findings in the present study was that degraded parafoveal previews seem to be easily detectable and to have a very strong impact on visual processing during reading. For this reason, our data cast doubt on whether such masks can be widely adopted to study parafoveal processing in adults during silent reading. Using degraded parafoveal previews may be less disruptive in children, but Marx et al. (2015, 2016) do not provide any data on display change awareness for their child participants. This in turn makes it difficult to assess the actual impact of degradation in their studies. In fairness to Marx et al. (2015), degraded previews may be further refined to make it harder to notice the degradation (for example, by putting the replaced pixels at the edge of letters, rather than scattering them freely around words). However, even then, it would still be necessary to demonstrate that parafoveal masks do not represent a "neutral" baseline for studying parafoveal preview effects in adult readers (see Marx et al., 2015, pp. 558-559). This is because in Experiment 1 we could not replicate the finding of decreased interference of letter masks that is shown by shorter fixation durations when the masks are less salient (i.e. increasingly degraded). At any rate, the present findings show that different preview manipulations can come with their own costs, as degraded previews were found to be highly detectible.

One interesting finding in the present research was the general lack of foveal load effects. Although an interaction between word N-1 frequency and word N preview was found in Experiment 2a, it was not found in the remaining experiments. One possibility is that the absence of such effects may be explained by the high awareness of display changes introduced by the degraded previews. White et al.'s (2005) results support this explanation because the authors observed a foveal load effect only in participants who were not aware of display changes. Of course, it is also possible that other factors beyond display change awareness played a role. For example, participants in Experiments 2a and 2b received different instructions, while participants in Experiment 1 did not have an additional display change detection task at all. Currently, it is not known how task demands influence the foveal load effect and this needs to be addressed by future research.

If one assumes a more sceptical view of the foveal load effect, a second possibility is that foveal load is simply an artefact of using parafoveal masks. Recently, Marx, Hawelka, Schuster, and Hutzler (2017) argued that foveal load effects observed with conventional parafoveal masks may be due to a combination of two things: 1) interference from pre-processing the parafoveal mask when foveal load is low; and 2) frequency spillover effect of the pre-target word when foveal load is high. This is an interesting prediction that needs to be tested by future research. More broadly, future studies may also wish to consider whether the foveal load effect can be reliably replicated with different types of parafoveal preview. Nevertheless, the proposal of Marx et al. (2017) is a less parsimonious explanation than the theoretical account of Henderson and Ferreira (1990) because it requires two processes (spillover effect and interference from the parafoveal mask) to explain why a foveal load effect may be observed with traditional letter masks.

If interference from the parafoveal mask contributes to the size of the foveal load effect, it can be further argued that visually degrading this mask may result in a smaller foveal load effect. Similar to Marx et al.'s (2017) proposal, this view requires that one should observe longer fixation durations after letter mask previews when foveal load is low compared to when it is high. This is because the low load condition should leave more attentional resources for pre-processing the letter mask in parafoveal vision. However, this prediction was not generally borne out in our data because fixation durations after non-degraded letter mask previews were not consistently longer when foveal load was low compared to when it was high. The two such differences that were found in GD were fairly small in size (4 and 7 ms, respectively). Therefore, while Marx et al.'s (2017) prediction remains to be tested in adult readers, our results are currently more in line with White et al. (2005), who showed that both foveal load and the traditional preview benefit effect are modulated by display change awareness.

Interestingly, Marx et al.'s (2017) recent study found no evidence of foveal load effect in fourth- and sixth-graders by using the incremental boundary technique (Marx et al., 2015). One surprising result in their study was that only 2 out of 238 children reported seeing display changes (<1%) when asked after the study. This is in contrast to the results of Experiment 1 where the detection rate of degraded previews in adults was 62%. Additionally, their child participants should have generated a parafoveal preview effect similar to or smaller than adults (see the discussion above). However, the authors reported parafoveal preview effects that are twice as big as the ones typically found in adults (cf. Vasilev & Angele, 2016, Table 2). This was due to long fixation duration measures in the degraded identical preview condition, which is a finding at odds with the contention that such previews are more neutral and involve less costs than traditional masks, but agree with the findings reported here. These two findings, together with the results of our experiments, suggest that visual degradation may interfere with reading by temporarily delaying attempts at linguistic processing of the target word.

In the framework of the parafoveal processing model proposed by Angele et al. (2016), display change detection occurs during an initial visual check stage that is pre-attentional. The current findings indicate that degraded text is highly apparent to this initial stage. Additionally, the proposed two-stage model posited that these detections were the cause of PoF effects found in some studies. However, the d-primes for degraded change detection in the current study were

5-6 times as high as those in Angele et al. (2016) while the PoF effects were absent. One major difference between these studies is that the detection of display changes in Angele et al. (2016) required detecting changes to orthographic information while those in the current study did not. In fact, the evidence from Experiment 2b indicates that participants were largely unable to detect changes to orthography in the presence of such obvious degraded text changes. Therefore, these results sharpen our two-stage model of parafoveal processing by suggesting that PoF effects are only observed in a change-detection task when the detection of changes is based on parafoveal information about upcoming words rather than on low-level parafoveal information about visual degradation.

In summary, we would recommend that future experiments using parafoveal visual degradation take the high salience and the strong impact of degraded parafoveal previews into account. Display change awareness is likely to be quite high in such studies and should at the very least be monitored, as it can lead to significant changes in eye movement behavior (Slattery et al., 2011; White et al., 2005). Finally, we were unable to confirm the claim made by Marx et al. (2015) that increasingly degrading parafoveal letter masks reduces their preview costs, as evidenced by increasingly shorter fixation durations. As such, we could not extend their original finding to adults reading silently.

References

Angele, B., & Rayner, K. (2013). Eye movements and parafoveal preview of compound words: Does morpheme order matter? *The Quarterly Journal of Experimental Psychology*, *66*(3), 505-526. doi:10.1080/17470218.2011.644572

Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic Bulletin & Review*, *23*(4), 1241–1249. doi:10.3758/s13423-015-0995-0

Ashby, J., Yang, J., Evans, K. H., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, & Psychophysics*, *74*(4), 634-640. doi:10.3758/s13414-012-0277-0

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*(4), 390-412. doi: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*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 813-840. doi:10.1037/0278-7393.19.4.813

Dienes, Z. (2016). How Bayes factors change scientific practice. J*ournal of Mathematical Psychology*, *72*, 78-89. doi:10.1016/j.jmp.2015.10.003

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

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621-636. doi: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*(4), 777-813. doi:10.1037/0033-295X.112.4.777

Gagl, B., Hawelka, S., Richlan, F., Schuster, S., & Hutzler, F. (2014). Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 588- 595. doi:10.1037/a0034408

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, USA: Cambridge University Press.

Häikiö, T., Bertram, R., & Hyönä, J. (2010). Development of parafoveal processing within and across words in reading: Evidence from the boundary paradigm. *The Quarterly Journal of Experimental Psychology*, 63(10), 1982-1998. doi:10.1080/17470211003592613

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102*(2), 167-181. doi:10.1016/j.jecp.2008.04.002

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 417-429. doi:10.1037/0278-7393.16.3.417

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: Evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*, 33. doi:10.3389/fnsys.2013.00033

Inhoff, A. W., & Radach, R. (2014). Parafoveal preview benefits during silent and oral reading: Testing the parafoveal information extraction hypothesis. *Visual Cognition*, *22*(3-4), 354-376. doi: 10.1080/13506285.2013.879630

Inhoff, A. W., Starr, M., Liu, W., & Wang, J. (1998). Eye-movement-contingent display changes are not compromised by flicker and phosphor persistence. *Psychonomic Bulletin & Review*, *5*(1), 101-106. doi:10.3758/BF03209463

Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics*, *57*(8), 1101-1110. doi:10.3758/BF03208367

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

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology*, *66*(3), 581-600. doi:10.1080/17470218.2012.658073

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R (Vol. 32)*. New York, USA: Springer Science & Business Media.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. doi:10.3389/fpsyg.2015.01171

Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2015). An incremental boundary study on parafoveal preprocessing in children reading aloud: Parafoveal masks overestimate the preview benefit. *Journal of Cognitive Psychology*, *27*(5), 549-561. doi:10.1080/20445911.2015.1008494

Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2017). Foveal processing difficulty does not affect parafoveal preprocessing in young readers. *Scientific Reports*, *7*, 41602. doi:10.1038/srep41602

Marx, C., Hutzler, F., Schuster, S., & Hawelka, S. (2016). On the development of parafoveal preprocessing: Evidence from the incremental boundary paradigm. *Frontiers in Psychology*, *7*, 514. doi:10.3389/fpsyg.2016.00514

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899 -917. doi:10.1037/h0037368

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Attention, Perception, & Psychophysics*, *17*(6), 578-586. doi:10.3758/BF03203972

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics, 25*, 221–224. doi:10.3758/BF03202990

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

Mulder, J., & Wagenmakers, E. J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". *Journal of Mathematical Psychology*, 72, 1-5. doi:10.1016/j.jmp.2016.01.002

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *36*(3), 507-518. doi:10.1080/14640748408402174

O'Malley, S., Reynolds, M. G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 451-458. doi:10.1037/0278-7393.33.2.451

Pagán, A., Blythe, H. I., & Liversedge, S. P. (2016). Parafoveal preprocessing of word initial trigrams during reading in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 411-432. doi:10.1037/xlm0000175

Pan, J., Laubrock, J., & Yan, M. (2016). Parafoveal processing in silent and oral reading: Reading mode influences the relative weighting of phonological and semantic information in Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1257-1273. doi:10.1037/xlm0000242

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65-81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65-81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211-236. doi:10.1016/0022-0965(86)90037-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506. doi:10.1080/17470210902816461

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. https://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*(1), 4–22. doi: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), 1–21. https://doi.org/10.3758/PBR.16.1.1

Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the EZ Reader model. *Psychological Science*, *17*(9), 742-746. doi:10.1111/j.1467-9280.2006.01775.x

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word n+ 1 during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 653-668. doi:10.1037/a0034997

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-237. doi:10.3758/PBR.16.2.225

RStudio (2015). RStudio: Integrated development environment for R (Version 0.99.491) [Computer software]. Boston, MA. Retrieved from http://www.rstudio.org/

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, *20*(4-5), 391-421. doi:10.1080/13506285.2012.670143

Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language, 83*, 118-139. doi:10.1016/j.jml.2015.04.005

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. J*ournal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1924-1938. doi:10.1037/a0024322

Sperlich, A., Meixner, J., & Laubrock, J. (2016). Development of the perceptual span in reading: A longitudinal study. *Journal of Experimental Child Psychology*, *146*, 181-201. doi:10.1016/j.jecp.2016.02.007

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, *52*(2), 121-128. doi:10.1177/070674370705200210

Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology*, *27*(5), 531-548. doi:10.1080/20445911.2014.999076

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190. doi:10.1080/17470218.2013.850521

Vasilev, M. R., & Angele, B. (2016). Parafoveal preview effects from word N+1 and word N+2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, 1-24. doi:10.3758/s13423-016-1147-x

White, S. J., Rayner, K., & Liversedge, S. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, *12*, 891-896. doi:10.3758/BF03196782

White, S. J., & Staub, A. (2012). The distribution of fixation durations during reading: Effects of stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 603-617. doi:10.1037/a0025338

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 274-296. doi:10.1037/0278-7393.33.2.274

Table 1

Frequency Information About Pre-Target Words (N-1)

	High-frequency words				Low-frequency words			
	Mean	SD	Min	Max	Mean	SD	Min	Max
SUBTLEX frequency	204	329	18	2304	2	2	0.03	16
SUBTLEX Zipf scale	5.08	0.38	4.26	6.36	3.14	0.42	1.54	4.19
BNC frequency	167	204	6	1421	4	4	0.18	26

Note: Frequencies reported as counts per million.

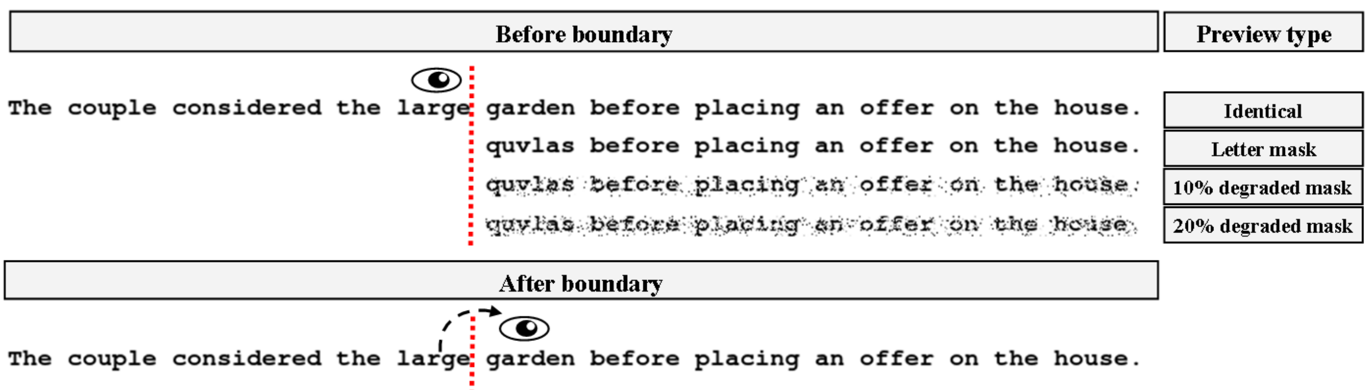

Figure 1. An illustration of the parafoveal preview manipulation of word N. The invisible boundary is represented by a dotted vertical line. In this example, word N-1 (“large”) is high frequency; in the low frequency condition, it was replaced by “weedy”.

Table 2

Fixation Durations for Word N-1 in Experiment 1 (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	10% degraded letter mask	20% degraded letter mask
FFD	High	228 (67)	233 (76)	233 (90)	235 (82)
FFD	Low	253 (81)	256 (93)	248 (92)	261 (96)
GD	High	261 (114)	273 (133)	266 (129)	263 (117)
GD	Low	319 (163)	334 (182)	306 (150)	329 (165)

Table 3

LMMs for Word N-1 in Experiment 1

Predictor	FFD				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.43	0.02	**240.03**	**< .001**	5.57	0.03	**174.35**	**< .001**
Preview (letter mask)	0.01	0.02	0.58	0.56	0.03	0.03	1.33	0.19
Preview (10% deg. mask)	-0.01	0.02	-0.57	0.57	-0.01	0.02	-0.40	0.69
Preview (20% deg. mask)	0.01	0.02	0.73	0.47	0.01	0.02	0.56	0.57
Freq	0.10	0.02	**4.24**	**< .001**	0.17	0.03	**5.35**	**< .001**
Preview (letter mask) * Freq	-0.01	0.03	-0.21	0.84	0.02	0.04	0.52	0.60
Preview (10% deg. mask) * Freq	-0.03	0.03	-0.90	0.37	-0.03	0.04	-0.68	0.50
Preview (20% deg. mask) * Freq	<-0.01	0.03	-0.13	0.89	0.02	0.04	0.48	0.63

Note: Statistically significant effects are formatted in bold. Preview: word N parafoveal preview. Freq: word N-1 frequency.

Table 4

Fixation Durations for Word N in Experiment 1 (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	10% degraded letter mask	20% degraded letter mask
FFD	High	238 (87)	268 (88)	268 (89)	280 (87)
FFD	Low	241 (84)	282 (103)	274 (98)	284 (102)
GD	High	268 (114)	303 (104)	301 (103)	314 (116)
GD	Low	270 (109)	330 (131)	310 (128)	317 (128)

Table 5

LMMs for Word N in Experiment 1

Predictor	FFD				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.42	0.02	**242.75**	**< .001**	5.51	0.03	**200.21**	**< .001**
Preview (letter mask)	0.14	0.02	**6.15**	**< .001**	0.18	0.03	**6.21**	**< .001**
Preview (10% deg. mask)	0.13	0.02	**5.74**	**< .001**	0.14	0.03	**5.37**	**< .001**
Preview (20% deg. mask)	0.16	0.02	**8.46**	**< .001**	0.17	0.02	**7.69**	**< .001**
Freq	0.01	0.02	0.55	0.59	0.02	0.03	0.64	0.52
Preview (letter mask) * Freq	0.03	0.03	0.84	0.40	0.05	0.03	1.55	0.12
Preview (10% deg. mask) * Freq	0.01	0.03	0.23	0.82	<0.01	0.03	0.03	0.98
Preview (20% deg. mask) * Freq	-0.02	0.03	-0.50	0.62	-0.03	0.03	-0.87	0.38

Note: Statistically significant effects are formatted in bold. Preview: word N parafoveal preview. Freq: word N-1 frequency.

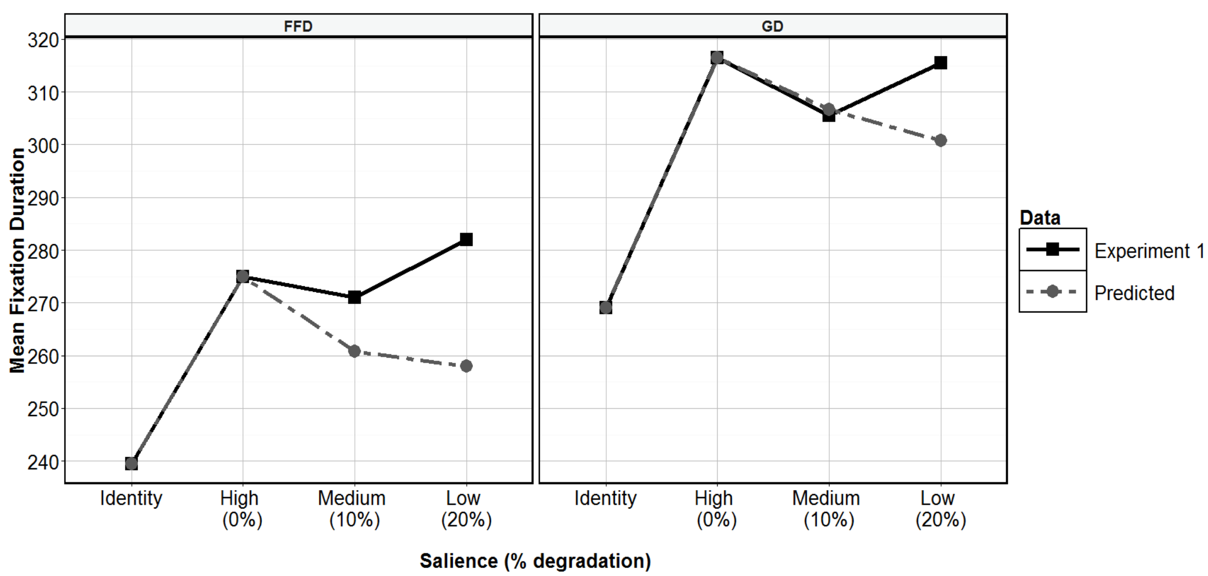

Figure 2. Fixation durations on word N as a function of salience (i.e. the amount of visual degradation). Plotted are the identity preview, letter mask preview with high salience (0% degradation), and letter mask preview with decreasing salience (i.e. degradation by 10 and 20%). Additionally, a predicted line for the degraded previews based on the effect sizes from Marx et al. (2015, Figure 3) is shown.

Table 6

Means for the Display Change Sensitivity Measures in Experiment 2a (SDs in parenthesis)

Word N preview	Word N-1 frequency	*d'*	False alarm rate	Hit rate	zFalse alarm	zHit
Letter mask	High	.55 (.57)	.05 (.05)	.14 (.09)	-1.77 (.37)	-1.22 (.48)
Letter mask	Low	.05 (.42)	.08 (.08)	.07 (.04)	-1.61 (.50)	-1.56 (.37)
20% degraded mask	High	2.91 (1.13)	.05 (.05)	.82 (.26)	-1.77 (.37)	1.14 (1.06)
20% degraded mask	Low	2.99 (.89)	.08 (.08)	.88 (.14)	-1.61 (.50)	1.38 (.64)
20% degraded identity	High	3.17 (.93)	.05 (.05)	.86 (.25)	-1.77 (.37)	1.41 (.96)
20% degraded identity	Low	2.86 (1.06)	.08 (.08)	.85 (.16)	-1.61 (.50)	1.25 (.68)

Table 7

Fixation Durations for Word N-1 in Experiment 2a (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	Degraded identical	Degraded letter mask
FFD	High	226 (75)	226 (85)	227 (73)	233 (79)
FFD	Low	246 (92)	247 (98)	246 (81)	258 (89)
GD	High	246 (97)	243 (102)	261 (115)	247 (90)
GD	Low	281 (156)	283 (136)	302 (127)	300 (120)

Table 8

Fixation Durations for Word N in Experiment 2a (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	Degraded identical	Degraded letter mask
FFD	High	216 (80)	242 (87)	256 (80)	290 (91)
FFD	Low	222 (60)	259 (97)	262 (80)	252 (81)
GD	High	243 (122)	283 (108)	274 (90)	317 (101)
GD	Low	246 (99)	276 (102)	297 (124)	278 (100)

Table 9

LMMs for Word N-1 in Experiment 2a

Predictor	FFD				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.42	0.03	**160.0**	**< .001**	5.52	0.04	**156.2**	**< .001**
Preview	0.01	0.02	0.66	0.51	-0.01	0.03	-0.33	0.74
Frequency	0.08	0.02	**3.62**	**.001**	0.14	0.03	**4.07**	**< .001**
Degradation	-0.02	0.03	-0.84	0.41	-0.06	0.03	-1.79	0.09
Preview x Frequency	0.01	0.04	0.21	0.83	0.02	0.05	0.37	0.71
Preview x Degradation	-0.03	0.04	-0.81	0.42	0.01	0.05	0.22	0.83
Frequency x Degradation	<-0.01	0.04	-0.07	0.94	-0.04	0.05	-0.77	0.44
Preview x Frequency x Degradation	-0.05	0.08	-0.65	0.51	-0.09	0.10	-0.91	0.37

Note: Statistically significant effects are formatted in bold.

Table 10

LMMs for Word N in Experiment 2a

Predictor	FFD				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.46	0.04	**155.9**	**< .001**	5.55	0.04	**149.19**	**< .001**
Preview	0.08	0.02	**3.67**	**< .001**	0.10	0.02	**3.96**	**< .001**
Frequency	<0.01	0.02	0.17	0.86	-0.01	0.02	-0.42	0.68
Degradation	-0.12	0.03	**-4.48**	**< .001**	-0.12	0.03	**-4.30**	**< .001**
Preview x Frequency	-0.08	0.04	**-2.18**	**0.03**	-0.14	0.04	**-3.35**	**< .001**
Preview x Degradation	0.08	0.04	**2.10**	**0.04**	0.09	0.04	**2.27**	**0.02**
Frequency x Degradation	0.12	0.04	**3.18**	**.001**	0.06	0.04	1.44	0.15
Preview x Frequency x Degradation	0.18	0.07	**2.48**	**0.01**	0.15	0.08	1.78	0.08

Note: Statistically significant effects are formatted in bold.

Table 11

Means for the Display Change Sensitivity Measures in Experiment 2b (SDs in parenthesis)

Word N preview	Word N-1 frequency	*d'*	False alarm rate	Hit rate	zFalse alarm	zHit
Letter mask	High	.36 (.56)	.07 (.07)	.14 (.10)	-1.60 (.46)	-1.23 (.50)
Letter mask	Low	.36 (.63)	.08 (.08)	.16 (.16)	-1.63 (.53)	-1.26 (.75)
20% degraded mask	High	.12 (.63)	.61 (.22)	.65 (.24)	.36 (.74)	.48 (.78)
20% degraded mask	Low	.37 (.72)	.60 (.25)	.70 (.25)	.35 (.83)	.72 (.90)

Note: For the 20% degraded mask, the false alarm rate is calculated with the identical degraded preview condition.

Table 12

Fixation Durations for Word N-1 in Experiment 2b (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	Degraded identical	Degraded letter mask
FFD	High	249 (73)	236 (80)	254 (106)	226 (77)
FFD	Low	250 (82)	263 (105)	249 (85)	259 (105)
GD	High	253 (74)	253 (90)	274 (135)	264 (160)
GD	Low	272 (113)	303 (145)	297 (159)	281 (139)

Table 13

Fixation Durations for Word N in Experiment 2b (SDs in Parentheses)

Measure	Word N-1 Frequency	Parafoveal preview of word N			
		Identical	Letter mask	Degraded identical	Degraded letter mask
FFD	High	234 (79)	257 (87)	289 (82)	294 (101)
FFD	Low	242 (100)	263 (106)	299 (95)	304 (108)
GD	High	255 (103)	294 (116)	328 (139)	321 (109)
GD	Low	255 (103)	290 (137)	318 (136)	345 (129)

Table 14

LMMs for Word N-1 in Experiment 2b

Predictor	FFD				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.45	0.03	**160.5**	**< .001**	5.51	0.05	**120.79**	**< .001**
Preview	-0.03	0.03	-1.19	0.24	-0.02	0.03	-0.62	0.54
Frequency	0.04	0.03	1.52	0.15	0.08	0.03	**2.77**	**0.01**
Degradation	0.03	0.03	1.01	0.33	0.03	0.04	0.67	0.51
Preview x Frequency	0.08	0.05	1.82	0.07	0.03	0.05	0.56	0.57
Preview x Degradation	0.04	0.05	0.79	0.43	0.10	0.05	1.94	0.053
Frequency x Degradation	<-0.01	0.05	-0.09	0.93	0.02	0.05	0.34	0.73
Preview x Frequency x Degradation	0.02	0.09	0.21	0.84	0.15	0.10	1.48	0.14

Note: Statistically significant effects are formatted in bold.

Table 15

LMMs for Word N in Experiment 2b

Predictor	FFD[1]				GD			
	b	S.E.	t value	p value	b	S.E.	t value	p value
Intercept	5.55	0.03	**216.33**	**< .001**	5.63	0.03	**187.86**	**< .001**
Preview	0.04	0.03	1.71	0.11	0.08	0.03	**2.90**	**0.01**
Frequency	0.02	0.02	1.04	0.32	0.01	0.03	0.31	0.76
Degradation	-0.19	0.04	**-4.94**	**< .001**	-0.21	0.04	**-4.73**	**< .001**
Preview x Frequency	<-0.01	0.04	-0.10	0.92	0.04	0.04	0.88	0.38
Preview x Degradation	0.07	0.04	1.78	0.08	0.08	0.04	1.92	0.06
Frequency x Degradation	0.01	0.04	0.04	0.97	<0.01	0.04	0.05	0.96
Preview x Frequency x Degradation	-0.02	0.08	-0.26	0.80	-0.13	0.09	-1.46	0.15

Note: Statistically significant effects are formatted in bold.

[1] Word N degradation was removed as a random slope for items due to convergence failure

Figure 3. Receiver operating characteristic (ROC) curves for Experiments 2a and 2b. The confidence levels are shown by the numbers.