This is a preprint of an article whose final and definitive form was published in Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 39(2), Mar 2013, 649-662. doi: 10.1037/a0029294

Processing *the* in the parafovea: Are articles skipped automatically?

Bernhard Angele and Keith Rayner

Department of Psychology

University of California, San Diego, USA

Correspondence to:

Bernhard Angele

Department of Psychology

University of California San Diego

9500 Gilman Drive #0109

La Jolla, CA 92093-0109

Email: bangele@ucsd.edu

Running Head: Skipping of articles

Abstract

One of the words that readers of English skip most often is the definite article *the*. Most accounts of reading assume that in order for a reader to skip a word, it must have received some lexical processing. The definite article is skipped so regularly, however, that the oculomotor system might have learned to skip the letter string *t-h-e* automatically. We tested whether skipping of articles in English is sensitive to context information or whether it is truly automatic in the sense that any occurrence of the letter string *the* will trigger a skip. This was done using the gaze-contingent boundary paradigm (Rayner, 1975) to provide readers with false parafoveal previews of the article *the*. All experimental sentences contained a short target verb, the preview of which could be correct (that is, identical to the actual subsequent word in the sentence, e.g. *ace*), a nonword (*tda*), or an infelicitous article preview (*the*). Our results indicated that readers tended to skip the infelicitous *the* previews frequently, suggesting that, in many cases, they seem to be unable to detect the syntactic anomaly in the preview and base their skipping decision solely on the orthographic properties of the article. However, there was some evidence that readers sometimes detected the anomaly, as they also showed increased skipping of the pre-target word in the *the* preview condition.

Readers do not fixate every single word in a sentence, but occasionally skip a word. Words are especially likely to be skipped if they are short, highly frequent, and predictable from the preceding sentence context (Koriat & Greenberg, 1994; for a summary of research on skipping effects, see Rayner, 1998, 2009). Some word skipping can also be explained by mislocated fixations (i.e. readers attempting to fixate a short word, but overshooting it and landing on the subsequent word instead; Nuthmann, Engbert, & Kliegl, 2005). It is also clear that word skipping is influenced by parafoveal processing (Fitzsimmons & Drieghe, 2011; Rayner, Slattery, Drieghe, & Liversedge, 2011; Schotter, Angele, & Rayner, 2012). The phenomenon of word skipping has the potential of providing insight into the time course of word identification and syntactic integration. More specifically, since the decision to skip a word has to be made relatively early during a fixation (within about 75 - 125 ms of fixation onset[1]), readers must make their decision quickly and may not take information from higher levels of processing into account. The definite article *the* is an ideal candidate for studying word skipping, since it is short, highly frequent, and highly predictable.

Previous research by O'Regan (1979) and Carpenter and Just (1983) demonstrated that three-letter function (closed-class) words such as articles or prepositions are more likely to be skipped by readers than three-letter content (open-class) words. Gautier, O'Regan, and Le Gargasson (2000) replicated O'Regan's finding that articles are skipped much more often than other short words even when they are not predictable from the context, as did Drieghe, Pollatsek, Staub, and Rayner (2008). This suggests that readers use parafoveal information in order to decide whether to skip a word, and do not rely exclusively on prior context. However, the nature of the parafoveal information that leads to the increased skipping rates in the *the*-condition is not clear: readers may skip *the* because it is easy to process, or they may have learned to skip articles by default, regardless of their processing status.

Another interesting issue is whether readers skip *the* automatically without actually seeing the word. Specifically, some recent research (Roy-Charland, Saint-Aubin, Lalande, Bélanger, & Klein, 2012) in which the missing-letter effect (MLE) was combined with a gaze-contingent moving window paradigm (McConkie & Rayner, 1975) suggests that this might be the case. In MLE studies (Healy, 1994; Koriat & Greenberg, 1993, 1994; Saint-Aubin & Klein, 2008), subjects have to detect a target letter in text. The well-known finding is that they miss more letters in frequent function words than in less frequent content words. In Roy-Charland et al.'s study, a moving window was used such that the fixated word was available for processing, but all words to the right of fixation were masked with X's. They found that readers were able to detect a target letter embedded in a word that was skipped. In such cases, the letter could only have been identified in post-view (to the left of fixation). More critically for the present issues, they found that *the* was skipped slightly more often than a three-letter content word (an 8% difference).

In Dutch, Drieghe, Brysbaert, Desmet, and De Baecke (2004) investigated the influence of context-driven expectations on the skipping of short words (but not *the*) in general. In particular, they compared target words that were highly predictable from the preceding context (e.g. "maakte het bed *op*"; English: "made the bed"; target word in italics) with neutral target words of either the same length ("maakte het bed *na*"; English: "imitated the bed") or a different length ("maakte het bed *vast*", English: "fastened the bed"). Drieghe et al. found an effect of contextual constraint on the probability of fixating the target word, with predictable words being skipped more often than unpredictable words. They claimed that the size of this effect, a nine percentage point difference, is among the largest that can be obtained with contextual constraints. Drieghe et al. also reported a main effect of word length, with short words being skipped more often than long words. There was, however, no effect of expected word length on skipping, i.e. the skipping probability of an unexpected word did not depend on whether it had the same length as the expected word. As a consequence, Drieghe et al. argued that visual features like word length and linguistic features like predictability might influence skipping separately.

In the present study, we tested the hypothesis that function words are automatically skipped by using a gaze-contingent preview manipulation (Rayner, 1975). This enabled us to differentiate between effects of parafoveal information and effects of the sentence context without overtly using syntactically illegal or unusual sentences. Specifically, we provided readers with a preview of the definite article *the* in a position where it can be expected to always be grammatically illegal (that is, in the position of a three-letter word used as a verb, such as *ace*). If readers only consider the upcoming parafoveal letters when making a skipping decision, we expected to find higher skipping rates for target verbs which had a infelicitous *the* preview than for target verbs which had a correct preview (e.g. *ace*). On the other hand, if readers do consider context information, they should be able to detect the anomaly and, as a consequence, be more likely to fixate the problematic word.

Whether readers are able to detect the anomaly inherent in the *the* previews should also determine their fixation durations on the target word and the surrounding words: If readers detect the anomaly, they should show longer fixation times on the pre-target word (i.e. a parafoveal-on-foveal effect), the target word, and, possibly, a spill-over effect on the post-target word in the *the* preview condition compared to the correct preview condition. If readers do not detect the anomaly until after they have skipped the target word, we should expect no difference in fixation times on the pre-target word. In contrast, there should be strong effects on the post-target words after inappropriate skips of the target word as well as a higher probability for regressions out of that word. Finally, on those trials where the target word is not skipped, we might also expect an effect of the incorrect previews on the target word itself.

In order to establish a baseline for the effects of unusual parafoveal information on fixation times and word skipping, we also included a condition in which the preview consisted of random letter strings (e.g. *fda*). We expected this condition to result in immediate effects of the unusual letter strings on fixation times on the target word, as well as possible effects on the pre-target and post-target words. However, random letter previews should not cause readers to skip the target word more frequently than the correct previews – on the contrary, we expected that readers would be more likely to fixate a word with a random letter preview.

Finally, in order to make sure that any observed effects are due to letter identity and not lower-level influences such as word shape, we presented all sentences in upper case for half of the subjects. We did not expect this to have a strong influence on parafoveal processing (see Slattery, Schotter, Berry, & Rayner, 2011).

Method

Subjects.

Sixty University of California, San Diego students participated in this experiment for course credit. All were native speakers of English, had either normal or corrected to normal vision, and were naïve concerning the purpose of the experiment.

Apparatus.

An SR Research Eyelink 1000 eyetracker was used to record subjects' eye movements with a sampling rate of 1000 Hz. Subjects read sentences displayed on an Iiyama Vision Master Pro 454 video monitor with a refresh rate of 150 Hz. Viewing was binocular, but only the right eye was recorded. Viewing distance was approximately 60 cm, with 3.8 letters equaling one degree of visual angle.

Materials.

Sixty-three experimental sentences were generated, each one containing a three-letter target word which was always used as a verb (e.g. *She was sure she would* ***ace*** *all the tests*, target word in bold; see the Appendix for all sentences)[2]. Acceptability ratings for each of the sentences were obtained from 46 University of California, San Diego undergraduates who participated for course credit and were native speakers of English to ensure that no sentence was unacceptable to the target population. On a scale from 1 (unacceptable) to 7 (perfectly acceptable), the average rating of the experimental sentences was 5.18. In order to ensure that any effects were not primarily due to word shape, half of the subjects read sentences in lower-case, while the other half read the sentences in all caps (see Figure 1).

Insert Figure 1 about here

Procedure.

The 63 experimental sentences were embedded in 60 filler sentences unrelated to the present study. Subjects were asked to read the sentences on the computer screen silently and press a button on the Eyelink button box when they were finished and felt that they understood the sentence content. During the presentation of the experimental sentences, the gaze-contingent boundary paradigm (Rayner, 1975) was used to manipulate the parafoveal preview of the target word. There were three preview conditions: The preview was correct, that is, identical to the target word (*ace*), a random-letter preview (*fda*), or a false, infelicitous preview of an article (*the*; see Figure 1). It is important to note that this procedure ensured that the false article preview always appeared in a position in which an article would be syntactically illegal.

After 20 out of the 63 sentences (31.7 %), subjects were presented with a two-alternative forced choice comprehension question and used the trigger buttons on the Eyelink button box to select the answer they thought was correct (see Appendix for a list of all questions used).

Results

For each of the critical words, we examined fixation time on the target word. Trials with track losses or display changes that completed after fixation onset as well as trials in which a blink occurred immediately before or during a fixation on the target word were eliminated (10.33% of the data). If a fixation was shorter than 80 ms and located within one character space (11 pixels) of another fixation, it was merged into that fixation. Otherwise, it was deleted, as were fixations shorter than 80 ms or longer than 800 ms (less than 1% of the data). All subjects answered at least 85% of the comprehension questions correctly.

Since we expected that frequent skipping of the three-letter target words and exclusion of delayed display changes would lead to unequal cell sizes, inferential statistics are reported based on linear mixed models (LMM) with subjects and items as crossed random effects (Baayen, Davidson, & Bates, 2008). In order to fit the LMMs, the lmer function from the lme4 package (Bates, Maechler, & Dai, 2009) was used within the R Environment for Statistical Computing (R Development Core Team, 2011). For each factor, we report regression coefficients (b), standard errors, and t-values. For binomial dependent variables such as fixation and regression probabilities, we report regression coefficients, standard errors, and z-values from generalized LMMs using a logit-link. We do not report p-values, since it is not clear how to determine the degrees of freedom for LMMs, making it difficult to estimate p-values. However, since our analyses contain a large number of subjects and items and only a few fixed and random effects are estimated, we can assume that the distribution of the t-values estimated by the LMMs approximates the normal distribution. We will therefore use the two-tailed criterion $|t| \geq 1.96$ which corresponds to a significance test at the 5% α-level. Of course, the z-values from the generalized LMMs can be interpreted in exactly the same way.

We fitted an LMM for each of the following dependent variables on each target word: first fixation duration (FFD), gaze duration (GD), go-past time (go-past), landing position, fixation probability, and the probability of making a regression out of the word. FFD is the mean duration of the first fixation on a word, regardless of whether there are subsequent fixations on that word or not. It can be considered a measure of early processing (Rayner, 1998, 2009). Mean GD is the sum of the duration of the first fixation on a word and the durations of all subsequent refixations before leaving the word. It is still a measure of early processing, but can capture some later processing difficulties that force a reader to refixate on a word. Mean go-past time includes all the fixations used to calculate GD, but additionally considers the durations of fixations that are made to the left of the word in question from the time a reader first enters that word from the left until the reader leaves the word to the right. As such, it is sensitive to integration difficulties that require regressions. Tables 1a, 2a, and 3a show the means and standard deviations of each dependent variable in each of the experimental conditions.

For fixation probability and probability of regressions out, logistic LMMs were used (Gelman & Hill, 2007). The analyses included two fixed effects, preview (correct vs. random letter vs. infelicitous *the*) and case (normal sentence vs. upper case) as well as their interaction (a 3 X 2 design), and random intercepts for subjects and items. We used two orthogonal contrasts to further explore differences between the preview factor levels. Contrast 1 compared the random letter condition with the mean of the correct and the infelicitous *the* condition – that is, it compared the condition in which the preview was a nonword with the conditions in which the preview was a word (random letters = 1, correct = -.5, infelicitous *the* = -.5). Contrast 2 then compared the correct and the infelicitous *the* condition (random letters = 0, correct = -1, infelicitous *the* = 1). Since we expected the infelicitous *the* condition to cause the strongest disruption on the post-target word (as opposed to the pre-target and target words where the random letter preview was expected to cause more disruption), we used slightly different contrasts for the analyses of the post-target word. In the post-target word analyses, Contrast 1 corresponds to the difference between the correct and the mean of the *the* and the random letters preview conditions (random letters = .5, correct = -1, infelicitous *the* = .5) and Contrast 2 tests for a difference between the random letters and the infelicitous *the* conditions (random letters = -

1, correct = 0, infelicitous *the* = 1). The LMM analyses included random intercepts for subjects and items as well as random preview effect slopes for subject and random case effect slopes for items. Model comparisons showed that none of the other possible random slopes (i.e. random preview effect slopes for items or random slopes for the interaction terms between preview and case) were justified by the data. Tables 1b, 2b, and 3b show the results for all models fitted on the pre-target, target, and post-target. We will now discuss the effects on each word in detail.

Pre-target word

Table 1a shows the means and standard deviations of all the dependent measures on the pre-target word. Random letter nonword previews resulted in longer GDs and Go-past times compared to the correct and the infelicitous *the* preview conditions (all coefficient estimates and t-values are presented in Table 1b). Compared to the other two conditions, in the random letter nonword condition there were also more regressions out of the pre-target word and there was a marginal effect indicating that the pre-target word was more likely to be fixated. These orthographic parafoveal-on-foveal effects are caused by the presence of a nonword letter string in the parafovea and are not necessarily related to lexical processing (for a review see Schotter et al., 2012). Somewhat surprisingly, there also was an effect of the infelicitous *the* preview (compared to the correct preview) on pre-target fixation probability, with the pre-target word being less likely to be fixated when the subsequent target word was *the.* This effect could be interpreted as subjects detecting the anomaly caused by the infelicitous *the* early and making a saccade towards it and could, as such, be classified as a lexical or syntactic parafoveal-on-foveal effect (Schotter et al., 2012). A similar effect of “magnetic attraction” to unusual information was observed by Hyönä and Bertram (2004; see also Hyönä, 1995). However, it is important to point out that the letter string “the” hardly constitutes unusual parafoveal information on its own.

Also, the size of this effect is quite small (3.4 percentage points in the lower case and 2.5 percentage points in the upper case condition). Finally, capitalization had an effect on landing positions on the pre-target word, with fixations occurring slightly further to the left of the word when the sentences were displayed in upper case. No other effects were significant.

Target word

Table 2a shows the means and standard deviations of all the dependent measures on the target word (the corresponding model coefficients, standard errors, and t-values are shown in Table 2b). Since the present experiment was designed to elicit skipping of the target word, we need to consider fixation probability on the target word and fixation times on the target word separately. Fixation probability on the target word is an indicator of whether the nature of the preview affected saccade target selection during the previous fixation. Our results show that this is the case: While an irregular letter preview caused readers to fixate the target word more often than in the other conditions, the infelicitous *the* preview condition led to a strong increase in skipping compared to the correct control condition – around 50% of the infelicitous *the* previews caused subjects to skip the target word when it was displayed in lower case. In the upper case condition, skipping rates were somewhat lower numerically, but the increase in skipping was still substantial. The interaction between capitalization and preview was not significant.

It is important to keep in mind that the remaining dependent measures – FFD, GD, Go-past time, landing position, and the probability of making a regression out of the target word – are a subset of the data and reflect the consequences of *NOT* skipping the target word. All measures show reliable effects of the irregular letter preview, with longer fixation times and a higher probability of regressions out of the target word in the irregular preview condition. This can be considered a preview benefit effect (Rayner, 1975; for a review, see Rayner, 1998). The infelicitous *the* condition resulted in a somewhat smaller preview benefit in GD and go-past time. There was a numerical trend for these effects to be stronger in the upper case condition, but the corresponding interaction term was not significant. As for landing position, when the nonword previews were not skipped, subjects' fixations landed further to the left in the target word than in the correct control condition. A significant interaction with capitalization indicates that this effect was enhanced when sentences were displayed in upper case.

In summary, the fixation probabilities on the target word show that the infelicitous *the* preview manipulation had the intended effect: subjects skipped the infelicitous *the* previews more often than random letter strings, and, importantly, more often than the actual subsequent word which fit into the sentence context. In those cases where subjects decided not to skip the target word, there was no evidence of processing disruption caused by irregular or infelicitous *the* previews beyond what is to be expected due to standard parafoveal processing effects. With respect to fixation time measures, first fixations on the target word following the infelicitous *the* previews were not longer than those following correct previews. GDs and go-past times following the infelicitous *the* previews were longer, however. In general, these effects were on the same order of magnitude as the effects of having had an irregular letter preview. This means that, in those cases where subjects decided not to skip the target word after an infelicitous *the* preview, its effects were comparable to a random letter preview and are essentially preview benefit effects.

Post-target word

Table 3a shows the means and standard deviations of all the dependent measures on the post-target word (the corresponding model coefficients, standard errors, and t-values are shown in Table 3b). There were no effects of any of the experimental manipulations on FFD, GD, and landing position on the post-target word. In go-past time, however, an effect of the target word preview manipulation emerged. First, go-past times on the post-target word were longer following a random letter or infelicitous *the* preview of the target word compared to the correct condition (a spillover effect). Additionally, when there had been an infelicitous *the* preview of the target word, go-past times on the post-target word were much longer than when the preview had consisted of random letters. This suggests that, while the infelicitous *the* preview did not have an effect on the early processing of the post-target word, it caused a major disruption of later processing stages, most likely on the syntactic integration of the post-target word (and the target word, if it was skipped) into the sentence structure. We observed a very similar effect on the probability of making regressions out of the post-target word, suggesting that the increase in go-past time is due to subjects re-reading earlier words in the sentence in order to arrive at a sensible interpretation of it.

Finally, we also observed significant effects of the preview manipulation on the probability of fixating the post-target word. Specifically, subjects were less likely to fixate the post-target word in the correct preview condition compared to the random letter and the infelicitous *the* condition and subjects were more likely to fixate the post-target word in the infelicitous *the* condition than in the random letter preview condition. The post-target word was also fixated more often when the sentences were displayed in upper case compared to the lower case condition. A similar effect was observed on landing positions, with fixations occurring further towards the left in the post-target words in the nonword and the infelicitous *the* conditions than in the correct control condition and fixations occurring further to the left in the infelicitous *the* condition than in the nonword condition This is likely a consequence of skipping the target word (see post-hoc analysis below).

Post-hoc analyses

Felicitous occurrences of the

The results presented above come with a potential limitation: as the preceding context determines whether an occurrence of *the* is felicitous or infelicitous, it is impossible to compare infelicitous and felicitous instances of *the* without changing the preceding context. However, most of our sentences contained at least one felicitous instance of *the*. Using a generalized LMM, we performed a post-hoc analysis of fixation probability on these felicitous occurrences of *the*, comparing the probability of fixating felicitous instances of *the* to the probability of fixating the target word in the control, nonword, and infelicitous *the* conditions. Sentence-initial occurrences of *the* were excluded from this analysis. Table 4 shows the fixation probability on the felicitous instances of *the* compared to the fixation probabilities in the experimental conditions. It is important to note that, in this analysis, we are no longer able to keep the preceding context constant between all conditions, as the felicitous instances of *the* can, by design, not occur in the target position. However, this post-hoc comparison between fixation rates for felicitous and infelicitous occurrences of *the* is vastly superior to an informal comparison between fixation rates for infelicitous instances of *the* in the present experiment and skipping rates observed for felicitous instances of *the* in previous studies. Importantly, our analysis found no significant difference between the probability of fixating infelicitous instances of *the* in the target position and felicitous instances of *the* elsewhere in the experimental sentences (b =,127 SE = ,117, z = 1.08). Also, there was no significant interaction of this contrast with case. This means that, at least as far as skipping decisions are concerned, subjects did not treat the infelicitous instances of *the* differently from felicitous instances of *the* that naturally occurred in the experimental sentences. Finally, we tested the possibility that processing of felicitous and infelicitous instances of *the* is influenced by their position within a sentence. We performed an analysis that included word position, preview, and the interaction between those two factors. The interaction between word position and the contrast between felicitous and non-felicitous instances of *the* was significant (b = -.193, SE = .071, z = -2.714). However, it is not clear how to interpret this interaction, which suggests that, at the beginning of a sentence, fixation rates for felicitous *the* were lower than those following infelicitous *the* previews (felicitous *the*: M = .461, SD = .149; infelicitous *the*: M = .593, SD = .0201), while there were lower fixation rates for infelicitous *the* previews compared to felicitous instances of *the* at the end of a sentence (felicitous *the*: M = .593, SD = .0193; infelicitous *the*: M = .495, SD = 0198). This pattern of data does not easily lend itself to interpretation and will have to be addressed by future research.

Costs and benefits of skipping the target word on the surrounding words

Since the preview manipulations had a strong effect on the probability of skipping the target word, the observed effects of the preview condition on eye-movement behavior on the pre- and post-target words may be confounded with the effects of skipping. For example, Kliegl and Engbert (2005) found that fixations prior to skipping of short or high-frequency words were shorter than fixations prior to normal forward saccades (skipping benefit), while fixations prior to skipping long or low-frequency words were longer than fixations prior to normal forward saccades (skipping cost). Kliegl (2007) reported that skipping benefits seem to be associated with skipping function words and skipping costs seem to be associated with skipping content words. We investigated this by performing a set of post-hoc LMMs on pre- and post target fixation time measures, fixation and regression probabilities, and landing positions. These LMMs included the same fixed and random main effects as the analyses (but not the interaction terms between preview and capitalization) reported above, as well as a factor indicating whether the target word was skipped in that trial (coded as -1 for fixated and 1 for skipped) and its interactions with preview and capitalization. A detailed account of the results can be found in Appendix B. In the following, we just highlight the most important result.

We found an effect of preview on go-past time, which remained highly significant even when target skipping was included as a predictor (Contrast 1: b = 34.12, SE = 5.82, t = 5.86; Contrast 2: b = 23.91, SE = 5.4, t = 4.472). This result is critical, as it shows that the effects of the infelicitous *the* preview on post-target word fixation duration associated with *the*-preview are not simply a consequence of the higher skipping probability associated with this condition.

Discussion

In the present study, we used the gaze-contingent boundary paradigm to present readers with previews of three-letter words which suggested that the upcoming word was the definite article *the*. This manipulation had the potential of affecting ongoing processing during fixations on the pre-target word in two ways: First, it made the parafoveal word appear to be extremely easy to process, which should increase the probability of readers skipping it. Second, since the sentences were constructed so that the target position (a verb) could never be occupied by an article, it caused the parafoveal word to be syntactically illegal given the preceding sentence context. This syntactic anomaly should make readers less likely to skip the target word. Our experiment therefore pitted parafoveal information and information about the preceding sentence context against each other.

Our results are quite straightforward: when context information and parafoveal information are in conflict, whether a reader will skip a word is strongly influenced by the parafoveal information, and not by the context – readers skipped the target much more often in the infelicitous *the* condition than in the correct preview condition in which the preview was compatible with the preceding context. While there was less skipping in the upper case condition, there was no interaction between the capitalization and preview conditions, suggesting that readers relied on letter identity (that is, the letter sequence "*T-H-E*"), not word shape, when making skipping decisions. There is, as noted above, the possibility that this effect is modulated by the position of the word in question within the sentence.

As expected, the infelicitous *the* preview had a disruptive effect – but this effect occurred quite late, most likely after the identification of the post-target word was completed, leading to increased go-past times on the post-target word and an increased probability of making a regression out of the post-target word. These effects were not just consequences of the increased target skipping rate, but appear to be genuine indicators of syntactic integration difficulty. Assuming that the article *the* was fully identified before the skipping decision, this difficulty could be the result of readers futilely attempting to fit the article *the* into an incompatible sentence context. If the infelicitous *the* preview was not fully identified but just to the extent that it can trigger a skipping eye movement, the disruption could be the result of skipping the target word inappropriately and, as a consequence, lead the reader to have to re-read earlier parts of the sentence.

From the above results, one might conclude that the sentence context has no influence at all on the decision to skip a word. A post-hoc analysis including felicitous instances of *the* confirms this, as there was no difference in fixation probability between an infelicitous target word *the* and felicitous occurrences of *the* in other parts of the experimental sentences. The low skipping rate for *the* in the present study is in contrast to other studies (Angele & Rayner, 2011; Drieghe et al., 2008), which report that felicitous articles were skipped around 80% of the time. Why subjects skipped felicitous instances of *the* less in the present experiment than previous experiments is not entirely clear. One possible explanation is that, in the sentence materials used in the present experiment, the article *the* was simply less predictable than in some previous studies. Another possibility is that, in the present experiment, subjects employed a more conservative strategy in their skipping decisions for *the.* Since every *the*-skipping in the infelicitous *the* condition resulted in considerable disruption on the subsequent word, subjects might simply decide to skip *the* less often in general. If subjects are capable of changing their response to the presence of a *the* preview, this suggests that *the*-skipping is not due to an automatic, rigid response to the letter string *t-h-e* but rather a relatively flexible reading strategy that can be modified if it leads to processing difficulties. There is evidence that older adults are able to reduce their skipping rates in expectation of difficult comprehension questions (Wotschack & Kliegl, 2012). Future research should be able to address the question of whether *the* skipping is a reading strategy and whether it is affected by processing difficulty.

A related question is whether this effect is specific to *the* or whether other function words might also be inappropriately skipped. Greenberg and Saint-Aubin (2004) found that subjects missed more occurrences of the letter *r* in the conjunction *or* than in control words even when they were presented in texts whose word order had been randomly scrambled. One possible explanation for this effect is that subjects tend to skip function words like *or* even when they appear in a syntactically inappropriate position. On the other hand, evidence for the hypothesis that *the* has a special processing status comes from a study by Koriat and Greenberg (1993) on the missing-letter effect (MLE). Koriat and Greenberg had subjects search for the Hebrew letter ה (*the*). They found that detection accuracy for ה was not influenced by syntactic structure, that is, subjects were just as good at detecting *the* within a sequence such as "and for the" or "in the" as they were at detecting *the* when it appeared on its own. This was in contrast to all other function words (e.g. *on* or *from*) whose letters were missed more often when they occurred in the first position (as in "on the") than when they were in the second position (as in "and on"). On the basis of our results, we believe that, at least in English, the definite article *the* is likely to be processed differently from any other function word due to its extremely high frequency and low semantic content, although it is possible that the indefinite article *a* or *an* might be processed in a similar way. In other languages such as French or German, which possess several different definite and indefinite articles (e.g. *le*, *la*, and *les; un, une,* and *des* in French), this may well be different. However, additional research is needed to fully settle this question.

Our finding that readers do not seem to take syntactic information into account when planning an eye movement suggests that either such information may not be available while readers make their skipping decision or they choose not to use it. This could be due to the time-course of linguistic processing: perhaps syntactic information just does not become available quickly enough to inform a skipping decision. However, predictability, which requires at least some expectations about the syntactic category of the upcoming word, has long been known to have an important effect on skipping probability (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996). It is possible that readers process the parafoveal word until a likely lexical candidate (given the parafoveally available letters) emerges. Due to the extremely high frequency of *the*, it should almost always be a likely lexical candidate. On the other hand, a less frequent content word might still emerge as a likely candidate if it is primed by the context. Further studies should investigate whether syntactic information can influence the decision to skip words other than *the*. If our hypothesis is confirmed, it might inform models of eye-movement behavior during reading such as E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006) and SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005). Finally, we found an effect of the preview manipulation on skipping the pre-target word. It remains to be seen whether this effect can be replicated. If this is the case, it will certainly inform models of eye-movements during reading as well.

In summary, readers greatly rely on parafoveal information in order to make a skipping decision. Future research is needed to clarify the role of prior context information in word skipping, which might be less important than previously assumed. This effect may be especially strong for extremely frequent function words such as *the*, which seem to be skipped regardless of the context.

References

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology. Human Perception and Performance*, *14*, 428–443.

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. doi:10.1037/a0023096

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. doi:10.1016/j.jml.2007.12.005

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

Bates, D., Maechler, M., & Dai, B. (2009). lme4: Linear mixed-effect models using S4 classes (R package Ver. 0.999375-28). *Vienna, Austria: R Foundation for Statistical Computing*.

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.

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology*, *16*, 79–103. doi:10.1080/09541440340000141

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*, 1552–1560.

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

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*, 621–636.

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

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulletin & Review*, *18*, 736–741. doi:10.3758/s13423-011-0105-x

Garside, R. (1987). The CLAWS word-tagging system. In R. Garside, F. Leech, & G. Sampson (Eds.), *The Computational Analysis of English*. London, UK: Longman.

Gautier, V., O'Regan, J. K., & Le Gargasson, J. F. (2000). "The-skipping"revisited in French: programming saccades to skip the article "les". *Vision Research*, *40*, 2517–2531.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Greenberg, S. N., & Saint-Aubin, J. (2004). Letter Detection for Homographs with Different Meanings in Different Language Texts. *Bilingualism: Language and Cognition*, *7*, 241–253. doi:10.1017/S1366728904001622

Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review*, *1*, 333–344. doi:10.3758/BF03213975

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 68–81.

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.

Kilgarriff, A. (2006). *BNC database and word frequency lists*. Retrieved from http://www.kilgarriff.co.uk/bnc-readme.html

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.

Koriat, A., & Greenberg, S. N. (1993). Prominence of leading functors in function morpheme sequences as evidenced by letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 34–50. doi:10.1037/0278-7393.19.1.34

Koriat, A., & Greenberg, S. N. (1994). The extraction of phrase structure during reading: Evidence from letter detection errors. *Psychonomic Bulletin & Review*, *1*, 345–356. doi:10.3758/BF03213976

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

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*, 2201–2217.

O'Regan, J. K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, *25*, 501–509.

R Development Core Team, R. (2011). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*.

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

Rayner, K. (1978). Eye movement latencies for parafoveally presented words. *Bulletin of the Psychonomic Society*, *11*, 13–16.

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

Rayner, K. (2009). 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.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720–732.

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology*, *40*, 473–483.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 514–528.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504–509.

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

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

Roy-Charland, A., Saint-Aubin, J., Klein, R. M., MacLean, G. H., Lalande, A., & Bélanger, A. (2012). Eye movements when reading: The importance of the word to the left of fixation. *Visual Cognition*, *20*, 328–355. doi:10.1080/13506285.2012.667457

Saint-Aubin, J., & Klein, R. M. (2008). The influence of reading skills on the missing-letter effect among elementary school students. *Reading Research Quarterly*, *43*, 132–146. doi:10.1598/RRQ.43.2.2

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

Slattery, T. J., Schotter, E. R., Berry, R. W., & Rayner, K. (2011). Parafoveal and foveal processing of abbreviations during eye fixations in reading: Making a case for case. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1022–1031. doi:10.1037/a0023215

Wotschack, C., & Kliegl, R. (2012). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2011.625094

Author Note

This research was supported by Grant HD26756 from NIH. We thank Ray Berry for his help in collecting the data and Jean Saint-Aubin, Alice Healy, and Reinhold Kliegl for their helpful comments and suggestions. Portions of the results were presented at the 51st annual meeting of the Psychonomic Society in St. Louis, MO and at the 10th Biannual Meeting of the German Society for Cognitive Science (KogWis 2010) in Potsdam, Germany.

Footnotes

1. Most sources estimate the minimum time to program a saccade at 150-175 ms (e.g. Abrams & Jonides, 1988; Rayner, 1978; E-Z Reader assumes a saccade programming time of around 125 ms, Reichle, Pollatsek, & Rayner, 2006). Given that the average fixation duration in reading is around 250 ms (Rayner, 1998), that leaves only the first 75-125 ms to make the decision to program a skipping saccade. There might be more time for saccade planning if one assumes that the target of a saccade is not decided until the very last stage of programming.

2 . Short target verbs such as hit, fit, tag, etc. almost always have homographs that are nouns or, in some cases, adjectives. We used a word frequency list with part-of-speech tags (Kilgarriff, 2006) based on the British National Corpus to determine the relative frequency of verb and noun usage for a given target word and found that, on average, our target words were more frequently used as nouns (55.3 %) than as verbs (37.63 %), with adjective and other usages making up the rest of the observations in the BNC. It is important to keep in mind that the part-of-speech tags in the BNC corpus were generated using the CLAWS tagger (Garside, 1987) and contain some ambiguity.

Table 1a: Condition means on the pre-target word.

		Fixation time measures (in ms)			Landing position	Probability	
	Preview	**FFD**	**GD**	**Go-past**	**Characters**	**Regressions out**	**Fixation**
Sentence case	Correct	210 (2.97)	227 (4.03)	251 (5.84)	2.16 (0.071)	0.0514 (0.00886)	0.754 (0.0173)
	Dissimilar	220 (3.11)	241 (4.36)	281 (7.47)	2.04 (0.0702)	0.0844 (0.0112)	0.773 (0.0169)
	Infelicitous "the"	214 (3.24)	232 (4.38)	268 (8.1)	2.14 (0.0719)	0.0494 (0.00865)	0.721 (0.0179)
Upper case	Correct	214 (2.85)	234 (4.1)	260 (6.26)	1.87 (0.072)	0.0526 (0.00921)	0.778 (0.0172)
	Dissimilar	214 (3.2)	240 (4.84)	275 (7.13)	1.98 (0.0737)	0.078 (0.0112)	0.789 (0.017)
	Infelicitous "the"	215 (3.54)	233 (4.49)	263 (7.33)	1.89 (0.0711)	0.0593 (0.00959)	0.753 (0.0175)

Standard deviations are in parentheses.

Table 1b: Fixed effects from the LMM analyses on the pre-target word. Each column represents a model fitted to one of the dependent variables.

		FFD			**Gaze duration**			**Go-past time**			**Landing Pos.**			**p(Regression out)**			**p(Fixation)**		
		b	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**z**	**b**	**SE**	**z**
	(Intercept)	214.22	3.45	**62.04**	233.26	4.46	**52.30**	264.83	7.39	**35.83**	1.96	0.07	**27.44**	-3.15	0.14	**-22.25**	1.49	0.16	**9.56**
Preview	Contrast 1: Nonword vs. Identical/'The'	3.20	1.69	1.89	6.15	2.44	**2.52**	11.95	4.17	**2.87**	0.01	0.04	0.30	0.32	0.10	**3.31**	0.13	0.07	1.89
	Contrast 2: Identical vs. 'The'	0.92	1.61	0.57	0.63	2.16	0.29	3.54	3.64	0.97	0.00	0.03	-0.08	0.07	0.10	0.77	-0.12	0.06	**-2.11**
Capitalization	lower vs. upper case	-0.77	2.99	-0.26	-2.09	3.87	-0.54	-1.06	6.09	-0.17	0.11	0.04	**2.81**	-0.17	0.13	-1.29	-0.11	0.10	-1.14
Interactions	Preview Contrast 1 x Capitalization	2.95	1.69	1.75	1.46	2.44	0.60	2.71	4.17	0.65	-0.07	0.04	-1.68	0.08	0.10	0.83	0.03	0.07	0.46
	Preview Contrast 2 x Capitalization	0.35	1.61	0.22	1.05	2.16	0.48	1.63	3.64	0.45	-0.02	0.03	-0.59	-0.04	0.10	-0.43	0.00	0.06	0.02
Random effects	Subject: (Intercept)	478.89	21.88	NA	688.88	26.25	NA	1788.11	42.29	NA	0.05	0.22	NA	0.52	0.72	NA	0.43	0.66	NA
	Subject: Preview Contrast 1	1.24	1.11	NA	27.66	5.26	NA	197.14	14.04	NA	0.01	0.07	NA	0.00	0.01	NA	0.04	0.20	NA
	Subject: Preview Contrast 2	28.77	5.36	NA	32.87	5.73	NA	158.36	12.58	NA	0.00	0.05	NA	0.01	0.08	NA	0.05	0.22	NA
	Item: (Intercept)	198.67	14.10	NA	410.59	20.26	NA	1253.94	35.41	NA	0.22	0.47	NA	0.35	0.59	NA	0.98	0.99	NA
	Item: lower vs. upper case	15.22	3.90	NA	101.74	10.09	NA	163.84	12.80	NA	0.00	0.03	NA	0.14	0.38	NA	0.05	0.22	NA
	Residual	3897.87	62.43	NA	7537.37	86.82	NA	19568.28	139.89	NA	2.09	1.45	NA	NA	NA	NA	NA	NA	NA

b: Regression coefficient, SE: standard error, t: test statistic (b/SE). Cell s with |t| >= 1.96 are marked in bold.

Table 2a: Condition means on the first target word

		Fixation time measures (in ms)			Landing position	Probability	
	Preview	**FFD**	**GD**	**Go-past**	**Characters**	**Regressions out**	**Fixation**
Sentence case	Correct	231 (3.62)	251 (4.11)	281 (6.06)	1.61 (0.0513)	0.0788 (0.0108)	0.711 (0.0182)
	Dissimilar	253 (4.26)	276 (5.11)	351 (9.01)	1.41 (0.0471)	0.164 (0.0149)	0.753 (0.0174)
	Infelicitous "the"	242 (4.94)	274 (6.37)	321 (9.31)	1.61 (0.0662)	0.0717 (0.0103)	0.486 (0.02)
Upper case	Correct	233 (3.84)	256 (4.76)	294 (7.38)	1.48 (0.0517)	0.0866 (0.0116)	0.75 (0.0178)
	Dissimilar	251 (4.19)	285 (5.2)	362 (9.22)	1.51 (0.0489)	0.158 (0.0152)	0.759 (0.0178)
	Infelicitous "the"	240 (5.02)	274 (6.15)	328 (10.3)	1.55 (0.0589)	0.0774 (0.0109)	0.598 (0.0199)

Standard errors are in parentheses.

Table 2b: Fixed effects from the LMM analyses on the target word. Each column represents a model fitted to one of the dependent variables.

		FFD			Gaze duration			Go-past time			Landing Pos.			p(Regression out)			p(Fixation)		
		b	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**t**	**b**	**SE**	**z**	**b**	**SE**	**z**
	(Intercept)	239.73	4.15	**57.78**	266.70	5.15	**51.76**	321.62	8.44	**38.11**	1.53	0.03	**50.75**	-2.51	0.13	**-19.21**	0.92	0.11	**8.43**
Preview	Contrast 1: Nonword vs. Identical/'The'	9.81	2.62	**3.75**	10.69	3.57	**3.00**	32.73	4.95	**6.61**	-0.07	0.03	**-2.00**	0.62	0.08	**7.76**	0.50	0.07	**7.04**
	Contrast 2: Identical vs. 'The'	3.41	2.21	1.54	9.55	3.06	**3.13**	18.78	4.35	**4.32**	0.02	0.03	0.53	-0.01	0.08	-0.18	-0.47	0.05	**-8.90**
Capitalization	lower vs. upper case	0.06	3.89	0.02	-3.01	4.77	-0.63	-4.92	7.75	-0.63	0.02	0.03	0.61	0.03	0.11	0.28	-0.14	0.10	-1.43
Interactions	Preview Contrast 1 x Capitalization	0.09	2.62	0.04	-2.38	3.57	-0.67	0.84	4.96	0.17	-0.07	0.03	**-2.18**	0.05	0.08	0.64	0.11	0.07	1.58
	Preview Contrast 2 x Capitalization	1.56	2.21	0.71	1.26	3.06	0.41	0.87	4.35	0.20	-0.02	0.03	-0.50	0.00	0.08	0.03	-0.07	0.05	-1.37
Random effects	Subject: (Intercept)	803.70	28.35	NA	1203.20	34.69	NA	2965.58	54.46	NA	0.02	0.13	NA	0.51	0.71	NA	0.53	0.73	NA
	Subject: Preview Contrast 1	107.04	10.35	NA	328.31	18.12	NA	234.13	15.30	NA	0.01	0.12	NA	0.03	0.17	NA	0.10	0.32	NA
	Subject: Preview Contrast 2	31.52	5.61	NA	184.18	13.57	NA	80.39	8.97	NA	0.02	0.13	NA	0.01	0.11	NA	0.05	0.22	NA
	Item: (Intercept)	137.57	11.73	NA	259.07	16.10	NA	928.42	30.47	NA	0.01	0.11	NA	0.33	0.58	NA	0.14	0.37	NA
	Item: lower vs. upper case	7.59	2.75	NA	20.86	4.57	NA	226.00	15.03	NA	0.00	0.04	NA	0.03	0.19	NA	0.03	0.16	NA
	Residual	6468.93	80.43	NA	9440.86	97.16	NA	25962.05	161.13	NA	1.11	1.05	NA	NA	NA	NA	NA	NA	NA

b: Regression coefficient, SE: standard error, t: test statistic (b/SE). Cell s with |t| >= 1.96 are marked in bold.

Table 3a: Condition means on the second target word

		Fixation time measures (in ms)			Landing position	Probability	
	Preview	**FFD**	**GD**	**Go-past**	**Characters**	**Regressions out**	**Fixation**
Sentence case	Correct	235 (4.43)	256 (5.54)	311 (9.56)	1.99 (0.0785)	0.103 (0.0122)	0.632 (0.0194)
	Dissimilar	226 (4.36)	243 (5.27)	339 (10.9)	1.91 (0.0799)	0.167 (0.015)	0.646 (0.0193)
	Infelicitous "the"	232 (4.05)	254 (5.24)	418 (12.3)	1.58 (0.062)	0.296 (0.0182)	0.734 (0.0176)
Upper case	Correct	224 (3.91)	253 (5.47)	307 (8.94)	1.78 (0.0691)	0.114 (0.0131)	0.747 (0.0179)
	Dissimilar	223 (3.86)	251 (5.35)	327 (9.8)	1.66 (0.0676)	0.133 (0.0142)	0.737 (0.0184)
	Infelicitous "the"	224 (3.66)	257 (5.18)	402 (11.8)	1.58 (0.0623)	0.265 (0.0179)	0.779 (0.0168)

Standard errors are in parentheses.

Table 3b: Fixed effects from the LMM analyses on the second target word. Each column represents a model fitted to one of the dependent variables.

		FFD			Gaze duration			Go-past time			Landing Pos.			p(Regression out)			p(Fixation)		
		b	SE	t	b	SE	t	b	SE	t	b	SE	t	b	SE	z	b	SE	z
	(Intercept)	226.25	3.98	**56.80**	249.17	5.06	**49.21**	347.03	10.24	**33.90**	1.66	0.07	**22.70**	-1.81	0.10	**-18.08**	1.18	0.16	**7.57**
Preview	Contrast 1: Correct vs. Nonword/'The'	-2.49	2.21	-1.12	-2.21	3.07	-0.72	42.62	6.22	**6.85**	-0.13	0.04	**-3.09**	0.56	0.08	**6.85**	0.14	0.06	**2.48**
	Contrast 2: Nonword vs. 'The'	1.79	2.13	0.84	4.43	2.56	1.73	39.64	6.10	**6.50**	-0.09	0.04	**-2.12**	0.47	0.07	**6.26**	0.17	0.06	**2.94**
Capitalization	lower vs. upper case	2.61	3.45	0.76	-2.59	4.20	-0.62	4.10	7.44	0.55	0.06	0.03	1.85	0.05	0.08	0.65	-0.28	0.08	**-3.40**
Interactions	Preview Contrast 1 x Capitalization	-1.84	2.22	-0.83	-2.77	3.07	-0.90	2.32	6.22	0.37	-0.03	0.04	-0.72	0.12	0.08	1.50	0.08	0.06	1.43
	Preview Contrast 2 x Capitalization	1.84	2.13	0.86	1.95	2.55	0.76	2.17	6.10	0.36	-0.05	0.04	-1.10	-0.02	0.07	-0.23	0.06	0.06	1.07
Random effects	Subject: (Intercept)	597.14	24.44	NA	800.16	28.29	NA	2214.72	47.06	NA	0.02	0.15	NA	0.27	0.52	NA	0.30	0.55	NA
	Subject: Preview Contrast 1	0.37	0.61	NA	45.50	6.75	NA	268.11	16.37	NA	0.02	0.15	NA	0.08	0.29	NA	0.00	0.02	NA
	Subject: Preview Contrast 2	61.99	7.87	NA	12.65	3.56	NA	765.63	27.67	NA	0.06	0.25	NA	0.17	0.41	NA	0.06	0.24	NA
	Item: (Intercept)	266.00	16.31	NA	561.38	23.69	NA	3318.88	57.61	NA	0.27	0.52	NA	0.21	0.46	NA	1.11	1.05	NA
	Item: lower vs. upper case	23.56	4.85	NA	66.58	8.16	NA	253.44	15.92	NA	0.01	0.08	NA	0.03	0.17	NA	0.04	0.19	NA
	Residual	6124.18	78.26	NA	10876.80	104.29	NA	42852.57	207.01	NA	1.76	1.33	NA	NA	NA	NA	NA	NA	NA

b: Regression coefficient, SE: standard error, t: test statistic (b/SE). Cell s with |t| >= 1.96 are marked in bold. Note that the contrasts for the preview factor are different from those used for the pre-target and the target word in order to make the results easier to interpret.

Table 4 - Post hoc analysis: Fixation probabilities for felicitous occurrences of *the* (fixation probabilities for the three target word preview conditions copied from Table 2a for comparison)

Capitalization condition	**Preview**	**p(Fixation)**
Normal	Correct	0.711 (0.0182)
Normal	Nonword	0.753 (0.0174)
Normal	Felicitous *the*	0.466 (0.0155)
Normal	Infelicitous *the*	0.486 (0.02)
All upper case	Correct	0.75 (0.0178)
All upper case	Nonword	0.759 (0.0178)
All upper case	Felicitous *the*	0.552 (0.0161)
All upper case	Infelicitous *the*	0.598 (0.0199)

Standard errors in parentheses.

Figure 1

Example stimuli and display change procedure

Capitalization condition	**Preview condition**	**Example**
Normal	Correct	She was sure she would ace all the tests.
Normal	Nonword	She was sure she would fda all the tests.
Normal	Infelicitous *the*	She was sure she would the all the tests.
All upper case	Correct	SHE WAS SURE SHE WOULD ACE ALL THE TESTS.
All upper case	Nonword	SHE WAS SURE SHE WOULD FDA ALL THE TESTS.
All upper case	Infelicitous *the*	SHE WAS SURE SHE WOULD THE ALL THE TESTS.

After readers fixated to the right of the invisible boundary (dashed line), the display changed to the correct preview condition corresponding to the capitalization.

Appendix A: Sentences used in the experiment. Target words are in italics.

1. The council voted to immediately *ban* cell phones in public buildings.

2. Everyone told him that he should *bow* to the emperor.

3. The members of the club will *box* every Friday to keep in shape.

4. After finishing the meal you must *bus* your table.

5. At the end of the book, all the villains will *die* when their hideout burns down.

6. The exhausted slaves must *fan* their emperor all day long.

7. If the pants don't *fit* we can hem them easily.

8. My brothers often *hum* a melody while they are working.

9. If you park here, they will *tow* your car immediately.

10. The honor student was sure she would *ace* all the tests.

11. The navy could not *man* every ship since they lacked sailors.

12. Since the grass is getting very tall, we must *mow* it next week.

13. Unfortunately, we still *owe* them a lot of money.

14. Before her illness my aunt *ran* five marathons a year.

15. They soon would *rue* their unfortunate decision.

16. On weekdays, the fast trains *run* every half hour.

17. If the button falls off we can *sew* it on again.

18. She retired after having finally *won* every prize in her field.

19. His enemies tried to *tag* him as a socialist.

20. This spring we must *lop* off all the dead branches from our trees.

21. Because the lights in their bar are bright, the owners always *dim* them at night.

22. Since it was very hot, he did not *don* his hat when he left.

23. You should not *fix* it unless it is broken.

24. The workers will *hoe* all day to prepare the field.

25. One should never *lie* about important matters.

26. Some people do not *tan* easily in the sun.

27. The crowd wanted to *tar* and feather the criminals.

28. Many people think the federal and state governments already *tax* them too much.

29. My nephew is just five years old, but he can already *tie* his shoelaces on his own.

30. When he was in town, the actor wanted to *eat* breakfast at his favorite restaurant.

31. If you are bored with hiking, you can *ski* around the area as well.

32. In order to hit the target you must *aim* very carefully.

33. Due to the growing unrest many people think they should *arm* themselves nowadays.

34. If you aren't careful, crooks will *con* you out of your savings.

35. Even experts can *err* on some issues.

36. Many fear the new high-rise buildings would *mar* our city's downtown.

37. In order to make the cake just *mix* all the ingredients on the list.

38. The impatient cows will *moo* when they want to be milked.

39. If they don't open the door, we will *ram* it with a sledgehammer.

40. At the start line, the drivers always *rev* their engines.

41. If the weather permits it, we can *row* our boat out onto the lake.

42. The witness described what she *saw* at the crime scene.

43. If you trespass on their property they will *sic* their dogs on you.

44. When catholic people *sin* they have to confess.

45. Every spring farmers *sow* their crops in the fields.

46. He said he would *sue* his employer after he was laid off.

47. In order to get the total we must *sum* up all the numbers.

48. If I had money, I would *use* it to buy a house.

49. The two parents will *vie* with each other in their attempts to gain the children's love.

50. In some religions, priests must *vow* to stay chaste.

51. This weekend we will finally *wax* our new car.

52. If the athlete wants a scholarship, she must *win* her next competition.

53. If you don't succeed, you should *try* using another method.

54. My father loves to *fry* all his food.

55. Before you can put it away, you must *dry* your wet laundry.

56. If you are happy with the service, you should *tip* the waiter generously.

57. All dogs will *wag* their tails when they are happy.

58. At nightfall we will *peg* our tents in a sheltered location.

59. The wrestler must *pin* his opponent down in order to win.

60. They did not know housing prices would *top* out very soon.

61. The paper bag burst with a loud *pop* that startled everyone.*

62. If you need something, you should *tap* gently on the door.

63. This month our sales will *hit* a new high due to the advertising campaign.

* Sentence was included due to experimenter error despite the target word being a noun instead of a verb. Excluding this sentence from the analyses did not result in a different pattern of effects.

Questions

Answer alternatives in parentheses. The first alternative was shown on the left of the screen, the second alternative was shown on the right of the screen. Correct answers are marked by asterisks.

1. Did they allow cell phones? (yes, no*)
2. Who was he supposed to bow to? (emperor*, king)
3. Do the club members exercise on Friday? (no, yes*)
4. Will the waiter clean your table? (no*, yes)
5. Did the villains survive the fire? (yes, no*)
6. Did the emperor have someone fanning him throughout the day? (yes*, no)
7. Is it easy to fix the pants if they don't fit? (no, yes*)
8. Do my brothers always work in silence? (no*, yes)
9. Will they tow your car if you park here? (no, yes*)
10. Was the student nervous about the tests? (no*, yes)
11. Was the navy understaffed? (yes*, no)
12. Is the grass short? (yes, no*)
13. Do they owe us a lot of money? - (no*, yes)
14. Did my aunt exercise a lot before her sickness? (no, yes*)
15. Were they going to regret their decision? (yes*, no)
16. Does the train only run twice on weekdays? (yes, no*)
17. Can we sew the button back on? (yes*, no)
18. Is she going to delay retirement further and keep working? (yes, no*)
19. Does he have enemies? (no, yes*)
20. Will we need to prune the trees this spring? (yes, no*)

Appendix B: Supplemental analyses on the costs and benefits of skipping the target word

Tables B1 and B2 show mean fixation times, landing positions, and regression as well as fixation probabilities for the pre-target (B1) and the post-target words (B2) conditional on skipping the target word. As there was no evidence of an interaction between target skipping and capitalization, the means in Table B1 are collapsed over the case condition. For the sake of brevity, we will only report those instances in which the pattern of effects observed in these post-hoc analyses diverged from that found in the analyses reported above.

With regard to fixation probability, skipping the target word was associated with a much higher probability of fixating the pre-target (b = 1.21, SE = .07, z = 17.26) and the post-target word (b = 1.29, SE = .07, z = 18.69). Since readers rarely skip two words at once, this is not unexpected. On the pre-target word, we also observed a skipping cost on FFDs (b = 4.09, SE = 1.38, t = -2.97), which is opposite to what Kliegl and Engbert (2005) found for single-fixation durations before short/high frequency words (see also Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner, Balota, & Pollatsek, 1986). This effect did not reach significance in any of the other fixation time measures. Additionally, the inclusion of target skipping as a predictor in the model caused the contrast measuring the difference between the nonword preview condition and the mean of the correct and the infelicitous *the* condition to reach significance with regard to pre-target fixation probability (b= -.33, SE = .06, z = -5.22), which can be interpreted as another orthographic parafoveal-on-foveal effect. There was a significant relationship between target skipping and landing position on the pre-target word, with fixations that preceded target skips located further towards the beginning of the pre-target word (b = .41, SE = .03, t = 13.88). However, this could simply be due to variations in the length of the pre-target word, as subjects tend to fixate a position several characters into a longer pre-target word and subjects are also more likely to initiate a target skip from a short pre-target word than from a long pre-target word. Interestingly, after including target skipping in the model, the main effect of preview on landing position became significant. Specifically, the nonword preview condition was associated with landing positions further into the pre-target word compared to identical and infelicitous *the.* Additionally, the infelicitous *the* condition was associated with landing positions further towards the beginning of the target word.

On the post-target word, there was a significant main effect of skipping the target word on GD, go-past time, landing position, regression probability, and fixation probability, with GDs being slightly lower when the target word had been skipped (b = 6.78, SE = 2.32, t = 2.92). A significant interaction between capitalization and target word showed that presentation of the sentence in upper case was only related to longer GDs when the target word had been skipped (b = -5.27, SE = 2.26, t = -2.33). In contrast, while there was no main effect of either capitalization or target skipping on FFD, the significant interaction indicated that lower case presentation was associated with longer FFDs, but only when the target had not been skipped. There was no such interaction for go-past times, which were substantially higher in that case (b = 74.75, SE = 4.41, t = 16.97). This effect was in addition to the effect of preview on go-past time, which remained highly significant even when target skipping was included as a predictor (Contrast 1: b = 34.12, SE = 5.82, t = 5.86; Contrast 2: b = 23.91, SE = 5.4, t = 4.42). The same was true for the increase in regression probability due to target skipping (b = 1.18, SE = .06, t = 20.75). A significant interaction term between capitalization and target skipping showed that capitalization only had an on regression probability when the target word had been skipped, with skips in the normal condition being more likely to be followed by regressions than skips in the upper case condition (b = .14, SE = .06, z = 2.45).

The effect of preview on the probability of fixating the target word disappeared once target skipping was taken into account. Surprisingly, skipping the target word was associated with a lower fixation probability on the post-target word (b = 1.29, SE = .07, z = 18.69). It is not quite clear what caused this, although it is possible that, given that both the target and the post-target word were frequently only three letters long, saccades which were intended to skip the target word only overshot and skipped the post-target word as well. Finally, target skipping also had an effect on landing positions on the post-target word, with subjects fixating further towards the beginning of the post-target word when the target word had been skipped (b = .64, SE = .03, t = -23.85).

In summary, target skipping was associated with a variety of changes in eye-movement behavior both on the pre- and the post-target word. Despite this, effects due to target skipping cannot explain the principal effects of the preview manipulation, especially the processing disruption effect on the post-target word.

Table B1: Condition means on the pre-target word conditional on skipping the target word (collapsed over capitalization conditions)

		Fixation time measures (in ms)			**Landing position**	**Probability**	
Target skipped	**Preview**	**FFD**	**GD**	**Go-past**	**Characters**	**Fixation**	**Regressions out**
No	Correct	202 (3.1)	228 (5.19)	255 (7.92)	2.58 (0.0922)	0.939 (0.0133)	0.055 (0.0126)
	Dissimilar	209 (4.19)	242 (7.18)	266 (9.43)	2.72 (0.101)	0.907 (0.017)	0.055 (0.0134)
	Infelicitous "the"	208 (2.85)	227 (3.91)	251 (6.49)	2.31 (0.0694)	0.892 (0.013)	0.0494 (0.00911)
Yes	Correct	217 (2.64)	232 (3.45)	256 (5.06)	1.74 (0.0574)	0.701 (0.0154)	0.0509 (0.0074)
	Dissimilar	221 (2.63)	240 (3.53)	283 (6.16)	1.73 (0.0551)	0.739 (0.0146)	0.0898 (0.00952)
	Infelicitous "the"	223 (4.02)	239 (5.09)	283 (9.15)	1.65 (0.0699)	0.605 (0.0189)	0.0584 (0.00908)

Table B2: Condition means on the post-target word conditional on skipping the target word (collapsed over capitalization conditions)

		Fixation time measures (in ms)			Landing position	Probability	
Target skipped	**Preview**	**FFD**	**GD**	**Go-past**	**Characters**	**Fixation**	**Regressions out**
No	Correct	226 (3.99)	258 (6.11)	351 (12.1)	0.965 (0.0685)	0.881 (0.018)	0.229 (0.0233)
	Dissimilar	220 (4.98)	243 (6.85)	429 (15.4)	0.919 (0.0727)	0.931 (0.0149)	0.44 (0.0291)
	Infelicitous "the"	226 (3.28)	255 (4.76)	481 (12.2)	1.2 (0.0568)	0.882 (0.0136)	0.485 (0.021)
Yes	Correct	231 (3.97)	253 (5)	286 (7.49)	2.36 (0.0616)	0.617 (0.0164)	0.0633 (0.0082)
	Dissimilar	227 (3.56)	249 (4.49)	286 (7.1)	2.2 (0.0618)	0.612 (0.0162)	0.0576 (0.00777)
	Infelicitous "the"	230 (4.5)	256 (5.72)	330 (10.6)	2.02 (0.0621)	0.65 (0.0185)	0.108 (0.012)