Reading Sentences of Uniform Word Length: Evidence for the Adaptation of the Preferred Saccade Length during Reading

Michael G. Cutter, Denis Drieghe, and Simon P. Liversedge

University of Southampton

Author Note

Michael G. Cutter, Denis Drieghe, and Simon P. Liversedge, School of Psychology, University of Southampton, Southampton, England.

Correspondence regarding this article should be addressed to Michael G. Cutter, School of Psychology, Shackleton Building, University of Southampton, Highfield Campus, Southampton SO17 1BJ. E-mail: M.G.Cutter@soton.ac.uk.

Word count: 12,828 (excluding tables, figures, and Appendix)

Abstract

In the current study we investigated the effect of removing word length variability within sentences on spatial aspects of eye movements during reading. Participants read sentences that were uniform in terms of word length, with each sentence consisting entirely of three, four, or five letter words, or a combination of these word lengths. Several interesting findings emerged. Adaptation of the preferred saccade length occurred for sentences with different uniform word length; participants would be more accurate at making short saccades while reading uniform sentences of three letter words, while they would be more accurate at making long saccades while reading uniform sentences of five letter words. Furthermore, word skipping was affected such that three and four letter words were more likely, and five letter words less likely, to be directly fixated in uniform compared to non-uniform sentences. It is argued that saccadic targeting during reading is highly adaptable and flexible towards the characteristics of the text currently being read, as opposed to the idea implemented in most current models of eye movement control during reading that readers develop a preference for making saccades of a certain length across a lifetime of experience with a given language.

Statement of Public Significance

The findings of the present study suggest that readers are able to rapidly adapt the way in which they move their eyes between words on the basis of the average word length within a sentence. Participants were more accurate at making short movements of the eyes when reading sentences which were entirely made up of three letter words, and more accurate at making relatively long eye movements when reading sentences which were entirely made up of five letter words. These findings demonstrate that the human perceptual system is highly sensitive to global word length information during reading, and that readers are able to

carefully tune their performance of motor behaviour on the basis of this global visual information.

Keywords: eye movements, reading, saccadic range error, preferred saccade length, saccadic targeting, word skipping.

During reading saccadic eye movements are made in order to fixate words in high-acuity foveal vision (see Rayner, 1998; 2009 for reviews). Typically these saccades will be targeted towards and originate from words of variable length. The length of the words within a sentence influences eye movement control during reading in several ways. First, the eyes will tend to land further into a long word than into a short word, necessitating saccades of differing lengths to move further when targeting long relative to short words (McConkie, Kerr, Reddix, & Zola, 1988). Secondly, the probability of skipping a word is dramatically affected by word length, with increased skipping of short compared to long words (see Brysbaert, Drieghe, & Vitu, 2005). In the current paper we investigated the effect of removing within sentence word length variability on all of these components of eye movement control. It is important to note that while word length also affects reading times, the focus of the current paper will be upon spatial, rather than temporal, aspects of eye movement control.

Saccadic targeting and the systematic range error

It is generally agreed that progressive, inter-word saccades are targeted towards the center of an upcoming word in spaced alphabetic languages. This location is considered the *optimal viewing position* for word recognition, with results from isolated word recognition studies showing that response latencies increase the further a fixation is from a word's center (O'Regan & Jacobs, 1992; O'Regan, Lévy-Schoen, Pynte & Brugaillère, 1984). Furthermore, in natural reading the probability of refixating a word increases with the distance of the first fixation from the central character (McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Rayner, Sereno, & Raney, 1996). These findings suggest a preference to fixate the center of a word. However, initial fixation landing positions tend to be normally distributed across the whole word (Rayner, 1979), indicating that the programming and execution of eye movements are not error-free, resulting in the eyes not always landing on the intended landing position. The peak of this normal distribution is slightly to the left of the word center, and is referred to as the *preferred viewing location* (Rayner, 1979). Thus, while the eyes are targeted towards a word's center in English reading, they tend to fall a small amount short of this, landing slightly to the left of the center.

McConkie, Kerr, Reddix, and Zola (1988) investigated factors influencing the distribution of initial fixation landing positions within a word. They showed that both the average initial fixation location and the standard deviation of fixations around this position was partly determined by the distance of the prior fixation from the center of the target word. Saccades launched from a greater distance resulted in a wider distribution of landing positions. McConkie et al. proposed that this occurred as a result of random motor error, with longer saccades being prone to increased motor error. More importantly for the current investigation, McConkie et al. also proposed a systematic saccadic range error during reading. They observed that English readers tended to overshoot the center of a word for saccades launched from less than seven characters away from the center, and undershoot for saccades launched from more than seven characters away. Thus, in English reading there is systematicity in the saccadic targeting system such that saccades tend to be seven characters in length, which we will refer to as the *preferred saccade length*. This preferred saccade length can be defined as the intended saccade length that is not biased to either under- or overshoot its intended target. McConkie et al. found that for each additional character that the saccade launch site was from seven characters away from the target word center, the mean landing position of the eye would shift approximately half a character before or beyond the center of the word (see Nuthmann, Engbert, & Kliegl, 2005, Paterson, Almabruk, McGowan, White & Jordan, 2015, and Yan et al., 2014, for similar findings in German, Arabic, and Uighur respectively).

McConkie et al. derived the idea of a systematic saccadic range error with a preferred saccade length from Kapoula (1985; see also Kapoula and Robinson, 1986). Kapoula (1985) had participants make saccades towards a target that appeared randomly in one of several set locations between 2.7 and 9.5 degrees of visual angle from a central launch site. When the target was at the lower end of this range participants' saccades typically overshot the target, and in the upper end of the range they often undershot the target. In a follow-up investigation, half of the original participants completed a similar task, but with targets ranging from 7 to 21.9 degrees away. Once again, participants showed a clear range bias. Crucially, a target that was far and was thus undershot in the first task (i.e. targets 7-11 degrees away) became a relatively near target in the second, and was overshot. Kapoula's findings suggested that in low-level oculomotor tasks the range bias is rapidly adaptable, with the range of targets in a stimulus set determining the preferred saccade length. The main aim of the current study was to determine whether the preferred saccade length is also adaptable during reading. As we detail below, people may adopt a preferred saccade length which is determined by their lifetime experience of average word length, due to reading being such a highly structured task in which people acquire substantial experience across their lifetime, compared to the relatively novel lab task conducted by Kapoula.

The systematic saccadic range error is implemented as an important parameter in the two most dominant models of eye movement control, the E-Z Reader model (Reichle, Rayner, & Pollatsek, 1999; Reichle, Rayner, & Pollatsek, 2003) and the SWIFT model (Engbert, Nuthmann, Richter, & Kliegl, 2005; Schad & Engbert, 2012). In both models the saccadic range error is vital for explaining landing positions within words, and whether certain words will be fixated at all. As well as these issues being important in and of themselves, they also affect each model's ability to explain other important phenomena. First of all, as mentioned above the probability of a reader re-fixating a word is strongly influenced by their initial fixation location, and this has also been implemented in both models. Secondly, both models include parameters whereby the processing rate of the fixated word (and any words that are processed beyond the fixated word) is influenced by the visual eccentricity of the letters within those words from the point of fixation. Thus, in both models not only does the systematic range error impact on initial landing position, but also the processing efficiency of word identification.

In addition, the systematic range error in the E-Z model is important in explaining so-called parafoveal-on-foveal effects (see Drieghe, 2011 for a review). These are effects whereby the characteristics of a word to the right of fixation influence the processing of the fixated word. These effects are highly controversial, in that they are seen as evidence in favor of multiple words being processed in parallel, a phenomenon which is viewed as being incompatible with the E-Z Reader model's approach to lexical processing, but very much in line with the SWIFT model. However, the idea of both random motor error and the systematic range error allows the E-Z Reader model to explain these effects through the idea that readers will sometimes fixate a word other than the one they intended to fixate. First of all, the random motor error observed by McConkie et al. will sometimes be great enough that the eye will land on one of the words to either side of the saccade target. Secondly, the systematic range error will sometimes cause a saccade to undershoot by enough characters that it lands on the word before the saccade target. This is especially likely when readers attempt to skip a word, since this requires a particularly long saccade. In these cases fixation durations on the fixated word are determined by the processing difficulty not of the word that is currently fixated but the word that was the original saccade target, thus leading to the observation of parafoveal-on-foveal effects (Drieghe, Rayner, & Pollatsek, 2008). As such, in the case of the E-Z Reader model the systematic range error, alongside random motor error, is important in providing an explanation for an effect that would otherwise be incompatible with this model's core theoretical assumption (i.e. serial lexical processing).

Given the importance of the range error for estimating initial fixation locations in both models, and the various impacts these landing positions have on the models' explanatory power, it is vital that these models accurately implement the range error. In both models the range error is currently programmed so that approximately 0.4 characters of under- or overshoot occurs for every character that a saccade was intended to move beyond or before the preferred saccade length. In E-Z Reader this preferred saccade length is seven characters, while in SWIFT it is 5.4 characters for progressive inter-word saccades. While the two models differ on the preferred saccade length (note that parameter fitting within these two models did not happen within the same language), this is currently treated as a fixed parameter in both. The models are agnostic about how this preference may have developed, but it seems reasonable to suggest that they could both, in principle, be modified to allow for a flexible preferred saccade length without violating any of their key theoretical assumptions.

Relatively little attention has been given to the issue of how exactly the preferred saccade length during reading develops. However, the general assumption seems to be that it develops due to a large amount of reading experience. For example, in a paper examining the development of oculomotor control during reading Reichle et al. (2013) state that it is possible that "…whatever 'tuning' of the oculomotor system occurs and is ultimately responsible for the systematic portion of saccadic error is not completely established in children" (p. 136). This suggests that the preferred saccade length develops over a long period of time as readers gain experience with a language, and as such should not easily be altered in a novel situation. In addition, a recent theoretical account of the systematic range error attempts to explain the preferred saccade length as a Bayesian prior which has developed as a result of experience of the typical distance between word centers during reading (Engbert & Krügel, 2010; Krügel & Engbert, 2014). This prior is combined with an estimate of the location of the center of the next word, resulting in a saccade that is programmed to travel between these two values, and thus the systematic range error.

Presumably, according to both of these approaches the range error should be fixed, and thus not adapt to novel conditions. In contrast, it has clearly been shown by Kapoula (1985) that the range error outside of the context of reading is not fixed. This gives us two competing hypotheses. First of all, it could be that the systematic range error is fundamentally different in reading to other oculomotor tasks, and is indeed determined by a lifetime's worth of reading experience. Alternatively, it could be that the existing assumptions about the range error during reading are erroneous, and that much like in simple oculomotor tasks the preferred saccade length will rapidly adapt to a novel distribution of word lengths. If the former position is true it suggests that the oculomotor system becomes tuned in reading in a way that it does not in other tasks. If the latter position is true then the range error as it is currently approached in models of eye movement control is incorrect, and this may have consequences for how different models are able to accurately predict the various phenomena discussed in the previous paragraphs. Clearly, it is important to determine which of these positions is correct.

Word length and launch site effects on word skipping

Word length also has significant effects on word skipping. Vitu, O' Regan, Inhoff, and Topolski (1995) found that readers skip three, four, and five letter words approximately 60%, 45%, and 30% of the time, respectively. This effect is driven by the spatial extent of the word, rather than the number of letters. Hautala, Hyönä, and Aro (2011) embedded a four or six letter word within a sentence either presented in a font in which the six letter word subtended a greater visual angle than the four letter word, or a font in which both words subtended the same amount of visual angle. When the six letter words subtended a greater visual angle they were skipped less often than the four letter words (6% vs. 21% of the time). In contrast, when both word lengths subtended the same visual angle there was no significant effect of the number of letters on skipping (14% of the time for both word lengths; see also McDonald, 2006). Thus, the decision of whether or not to skip a word seems to be largely driven by a word's spatial extent (see also Hermena, Liversedge, & Drieghe, 2016 for evidence from Arabic). Importantly, words with a greater spatial extent will extend further into the parafovea, and thus will fall in lower acuity vision and will be less likely to be processed as fully as when they are directly fixated (and hence less often skipped).

It is also worth considering whether the decision to skip a word may be partially determined by its relative length within a sentence, rather than simply its absolute length. It may be that the decision about which words to skip will partially be based upon whether certain words are likely to be of high informational value. Informational value refers to the amount of information a word communicates within a sentence; essentially a word which is entirely predictable given the preceding sentential context is of low informational value, while a word which is entirely unexpected given the preceding context is of high informational value. It could be costly for a reader to fixate a word of low informational value, since readers will remain in this position until they can program a new saccade despite there being very little linguistic processing to do either in terms of word identification or sentential integration. Readers may assume that short words are likely to be of lower informational value due to the way in which speakers (and writers) attempt to regulate the rate of information conveyance during communication, to avoid either communicating information too quickly or slowly. One way in which speakers may regulate the rate of information conveyance is through the use of word length. Through analyses of large text corpora Piantadosi, Tily, and Gibson (2011) demonstrated that in normal sentences shorter words typically tend to be of lower informational value than longer words, partly because many short words are function words, which tend to carry very little semantic information. However, this relationship also holds for content words. For instance, when a word is of low informational value (i.e. highly predictable) within a sentence an abbreviated form (e.g. *chimp* for *chimpanzee*) is more likely to be used (Mahowald, Fedorenko, Piantadosi, & Gibson, 2013). Clearly, both *chimp* and *chimpanzee* carry identical semantic information, and so the only reason to use one form over the other is the variation in length, with the longer version slowing down information conveyance and the shorter version speeding up information conveyance. Furthermore, Mahowald et al. demonstrated that this relationship forms a part of people's abstract linguistic knowledge. Given that there is a clear tendency for short words to carry less information in natural language, and the potential cost of fixating a low information value word, it might make sense for readers to skip shorter words. In Hautala et al.'s (2011) study these cues were absent with six-letter words extending across the same spatial distance as four-letter words. The lack of a difference in skipping of short and long words in this study may have partially been due to the absence of relative length information. We will return to this issue below.

As mentioned above, the systematic range error also has an effect on whether participants will fixate a word, and as such affects word skipping rates. A consequence of saccadic overshoot or undershoot is that a word will sometimes receive a mislocated fixation, such that a word that was not the intended saccade target is fixated (Nuthmann, Engbert, & Kliegl, 2005). For example, if a saccade is targeted towards a three letter word from the final character of the prior word, then saccadic overshoot may cause the eye to land on the following word. This would constitute an accidental skip. Furthermore, saccadic undershoot can sometimes cause the eye to land on the word before the saccade target, as discussed above in relation to parafoveal-on-foveal effects. Through simulations, Nuthmann et al. were able to estimate the probability for words of a certain length either receiving a mislocated fixation or being accidentally skipped. While words of all lengths were approximately equally likely to be accidentally fixated or undershot, some were more likely than others to be accidentally skipped (see Table 6 on page 2212 of Nuthmann et al.). While the probability of accidentally skipping a three-letter word was 0.28 this decreased to 0.16 and 0.08 for four- and five-letter words. Thus, due to the saccadic range error short words will accidentally be skipped more often than longer words.

The current study

In order to investigate whether the preferred saccade length is fixed or flexible, we presented participants with sentences which varied in terms of which saccade length would be the optimal preferred saccade length. Participants read four types of sentences. In three of these all of the words were of uniform length, with all words being three-letters long (referred to as U3; e.g. *The sad boy had not had any fun all day*), four-letters long (U4; e.g. *They went over some very hard sums last week*) or five-letters long (U5; e.g. *David often plays awful death metal music about Satan*). In a non-uniform condition, participants read sentences comprised of words with a mixture of these word lengths (NU; e.g. *Tim can often leave work about one hour early*). On average the words within these sentences were 3.94 letters long. Participants read thirty sentences of each type in a block. We hypothesized that if the preferred saccade length which determines the systematic saccadic range error during reading is adaptable then it should vary between these different types of sentences. Across the different types of sentence the average distance between saccade targets will systematically vary, with inter-word saccades of an average of four, five, and six characters being required to move between the centers of words in uniform sentences consisting of three, four, and five letter words, respectively (i.e. the length of each word, in addition to a space). In the non-uniform sentences the distance between saccade targets will be more variable, but should on average be about five characters (i.e. 3.94 letters per word, in addition to a space). Naturally, as the distance between saccade targets decreases readers would be expected to attempt to make shorter saccades, and so a change in the average saccade amplitude across our sentences would be entirely uncontroversial. The more interesting issue is whether participants become more accurate at making saccades of this average length for a certain sentence type. For example, participants will generally attempt to make four character saccades more often in the uniform sentences of three letter words than the uniform sentences of five letter words, due to this being the distance between the center (i.e. a common landing position and thus launch site) of one three letter word and the center of the next, but the distance between the final character of one word (i.e. an uncommon landing position and thus launch site) and the middle of the next for five letter words. It could be that the preferred saccade length is fixed at seven characters. Assuming 0.4 characters of error for every character of deviation between the preferred saccade length and intended saccade length, as E-Z Reader and SWIFT do, a saccade launched from four characters away from an upcoming word center will overshoot by 1.2 characters in both sentence types if the preferred saccade length is fixed at seven characters. Conversely, if the preferred saccade length is flexible towards the average word length in a sentence then participants will not overshoot at all for these saccades in uniform sentences of three letter words with a preferred saccade length of four characters, and will only overshoot by 0.8 characters when making these saccades in uniform sentences of five letter words with a preferred saccade length of five characters. Essentially, we predicted that if the preferred saccade length was flexible and readers adjusted it contingent on the length of words in the text being read then it would be smallest when participants read sentences consisting of three letter words, slightly larger when reading the non-uniform sentences and the uniform sentences of four letter words, and larger still when reading uniform sentences of five letter words. Alternatively, as models of oculomotor control and accounts of the systematic range error in reading currently assume, it could be that the preferred saccade length has been determined by a lifetime's worth of reading experience. If this was the case then participants are unlikely to adopt different preferred saccade lengths for our different sentences.

We also predicted significant effects of our manipulations on skipping rates. First, we expected standard word length effects, such that for sentences comprised of short words we would see increased word skipping. We also made a hypothesis contingent upon our saccadic adaptation effect. As mentioned, short words are accidentally skipped due to the systematic saccadic range error more often than long words. Presumably, if readers become more adept at making short saccades in the uniform sentences of three letter words then there should be less accidental skipping of three letter words in these sentences than in the non-uniform sentences due to saccadic overshoot. Experimentally determining whether a word was skipped accidentally or on purpose is not possible, and as such we can only assess this hypothesis on the basis of whether the preferred saccade length adapts, and whether this is accompanied by systematic variation in word skipping rates.

We were also curious to see if sentence uniformity had a further effect on word skipping due to the absence of relative word length cues within sentences. First of all, we expected to observe a substantial decrease in the skipping of three letter words appearing in a uniform compared to non-uniform sentence, simply due to the fact that maintaining a global skipping rate of 60% (as it typically observed for three letter words in non-uniform sentences, e.g. Vitu et al,. 1995) across an entire sentence would involve encoding visual information from very few positions in that sentence. As such it seems unlikely that participants would skip three letter words as often in a sentence consisting exclusively of words that length, than when these words appeared in sentences of mixed word length. Furthermore, if readers do use relative word length information as a cue to the informational content of a word, and therefore the necessity for direct fixation, we may expect that while a three letter word may be more likely to be of low informational value in a non-uniform sentence, this would be less the case in a sentence comprised entirely of three letter words. Consequently, three-letter words may be less frequently skipped in uniform than non-uniform sentences. The converse may be true for five letter words, with these typically being more likely to be of increased informational value in non-uniform sentences, but less so in a sentence comprised entirely of five letter words.[1] In short, we may observe decreased skipping of three letter words and increased skipping of five letter words in uniform relative to non-uniform sentences.

Method

Participants

24 students at the University of Southampton with normal or corrected to normal vision participated in return for course credits.

Apparatus

Participants' eye movements were monitored using an SR Research Eyelink 1000 system with a sample rate of 1000 hertz. Only the right eye was tracked. Sentences were displayed in black on a grey background, on a single line of a ViewSonic p227f CRT monitor. Viewing distance was 78cm, with 1° of visual angle being occupied by 2.9 characters of monospaced Courier font.

Materials and Design

Four sets of thirty sentences were created for the current experiment. In three of these sets all words in a sentence were of uniform length, such that the sentences were comprised entirely of three letter words, entirely of four letter words, or entirely of five letter words. The fourth set consisted of sentences made of a combination of three, four, and five letter words.

A within subjects design was used, such that all participants saw all 120 sentences. The sentences consisted of between eight and twelve words.

Eighteen participants rated our sentences for naturalness on a scale from one (very unnatural) to five (perfectly natural). The mean rating for the uniform sentences of three, four, and five letter words, and the non-uniform sentences were 3.69, 3.45, 3.47, and 3.87, respectively. A one-way ANOVA revealed significant differences between the four conditions, $F(3, 51) = 23.46$, $p < .001$. Follow up t-tests comparing the non-uniform sentences to each uniform sentence type revealed that the non-uniform sentences were rated as being significantly more natural than all three uniform sentence types (NU vs. U3 $t(17) = -2.58$, $p < .01$; NU vs. U4 $t(17) = -5.78$, $p < .001$; NU vs. U5 $t(17) = -5.34$, $p < .001$). Consequently, potential effects of naturalness were incorporated in post-hoc analyses.

We also examined the frequencies of the words making up our sentences, as a function of word length and uniformity. We conducted t-tests to examine whether the frequency of the words of one length in the uniform sentences was the same as words of that length in the non-uniform sentences. Frequencies were based on the Zipf scale introduced by van Heuven, Mandera, Keuleers, and Brysbaert (2014). There were no differences in the frequency of five letter words appearing in a uniform ($m = 5.02$) relative to non-uniform ($m = 5.09$) sentence ($t(109.5) = -0.62$, $p > .1$), or of four letter words in a uniform($m = 5.60$) relative to non-uniform ($m = 5.75$) sentence ($t(185.4) = -1.49$, $p > .1$). However, three letter words appearing in uniform sentences ($m = 5.87$) were significantly less frequent than three letter words appearing in a non-uniform ($m = 6.38$) sentence ($t(112.9) = -3.65$, $p < .001$). Furthermore, it should be clear from the above means that differences in frequency did exist as a function of word length with 3-letter words being more frequent than 4-letter words, which in turn were more frequent than 5-letter words. Therefore, we assessed whether the effects of word length and sentence uniformity were modulated by differences in word frequency across conditions.

Procedure

Participants completed the experiment in two sessions across two days, each lasting twenty minutes. In each session participants were presented with two blocks of thirty experimental sentences, preceded by six practice trials. All sentences from a single condition were presented in the same block. We adopted a blocked design to maximize our chances of obtaining adaptation effects.

Upon arrival at the first session participants were presented with a consent form and information sheet. They were seated in front of the eye tracker and a head rest was used for stabilization. A three point horizontal calibration grid was used, with an acceptance criterion of an average error below 0.25 degrees.

After a successful calibration the experiment began. Each trial began with two drift checks. The first drift check was in the center of the screen, while the second was on the left in the position of the central character of the first word of the sentence. If either indicated more than 0.3 degrees of error the participant was recalibrated. After the drift checks had been completed the experimental sentence appeared. Participants were instructed to read for comprehension, and press a button once they had read the sentence. On 33% of the trials the experimenter read the participant a yes/no comprehension question, and the participant responded using a button box. Across all participants 91% of comprehension questions were answered correctly. At the end of the second session the experimenter asked the participant if they had consciously noticed the experimental word length manipulation. Only one participant had.

Results

We computed global and local measures pertaining to each theoretical issue outlined above. The global measures were calculated across whole sentences (see Table 1). The four sentence types were compared using linear mixed effects models, constructed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (R Development Core Team, 2013). Uniform sentences of four letter words were treated as the baseline, due to being the most appropriate condition to compare our non-uniform sentences with due to the similar average word length, and to allow us to compare these sentences to uniform sentences with shorter and longer words. Model output is shown in Table 2. For all models we initially adopted a full random structure, treating both participants and sentences as random factors, with random intercepts and slopes (see Barr, Levy, Scheepers, & Tily, 2013). However, due to the models for some measures failing to converge, random slopes were sometimes removed.

Our local measures examined the effect of our manipulation on each individual word within our sentences. The interest area used for each word included the space preceding the word.[2] After removing data for the first or last word of a sentence, and cases of tracker loss, 21592 words were available for analysis (out of a full data set of 28176). Of these 21592 words, 17080 received a direct first pass fixation, allowing them to be used in the landing position analysis, as well as the reading time measures presented in the Appendix. For each measure that we examined we excluded values that were more than 2.5 standard deviations from the mean of each participant in each condition as outliers. The amount of data excluded due to this final criterion varied between measures, with a maximum of 3.2% of the remaining data being lost. All models described treated both subjects and individual words as random factors, with both random slopes and intercepts. Similar to our global analyses, the models for some measures failed to converge with the full random structure, and as such random slopes were sometimes removed.

Saccadic targeting

At a global level our manipulation had a clear effect on saccade metrics (see Table 1). The average saccade amplitude increased with the average word length in a condition, with the shortest average saccade length being observed in the uniform sentences of three letter words, and the longest in the uniform sentences of five letter words. Unexpectedly, participants made longer saccades in the non-uniform sentences compared to the uniform sentences of four letter words. This is most likely due to differences between these conditions in word skipping, which will be discussed in detail below. The linear mixed-effects models (see Table 2) revealed that the differences between the uniform sentences of four letter words and the other conditions were significant. Clearly as the distance between the centers of two words decreased, so did the amplitude of saccades moving between them. As mentioned above, our uniform sentences were rated as being slightly less natural than our non-uniform sentences. In order to ensure that this had not affected the length of the saccades made in each sentence type we performed an additional analysis of our data for average saccade amplitude in which we included naturalness as a predictor variable. This did not improve the fit of our model to the data [$\chi 2(1) = 2.56$, $p > .1$], suggesting very little effect of naturalness on saccadic targeting.

Given that saccade amplitudes varied between conditions, we next considered whether this affected the systematic range error. Primarily, we were interested in seeing whether the preferred saccade length was constant across our different types of uniform sentences, or whether it increased systematically with the length of words within these sentences. Similarly, as a control, we were interested in seeing whether the preferred saccade length for words of different lengths within non-uniform sentences was near constant, as in McConkie et al.’s (1988) original investigation. If the preferred saccade length did adapt then it should have increased alongside the length of words between the three different uniform sentence types, while any differences between the word lengths in non-uniform sentences should be minimal. Given this theoretical question, and that the preferred saccade length is the distance between the saccade launch site and the center of the word on which the eyes land at which saccades are most accurate, all the analyses below calculated the saccade launch site relative to the center of the word to which a saccade was made.[3]

Linear mixed effects models. As a first step in our analysis, we constructed linear mixed effects models with the fixation landing position within each word as a dependent variable. The landing position was calculated relative to the word center, such that a landing position of 0 corresponded to a saccade landing perfectly in the center of the word, while landing positions of 1 or -1 corresponded to a saccade over- or undershooting the center of the word by one character, respectively. This dependent variable was examined using a linear mixed-effect model with word length, sentence uniformity, and saccade launch site as fixed factors, in addition to two- and three-way interactions between these factors. Words with a launch site greater than 12 characters were not included in this analysis due to visual inspection of the launch site distributions identifying them as outliers, and the fact that it is unlikely that the fixated word was in high enough acuity vision for readers to make an accurate estimate of the word center. This exclusion accounted for 1.63% of the remaining data.

There were significant main effects of word length and launch site (see Table 3 for model output, and Figure 1 for estimated effects). As the launch site distance decreased, fixations landed further into a word, thus replicating McConkie et al.'s original observation of the effect of saccade launch site on fixation landing positions. Word length also had a significant main effect, such that the eyes would land further into a long word.

In addition to these main effects, the model also revealed several significant interactions. Sentence uniformity interacted with launch site in a two-way interaction, and with both word length and launch site in a significant three-way interaction. Clearly the word length uniformity within our sentences had an effect on the systematic range error, with it modulating the effect of launch site on landing positions. There was also a significant two-way interaction between word length and launch site. The nature of all these interactions can most clearly be seen by examining Figure 1. In the case of saccades made from a near launch site (i.e. 3 or 4 characters) in non-uniform sentences, participants overshot the center of five letter words to the greatest extent and the center of three letter words to the smallest extent. For saccades made from far launch sites (i.e. 7 or more characters) in the non-uniform sentences participants undershot the center of five letter words to the greatest extent, and three letter words to the smallest extent. For intermediate launch sites within non-uniform sentences participants neither undershot nor overshot the word center. Crucially, in terms of the preferred saccade length, the linear-mixed model predicted that, in the case of all three word lengths in the non-uniform sentences, participants would accurately fixate the word center when making saccades of between 5 and 6 characters. To obtain estimates of the point at which participants would accurately fixate the word center we used the Effects library (Fox, Weisberg, Friendly, & Hong, 2015) in R, and examined the output for the launch site that predicted a landing position of 0. The model predicted that participants would land directly in the word center (i.e. a landing position of 0) given launch sites of 5.29, 5.63, and 5.85 characters for three, four, and five letter words, respectively. This suggests that within our non-uniform sentences the preferred saccade length was between five and six characters, with any differences on the basis of word length being reasonably small. The average of these three preferred saccade lengths was 5.59 characters. While this preferred saccade length differs quantitatively from that obtained by McConkie et al. (1988), potentially due to the lower average word length in our non-uniform sentences, our results do qualitatively replicate the original finding of a similar preferred saccade length for words of different lengths in normal English sentences.

A different pattern of results emerged for saccades made within our uniform sentences. Similarly to the non-uniform sentences, participants overshot the center of a five letter word from near launch sites to a greater extent than the center of three letter words. However, while in the non-uniform sentences the effect of word length quickly reversed with an increase in launch site, such that people undershot the center of a five letter word to a greater extent than the center of a three letter word from a far launch site, this was not the case across the uniform sentences. What is more interesting to our investigation of the adaptability of the preferred saccade length is the launch site from which participants would on average land perfectly in the center of words of each length, and the way in which this varied across our three different types of uniform sentences. To estimate this we again used the Effects package in R. Participants would land directly in the center of three, four, and five letter words from launch sites of 4.52, 5.41, and 6.14 characters, respectively. Thus, while the preferred saccade length was relatively similar for words of different lengths in the non-uniform sentences, it systematically varied across our three uniform sentence types. This is highly suggestive of adaptation in the preferred saccade length dependent on sentential word length context.

In addition to the main analysis presented above, we briefly considered whether our adaptation effect increased throughout an experimental block, or whether it was rapidly established, remaining constant across the majority of trials. To investigate this we constructed linear mixed effects models with trial index included as an additional predictor. We added this first as a main effect, and then as part of interactions with other factors. Model comparisons were conducted with each increase in model complexity. While a model including only a main effect of trial index significantly improved the fit of the model to our data [$\chi 2(1) = 1995$, $p < .001$], allowing trial index to interact with other factors in two-, three-, or four-way interactions did not further improve our model. Furthermore, while the inclusion of trial improved the fit of our model, this effect was not significant within the model. The lack of interaction between trial index and other factors suggests that our adaptation effects were very rapidly, if not immediately, established.

As mentioned above there were small differences in the frequencies of the words in our different sentence types. To assess whether this had any influence on our results we included frequency in our model as a main effect, and in several interactions. The most complex model which led to an improvement in the fit to the data included an interaction between frequency and word length. All of the effects from our original model remained significant in this model, alongside a main effect of frequency (b=-0.11, SE=0.05, t=-1.96) and the interaction between frequency and word length (b=0.05, SE=0.01, t=3.47). Effects estimates showed that as frequency increased, landing positions moved further into a word. This effect was largest for five letter words such that the landing position increased by 0.5 characters between words with a zipf frequency of three and seven. Frequency effects for the three letter words were negligible, such that mean landing position only increased by 0.13 characters between words with a zipf frequency of three and seven. Recall that it was only for three letter words that mean frequency differed between uniform and non-uniform sentences; thus, frequency had very little effect on landing positions for the word length for which this could have become problematic. Furthermore, we obtained estimates of the preferred saccade length from this new model, and obtained very similar results as from our original model; three, four, and five letter words in uniform sentences had preferred saccade lengths of 4.46, 5.45, and 6.25, whereas in non-uniform sentences the preferred saccade lengths were 5.14, 5.6, and 5.9. Thus, our adaptation effects were still present when accounting for word frequency.

Fixation landing position distributions. In addition to our linear mixed effects analyses, we conducted a further analysis which was more in line with that presented by McConkie et al. (1988). To conduct this analysis, we transformed our continuous landing position data into discrete values, such that a fixation was categorised as landing on a certain letter within a word (or the space preceding it). Figure 2 shows the distribution of initial fixations on each letter of a word, as a function of word length, sentence uniformity, and saccade launch site. Saccade launch sites of 1 to 7 characters from the left of the space preceding the fixated word were used, as in McConkie et al. A landing position of 0 represents the space preceding the word, a landing position of 1 the first letter of the word, and so on.

The standard launch site effect can be seen in Figure 2, with a greater proportion of fixations being made on the earlier characters of a word as the distance of the launch site increased relative to the start of the targeted word. This is the case for all three word lengths, in both uniform and non-uniform sentences. Thus, we replicated McConkie et al.'s original findings.

It could be argued that differences should be present for words appearing in uniform relative to non-uniform sentences in these landing position distributions. The logic behind this argument is that the landing position of a fixation within a word should be influenced by the distance from which the saccade was launched, and the extent to which this launch site distance deviates from the preferred saccade length. Furthermore, our linear mixed effects model demonstrated that the preferred saccade length systematically varied between our different sentence types, such that a word of a specific length is subject to a different preferred saccade length when it appears in a uniform relative to non-uniform sentence. Due to this, the distribution of landing positions from each launch site should differ depending upon sentence uniformity, since the deviation of the launch site from the preferred saccade length will differ under these two different conditions. Despite this, there is very little evidence in our landing site distributions for these differences.

While this seems to run contrary to the idea that we observed adaptation in the preferred saccade length, when taking into account both the typical magnitude of the systematic range error and the size of the adaptation in the preferred saccade length, this outcome is not particularly surprising. As discussed above, the range error leads to 0.40 characters of systematic error for each character that the planned saccade length deviates from the preferred saccade length. We can gain a rough estimate of how much the preferred saccade length changed for a word of a certain length by taking the estimate of the preferred saccade length for that word in a uniform sentence from the preferred saccade length for a word of the same length in the non-uniform sentence. In the case of four letter words this equates to a change of 0.22 characters, and for five letter words a change of 0.29 characters. From any given launch site these changes in the preferred saccade length will lead to differences of 0.09 and 0.12 characters of systematic error, respectively. It is not surprising that a shift of such a small magnitude does not lead to a large shift in landing site distributions, especially considering that such subtle effects could have been lost when transforming our continuous landing position data into discrete, whole character values.

Relative proportions of saccades under and overshooting. As a final step in our analysis, we also considered the relative frequency with which participants under and overshot the central character of a word from any given launch site in each condition. To examine this issue we categorized our landing position data in terms of whether a fixation landed within a character sized region in the center of a word, in the area to the left of this central region (i.e. undershoot), or in the area to the right of this central region (i.e. overshoot), and plotted this as a function of word length, sentence uniformity, and launch site (see Figure 3). While this categorisation is a coarser measure of fixation landing positions compared to the more typical distributions presented in Figure 2, it allows us to observe any systematic variation in the saccadic range error and preferred saccade length more clearly. By considering the launch sites between which participants make a transition from predominantly undershooting the center of a word to predominantly overshooting that point, it is possible to obtain further confirmation of the variations in preferred saccade length across different sentence types obtained from our linear mixed effects models.

Predictions of the preferred saccade length derived from our linear mixed effects models are as follows. Recall that the linear mixed effects models predicted preferred saccade lengths of 5.29, 5.63, and 5.85 characters for words in these sentences. As such, it seems reasonable to predict that participants should predominantly overshoot for launch sites nearer than these values (i.e. five characters or less) and undershoot for launch sites further away (i.e. six characters or more). As such, readers should transition from over- to undershooting between launch sites of 5 and 6 characters. In contrast, the transition point will vary across the three different uniform sentence types, with the predictions of the preferred saccade length from our linear mixed effects models being 4.52, 5.41, and 6.14 characters for three, four, and five letter words. For each uniform sentence type we should expect the transition point to vary around these values, and as such it should occur between launch sites of 4 and 5, 5 and 6, and 6 and 7 characters for uniform sentences of 3, 4, and 5 letters, respectively.

For the most part, we observed the exact pattern of effects we predicted, with the transition from predominantly overshooting to undershooting occurring at the point we predicted for all three word lengths in uniform sentences, and the four and five letter words in non-uniform sentences. In the case of three letter words in the non-uniform sentences the transition actually occurred slightly later, with readers still overshooting very slightly more often than undershooting from a launch site of six. The fact that the transition point was very similar for different word lengths in non-uniform sentences, but increased systematically for different word lengths in the uniform sentences is entirely consistent with the findings from our linear mixed effects model.

Overall saccadic accuracy and refixation probabilities. In addition to considering whether our manipulation led to an adaptation of the preferred saccade length, we examined whether this resulted in more accurate overall saccadic targeting. We did this by considering the mean landing position and proportion of initial fixations made on the central character of words of different length in uniform relative to non-uniform sentences. The mean landing position tended to be slightly to the left of the word center across all conditions, with very little difference between words appearing in a uniform versus non-uniform sentence (see Table 4). Furthermore, the proportion of fixations landing on a word's central character also varied very little between uniform and non-uniform sentences (see Table 4). Thus, it seems that an adaptation in the preferred saccade length does not translate into more accurate saccadic targeting overall. This may well have occurred as a result of random motor error remaining relatively constant across conditions, and thus leading to a similar distribution of landing positions around the preferred viewing location. Furthermore, these landing positions would then function as the next launch site, which would presumably be non-optimal for the preferred saccade length within the sentence being read.

In summary our manipulation of sentence uniformity and the length of words within these sentences had significant effects upon saccadic targeting. First, saccade amplitudes were affected by the length of words within a sentence, such that longer saccades were made for sentences comprised of longer words. Furthermore, variation in saccadic targeting led to an adaptation of the systematic range error, as demonstrated across three different approaches to analyzing the landing position data. Participants were more attuned towards making short saccades in the uniform sentences of three letter words, and more attuned towards making long saccades in the uniform sentences of five letter words, with the preferred saccade length increasing across the three different uniform sentence types. This pattern of effects was very consistent across three different approaches we adopted in our analysis.

Fixation probabilities and word skipping

In order to examine the effect of our manipulation on word skipping we calculated the proportion of words fixated in each sentence (see Table 1). This global measure was calculated as instances when a word neither received a direct first pass fixation, nor was regressed to later on. Clear word length effects were observed, such that a lower proportion of words were directly fixated in uniform sentences consisting of short words as opposed to long words. However, skipping was no greater in the uniform sentences comprised of three letter words than non-uniform sentences, despite the non-uniform sentences on average containing longer words (*m*=3.94). Thus, while word skipping increased when there were more short words in a sentence, the effect is limited.

We also examined skipping of individual words as a function of their length, whether they appeared in uniform or non-uniform sentences, and their frequency. In the case of this local measure a word was classed as skipped whenever readers did not directly fixate it during first pass reading, regardless of whether they subsequently regressed onto it. As a reminder, we expected a large decrease in the skipping of three letter words in uniform relative to non-uniform sentences, and a slight increase in the skipping of five letter words in uniform relative to non-uniform sentences. We also predicted main effects of length and frequency, two variables that are established as important predictors of word skipping (see Brysbaert, Drieghe, & Vitu, 2005). Highly frequent words and short words were skipped significantly more often than less frequent words and longer words (see Table 4 for means and Table 5 for LME output). There was also a significant effect of sentence uniformity, which interacted with word length, such that three and four letter words were skipped more often while five letter words were skipped less often in non-uniform sentences than in uniform sentences.

Discussion

In the current study we examined the effect of within sentence word length uniformity on eye movement control during reading, with a particular focus on how this manipulation affects saccadic targeting. Several novel effects were observed. First, there was clear evidence of an adaptation of the systematic range error first observed by McConkie et al. (1988), with the preferred saccade length within a sentence being systematically modulated by the length of words in each sentence type. This finding has implications for models of oculomotor control, and accounts of the systematic range error during reading. Furthermore, our manipulation affected word skipping, such that the probability of skipping short words dramatically decreased and the probability of skipping long words slightly increased in uniform relative to non-uniform sentences.

Saccadic targeting and the systematic range error

The observation of the systematic saccadic range error by McConkie et al. (1988) has had long lasting implications for the understanding of eye movement control during reading. It is an important parameter in explaining the landing positions of fixations, and is thus implemented in both the E-Z Reader and SWIFT models of eye movement control. The accurate implementation of the systematic range error within both models is vital for explaining phenomena such as re-fixation rates, word processing speeds, and in the case of the E-Z Reader model, also parafoveal-on-foveal effects. Currently both models incorporate the preferred saccade length as a fixed parameter, optimized for saccades traversing the average distance between the centers of two words in English (E-Z Reader) and German (SWIFT).

Contrary to the implemented versions of both computational models discussed above, our results show that there was clear adaptation in saccadic targeting across our uniform sentences in terms of the preferred saccade length. The estimates from our linear mixed models demonstrated systematic variation on the basis of the mean word length within the sentence being read. The preferred saccade length increased across the three different uniform sentence types, with preferred saccade lengths of 4.52, 5.41, and 6.14 for uniform sentences of three, four, and five letter words. In contrast, any variations between these three word lengths in the non-uniform sentences were more modest, with preferred saccade lengths of 5.29, 5.63, and 5.85. This represents a clear and systematic modulation of the preferred saccade length by the current word length context. Our results replicated McConkie et al.'s original finding that landing positions within words are driven by the saccade launch site, in addition to extending those findings by demonstrating an adaptation of the preferred saccade length under different word length contexts. This adaptive modulation of the preferred saccade length is something that should be considered in future implementations of models of eye movement control, due to the large range of phenomena for which an accurate implementation of the systematic range error is necessary for explaining.

It is worth considering the launch site from which on average the preferred saccade length was launched from for our three uniform sentence types. Figure 4 shows an example of part of each uniform sentence type, with the preferred saccade length to the word center plotted above. The quantitative adaptation in the preferred saccade length essentially allowed participants to perform the same qualitative behavior in all three uniform sentence types, with participants adapting to move from the preferred viewing location of one word to the center of the next. This aspect of our results supports our hypothesis that participants would adapt their preferred saccade length alongside the distance between the centers of two words in a sentence. Clearly, the saccadic targeting system tuned the preferred saccade length in order to optimize performance in terms of fixation locations.

Our results contribute to a literature suggesting a high degree of adaptability during reading. Kaakinen and Hyönä (2014) demonstrated that the extent of the effective visual span during reading is modulated by task demands, and the relevance of a particular piece of text to these task demands. Schotter, Bicknell, Howard, Levy, and Rayner (2014) have shown that the extent to which word frequency and predictability influence fixation durations can be independently modulated by different proofreading tasks, in order to perform these tasks as efficiently as possible. The results from these two studies suggest that the deployment of attention and nature of linguistic processing can adapt during reading based on explicit task demands. Our findings demonstrate that this adaptation extends all the way down to the parameters influencing the execution of saccades. Furthermore, the adaptation in our study occurred despite participants not being conscious of the manipulation, suggesting that these effects are highly automatic.

One surprising aspect of our findings is the time course across which adaptation occurred. An additional analysis of our data showed that allowing trial number to interact with the other variables in our model for landing position did not improve the fit of the model. This could suggest that our effects were instantaneous rather than a gradual adaptation process, though this would be surprising, since presumably at least some practice is required for a change to occur. Due to this, we believe that adaptation very likely occurred across a small number of trials, and quickly plateaued (i.e., adaptation was very rapid, rather than instantaneous). It is clear that future work is needed to examine the time course of our effects.

Our adaptation effects may be difficult to account for within current theories of the systematic range error. As discussed above, the accepted view generally seems to be that the preferred saccade length develops as a result of a large amount of reading experience. One such theory describes the systematic range error in terms of Bayesian decision theory (Engbert & Krügel, 2010; Krügel & Engbert, 2014). According to this approach the oculomotor system makes use of prior knowledge about the distribution of the distance of previous saccade targets from the saccade launch site when estimating the location of the next saccade target. The prior in this theory is essentially equivalent to the preferred saccade length. This prior is combined with a sensory estimate of where the center of the next word is to yield an optimal estimate of the distance of the saccade target. This combination results in the systematic range error. When a target (and thus its sensory estimate) is far away relative to the prior distribution then the Bayesian combination results in the optimal estimate being nearer than it actually is, and thus saccadic undershoot. When a target is near relative to the prior then the optimal estimate will lead to a saccadic overshoot.

While our findings do not directly contradict the assumptions of this theory, they do suggest that further specification may be required regarding the establishment of the prior. Currently, this theory does not specify the time course across which prior information is taken into account. The simplest assumption would be that the prior is determined over an extended period of reading experience, rather than constantly being reset due to slight changes in average word length. Under this assumption it is unlikely that this theory can explain our findings. In our study each condition consisted of only thirty trials, and analyses taking trial index into account suggested that the adaptation occurred across the early trials, rather than gradually across all thirty trials. Even if the prior was instantaneously updated with the new information, it seems unlikely that enough information would have accumulated in the early trials of a block to alter a prior established over years of reading experience.

Rather, we would expect to see a gradual adaptation across the entire block, as more and more information is integrated into the existing prior. Under such circumstances it is unclear how a substantial adaptation of the preferred saccade length would occur, since the optimal estimate of a saccade target would be biased to a considerable extent by the same prior. In order for this approach to explain our findings it may be necessary for the prior to be based on only a recent number of trials, or to allow multiple priors, which readers switch between dependent upon the current distribution of target distances. Berniker, Voss, and Kording (2010) recently demonstrated that it is possible for people to rapidly estimate multiple Bayesian priors for the same perceptual task. In this study participants had to determine the location of a target on the basis of a noisy perceptual cue and their estimate of the prior distribution of the target's location. The mean of the target's position (and thus the prior) was varied between one of two values, with the current distribution switching throughout the experiment. Participants performed in a manner congruent with Bayesian decision theory, and their performance suggested that they estimated the mean of each prior within ten trials. Furthermore, once both priors had been learnt, participants were able to detect a switch in target distributions within two trials, and begin using the appropriate prior. Thus, it is certainly possible, within the context of Bayesian decision theory, for participants to learn and make use of multiple priors. Given our findings, this is something that should be considered within Engbert and Krügel's (2010) model. It should also be noted that if the systematic range error does occur as a result of Bayesian combination of a prior distribution and noisy perceptual cue, then our study extends the findings of Berniker, Voss, and Kording (2010). In their study, they demonstrated that a Bayesian prior that had been built up during a novel laboratory task could adjust to a new distribution of target locations, whereas we have shown that a prior that has built up over many years can adjust.

The fact that the preferred saccade length adapts should perhaps be unsurprising, considering that readers will regularly encounter texts differing in terms of the average of and variability in word length. Within a language, mean word length will vary across texts produced with different purposes in mind. For example, a scientific paper about particle physics will contain longer words than an article in a tabloid newspaper, and the words in a tabloid newspaper may be longer than in something that is read to a child as a bedtime story. It is plausible that all three types of writing will be read by a single person in a day; as such a flexible and rapidly adaptable preferred saccade length is necessary in order for readers to efficiently process a range of texts written for different purposes.

Adaptation may also be necessary for multilingual readers. Different languages vary in the length of the words used. For example, dramatic differences exist between Chinese, English, and Finnish, both in terms of the variations in word length and the average word length. Liversedge et al. (2016) translated passages of text between these three languages, finding differences in the word lengths in each language. In Chinese, English, and Finish the average word length was 1.55 Chinese characters, 5.63, and 8.32 alphabetic characters with standard deviations of 0.20, 0.80, and 1.44 respectively. Furthermore, the saccade length of participants reading in each language varied, with average rightward saccades of 3.19, 8.53, and 9.35 characters. Clearly, different preferred saccade lengths would be appropriate for these different situations, and flexibility would be vital for bi- or trilingual readers of these languages. It might even be the case that the adoption of a preferred saccade length is more appropriate in a language such as Chinese with relatively little variation in word length, relative to a language like Finnish with highly variable word length.

Further research is required to fully understand the nature of the adaptation observed in the current study. Some issues that need to be addressed include the time course of the adaptive process, the extent to which adaptation occurs when average word length varies between sentences but is not uniform within sentences, and the extent to which our effects were driven by word length in letters as opposed to spatial extent (see Yao-N'Dré, Castet, & Vitu, 2014). Investigation of these issues will allow the development of a more comprehensive theory of the systematic range error, as well as a greater understanding of the manner in which the saccadic targeting system adapts to the demands of a specific task.

Skipping probabilities

We observed clear word length effects on skipping. The proportion of words being skipped was higher in the uniform sentences of short words than of long words. Furthermore, within the non-uniform sentences, short words were skipped more than long words. There was also an effect of uniformity. The proportion of words skipped was as high in the non-uniform sentences as in the uniform sentences of three letter words, despite the latter condition consisting entirely of words that should be highly likely to be skipped. Furthermore, skipping of three and four letter words was lower in the uniform than non-uniform sentences, while skipping of five letter words was actually slightly higher.

It should perhaps not be surprising that the skipping of three letter words decreased so dramatically in uniform relative to non-uniform sentences. While three letter words were skipped 40% of the time in non-uniform sentences the skipping rate in these sentences across all word lengths was 20%. A skipping rate of 40% across an entire sentence may leave too few fixations to visually encode information. Thus, it is likely that some of the decrease in skipping of three letter words was due to a need to encode visual information from multiple points in the sentence. While this certainly explains a large amount of our observed effect, it is worth considering whether other factors proposed earlier may also have contributed.

One factor was that readers might use the relative length of a word in a sentence to make inferences about its informational value, and whether it requires a direct fixation. In non-uniform sentences this may contribute to the greater skipping of short words (i.e. three and four letters) relative to long words, since short words are likely to be less informative. When this cue is removed in the uniform sentences, the proportion of skipping driven by this source of information would no longer occur, explaining some of the change in skipping rates. It should be noted that if this argument is correct then this would constitute a demonstration of the adaptability of the cognitive system during reading. Additionally, prior research examining the link between word length and informational value has generally focused on language production, by examining whether speakers or writers regulate informational flow using word length. However, to the best of our knowledge research has not focused upon whether readers make use of cues to informational value in order to regulate the rate of information intake. Our data suggest this may be the case, with readers potentially avoiding fixating on words that are of low informational value. Further research is required to evaluate whether this account is correct, or if the effect of relative word length can be explained solely by the need to fixate a minimum proportion of words in a sentence.

A decrease in both accidental and failed skipping due to the saccadic range error may also have contributed to the alteration in skipping behavior. Often a three letter word will be skipped after an intended short saccade as a result of saccadic overshoot, while a five letter word may sometimes be accidentally fixated after a long saccade targeted towards the following word, as a result of saccadic undershoot (see Nuthmann, Engbert, & Kliegl, 2005). In the uniform sentences of three letter words the preferred saccade length was shorter than in the non-uniform sentences, while in the uniform sentences of five letter words it was greater. As such, a saccade targeted towards a three letter word from a near launch site was less likely to overshoot in these uniform sentences than the non-uniform sentences, while saccades programmed to skip a five letter word from any given launch site would be less prone to undershooting in the uniform sentences. This may have contributed to the decreased skipping of three letter words and increased skipping of five letter words in uniform sentences. It is difficult to be certain of the exact extent to which the adaptation in the preferred saccade length may have affected skipping rate due to mislocated fixations, but it seems reasonable to assume that it had some influence.

Conclusion

In closing, the current paper examined the effect of within sentence word length uniformity on eye movement control during reading. The most striking effect of this manipulation was on the saccadic targeting system, with the preferred saccade length which determines the systematic range error adapting contingent upon the length of the words within a sentence. The main implications of this effect are that assumptions that the preferred saccade length is determined by an extensive amount of experience and is thus fixed are incorrect; as such models of eye movement control currently attempt to explain landing positions (and, consequently, refixation rates, processing efficiency, and parafoveal-on foveal effects) on the basis of an erroneous assumption. We also observed effects of our manipulation on the skipping probabilities of words within the sentences. We argued that the effect on skipping was due to a combination of needing to fixate a minimum number of words in a sentence, as well as relative word length information and the adaptation of the preferred saccade length. Altogether our findings suggest a high degree of adaptability in the saccadic targeting system during reading.

References

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4 classes* (R Package Version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Berniker, M., Voss, M., & Kording, K. (2010). Learning priors for Bayesian computations in the nervous system. *PLoS One, 5*: e12686. doi:10.1371/journal.pone.0012686

Brysbaert, M., Drieghe, D., Vitu, F. (2005). Word skipping: Implication for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53-77). Oxford: Oxford University Press.

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 839-855). New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199539789.013.0046

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

Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science, 21,* 366-371. doi:10.1177/0956797610362060

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

Fox, J., Weisberg, S., Friendly, M., & Hong, J. (2015). *Effect Display for Linear, Generalized Linear, and Other Models* (R Package Version, 3.0-4). Retrieved from http://CRAN.R-project.org/package=effects

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research, 51,* 1719-1727. doi: 10.1016/j.visres.2011.05.015

Kaakinen, J. K., & Hyönä, J. (2014). Task relevance induces momentary changes in the functional visual field during reading. *Psychological Science, 25,* 626-632. doi: 10.1177/0956797613512332

Kapoula, Z. (1985). Evidence for a range effect in the saccadic system. *Vision Research, 25,* 1155-1157.

Kapoula, Z., & Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research, 26,* 735-743. doi: 10.1016/0042-6989(86)90087-8

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262-284. doi:10.1080/09541440340000213

Krügel, A., & Engbert, R. (2014). A model of saccadic landing positions in reading under the influence of sensory noise. *Visual Cognition, 22,* 334-353. doi:10.1080/13506285.2014.894166

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition, 147*, 1-20. doi: 10.1016/j.cognition.2015.10.013

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition, 126*, 313-318. doi: 10.1016/j.cognition.2012.09.010

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words.Vision research, 28, 1107-1118.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. Perception & Psychophysics, 46, 245-253.

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition, 13*, 89-98. doi: 10.1080/13506280500143367

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research, 45,* 2201-2217. doi:10.1016/j.visres.2005.02.014

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance, 18.* 185-197. doi:10.1037/0096-1523.18.1.185

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. É. (1984). Convenient fixation location within isolated words of different length and structure. Journal of

Experimental Psychology: Human Perception and Performance, 10, 250-257. doi:10.1037/0096-1523.10.2.250

Paterson, K. B., Almabruk, A. A. A., McGowan, V. A., White, S. J., & Jordan, T. R. (2015). Effects of word length on eye movement control: The evidence from Arabic. *Psychonomic Bulletin & Review.* (Advance online publication). doi:10.3758/s13423-015-0809-4

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108,* 3526-3529. doi: 10.1073/pnas.1012551108

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.Rproject.org/

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8,* 21-30. doi: 10.1068/p080021

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

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22(5),* 1188-1200.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review, 33,* 110-149. doi: 10.1016/j.dr.2013.03.001

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: accounting for initial fixation locations and refixation within the E-Z Reader model. *Vision Research, 39,* 4403-4411. doi:10.1016/S0042-6989(99)00152-2

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445-476. doi: 10.1017/S0140525X03000104

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

Schotter, E R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131,* 1-27. doi: doi:10.1016/j.cognition.2013.11.018

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

Vitu, F., O'Regan, K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics in scanning letter strings and reading texts. *Perception & Psychophysics, 57,* 352-364. doi:10.3758/BF03213060

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krugel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition, 132,* 181-215. doi: 10.1016/j.cognition.2014.03.008

Yao-N'Dré, M., Castet, E., & Vitu, F. (2014). Inter-word eye behaviour during reading is not invariant to character size: Evidence against systematic saccadic range error in reading. *Visual Cognition, 22,* 415-440. doi: 10.1080/13506285.2014.886652

Footnotes

[1] We feel it is important to point out that we realize that the probability that a word will be a content word (as opposed to a function word) is fixed and does not change in relation to word length context. Instead, we are suggesting that a uniform word length constraint changes the likelihood that function or content words of that length will be present in those sentences.

[2] Some researchers may prefer to exclude fixations made on the space preceding a word when conducting a landing position analysis, since these fixations did not technically land on the word. While our approach is in line with the classic papers on landing positions, we also conducted all analyses involving landing position as a dependent or predictor variable on a restricted dataset in order to assuage any concerns arising from this point. In this restricted dataset we excluded words in which the first fixation was on the space preceding the word. The differences between the two analyses were minimal. Excluding trials in which the initial fixation was on the space preceding a word also allowed us to scale landing position relative to word length. This transformation also made minimal difference to the outcome of our analyses. We also acknowledge that typically the term ‘target word’ would be used to refer to specific words that had been manipulated within a sentence for certain characteristics. However, for the purpose of the current investigation we use this tem to refer to any individual word within our sentences.

[3] Many papers investigating the saccadic range error also include an analysis in which the launch site is calculated relative to the space before the target word. We did also construct a linear mixed effect model with this measure calculated from the start of the word, and calculated the preferred saccade length values from this model. Unsurprisingly, the same pattern of effects emerged.

Table 1

Mean Global Reading Measures for each Sentence Type. Standard Deviations are Presented in Parentheses.

	U3	U4	U5	NU
ASA (degrees)	1.68(0.40)	1.88(0.40)	2.13(0.50)	2.00(0.50)
ASA (letters)	4.87(1.16)	5.45(1.16)	6.18(1.45)	5.80(1.45)
PWF	0.77(0.12)	0.82(0.12)	0.89(0.10)	0.78(0.12)

Note. TPW= reading time per word; ASA= average saccade amplitude; PWF= the proportion of words fixated in a sentence. U3, U4, and U5 refer to uniform sentences of three, four, and five letter words respectively, and NU to non-uniform sentences.

Table 2

Linear Mixed Effects Models for Global Reading Measures.

	ASA			PWF		
Model	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	0.61	0.03	18.28	0.82	0.02	47.84
U3	**-0.12**	**0.02**	**-5.92**	**-0.04**	**0.01**	**-5.33**
U5	**0.12**	**0.02**	**6.82**	**0.07**	**0.01**	**6.69**
NU	**0.06**	**0.02**	**2.39**	**-0.04**	**0.01**	**-3.41**

Note. Significant factors are presented in bold. ASA= average saccade amplitude in degrees; PWF= the proportion of words fixated in a sentence. U3, U4, and U5 refer to uniform sentences of three, four, and five letter words respectively, and NU to non-uniform sentences.

Table 3

Linear Mixed Model Analyses for Fixation Landing Position Data.

	LP		
Model	*b*	*SE*	*t*
Intercept	0.01	0.23	0.05
SU	0.22	0.36	0.59
WL	**0.64**	**0.05**	**12.51**
SU*WL	-0.09	0.09	-1.04
LS	**-0.19**	**0.04**	**-5.11**
LS*WL	**-0.07**	**0.01**	**-10.81**
LS*SU	**0.17**	**0.05**	**3.15**
LS*SU *WL	**-0.03**	**0.01**	**-2.48**

Note. Significant terms are presented in bold. LP = landing position; SU= effect of moving from uniform to non-uniform sentences; WL= word length; LS= launch site.

Table 4

Fixation Landing Position and Skipping Probabilities as a Function of Word Length, Sentence Uniformity, and the Difference (D) Between each Word Length in Uniform and Non-uniform Sentences.

	Three letter words			Four letter words			Five letter words		
	U	UN	D	U	NU	D	U	NU	D
LP	-0.45	-0.40	-0.05	-0.41	-0.33	-0.08	-0.34	-0.54	0.20
CLP	0.27	0.25	0.02	0.23	0.23	0.00	0.23	0.22	0.01
SP	0.26(0.43)	0.40(0.49)	-0.14	0.20(0.40)	0.22(0.42)	-0.02	0.13(0.33)	0.11(0.32)	0.02

Note. U= uniform; NU= non-uniform; D= difference; LP= mean landing position relative to the word center; CLP= proportion of fixations landing in the word center; SP= skipping probability.

Table 5

Linear Mixed Model Analyses for Skipping Probabilities.

	SP		
Model	*b*	*SE*	*z*
Intercept	-1.5	0.25	-5.86
SU	**1.5**	**0.29**	**5.14**
WL	**-0.49**	**0.04**	**-12.68**
Frequency	**0.17**	**0.01**	**15.45**
SU*WL	**-0.33**	**0.07**	**-4.48**

Note. Significant terms are presented in bold. SP= skipping probability; SU= effect of moving from uniform to non-uniform sentences; WL= word length; LP = landing position.

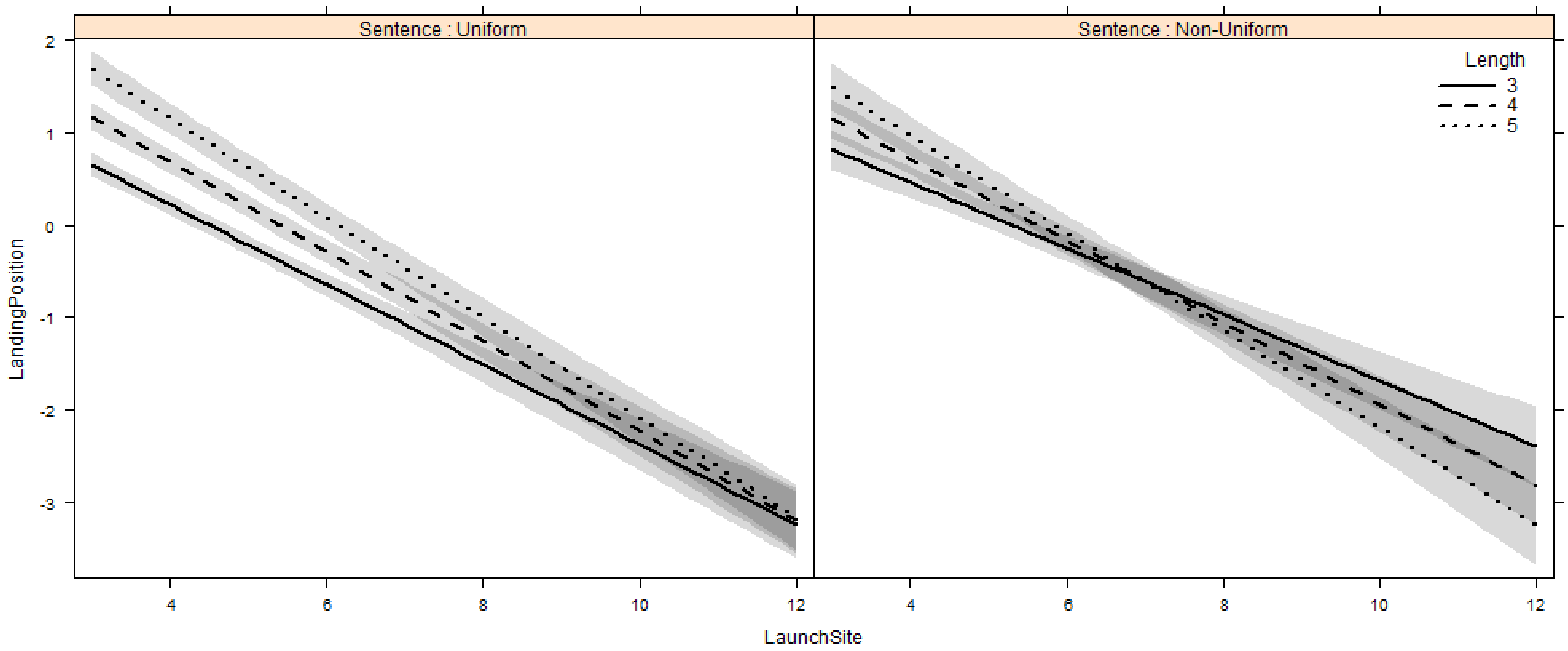

Figure 1. Fixed effects estimates from the linear mixed effect model for initial fixation landing position as a function of sentence uniformity, word length, and saccade launch site. Both landing position and launch site were calculated relative to the word center. 95% confidence bands are presented around the lines of estimated effects.

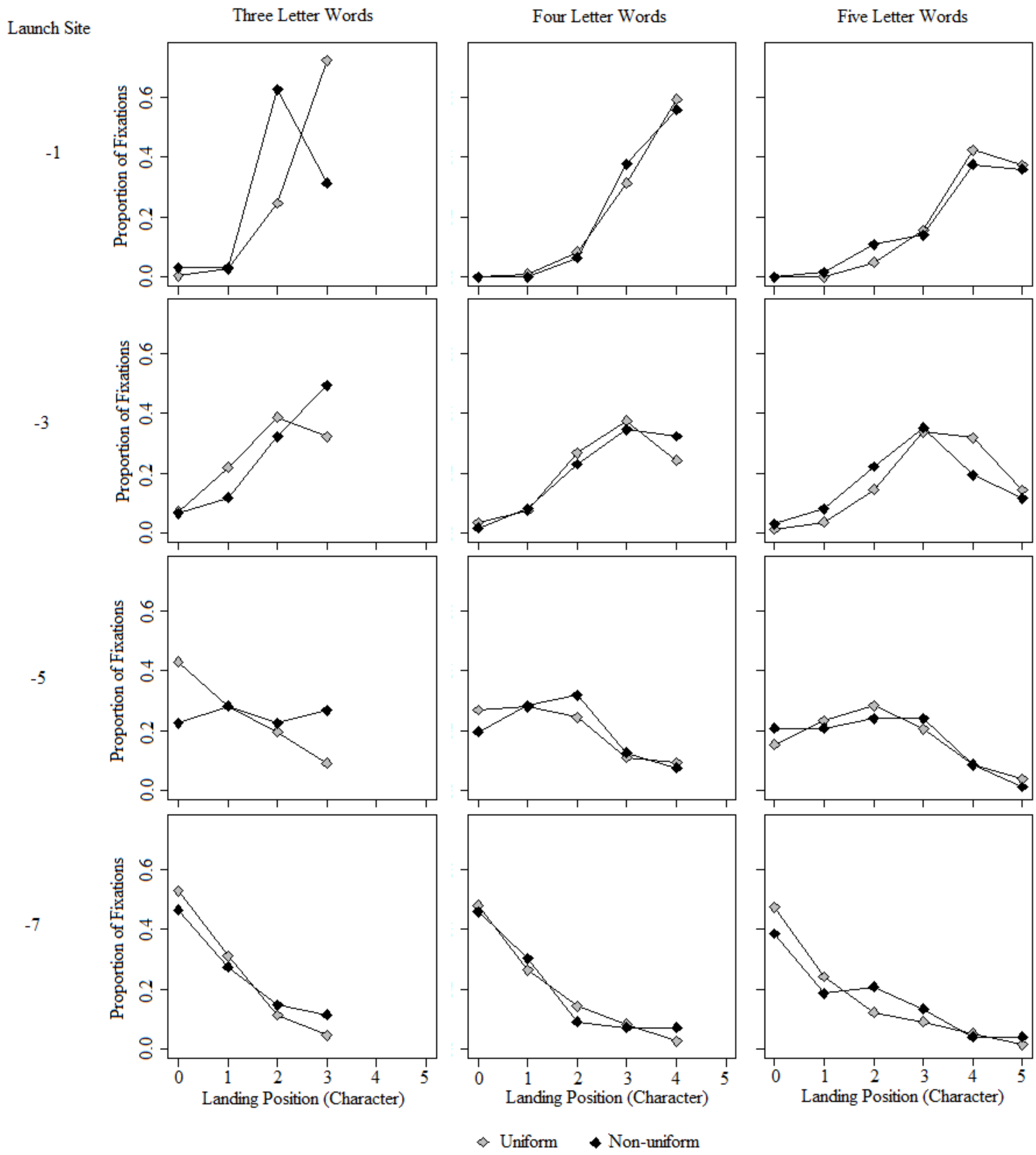

Figure 2. Fixation landing position distributions as a function of word length, sentence uniformity, and launch site. Both launch site and fixation landing position were calculated relative to the space on the left of the word. Launch sites between -1 to -7 were chosen in order for equivalence with McConkie et al. (1988).

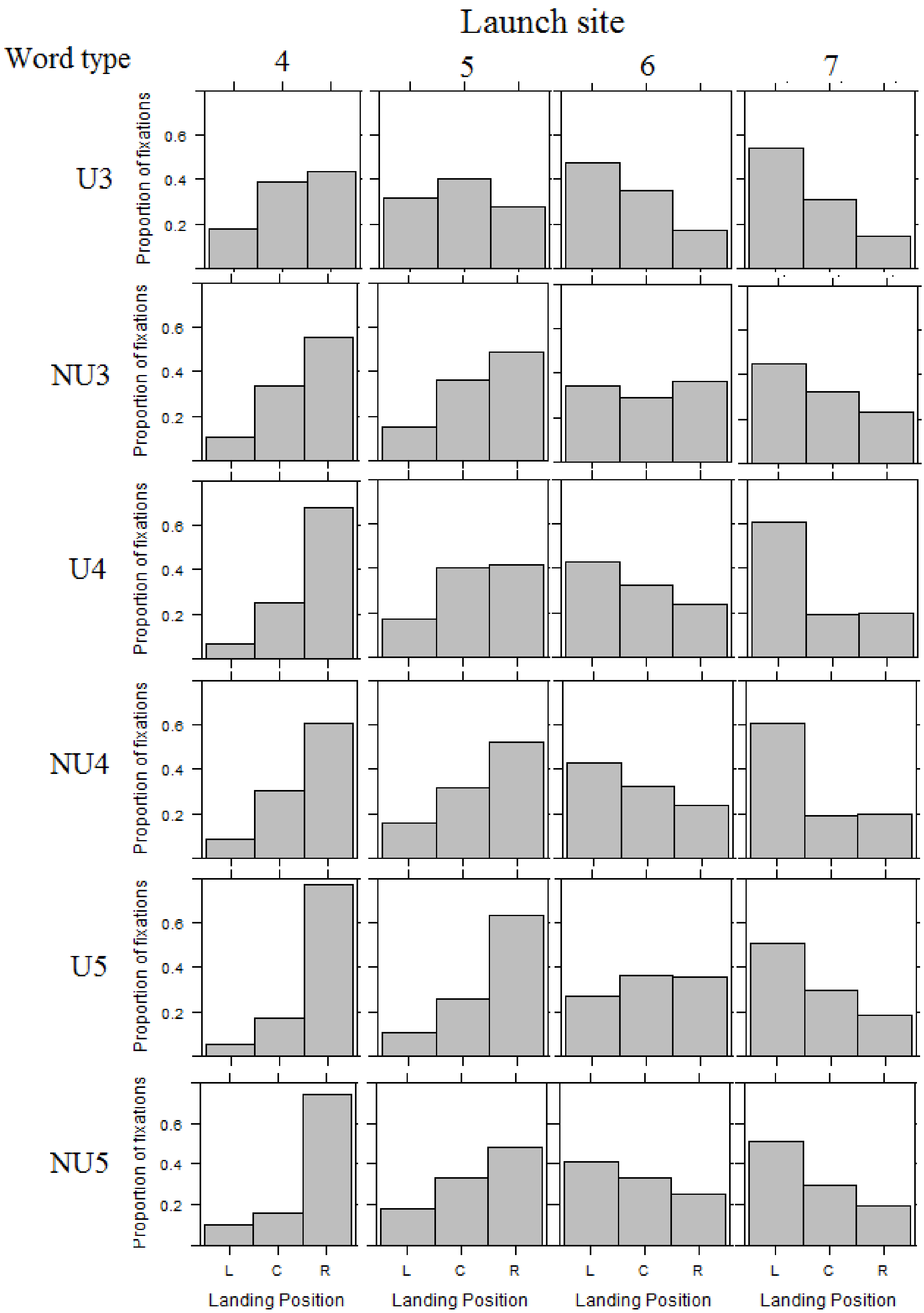

Figure 3. Fixation landing positions as a function of word length, sentence uniformity, and saccade launch site. Launch sites are relative to the central point of the fixated word. Launch sites of four to seven characters are presented due to our predictions of when readers would transition from over to undershooting the center of a word.

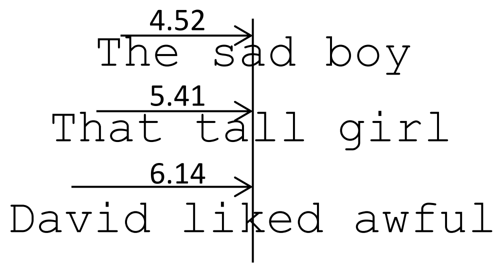

Figure 4. The preferred saccade lengths obtained from our linear mixed effects models mapped onto each uniform sentence type. The vertical black line bisects the center of a word in each sentence type, with the arrow above showing the origin of the preferred saccade length towards this location.

Appendix

The following Appendix contains reading time measures (which were not the primary focus of our main analyses). In order to examine the effect of our manipulation on reading times at a global level we examined the reading time per word across our different sentence types (see Table A1 for descriptive statistics and Table A2 for inferential statistics). Across the three uniform conditions reading times increased alongside word length, with the words within uniform sentences of three letter words and uniform sentences of five letter words being read significantly faster and slower, respectively, than the words within uniform sentences of four letter words, replicating standard word length effects. Furthermore, the reading time per word in the non-uniform sentences was significantly shorter than in the uniform sentences of four letter words, and numerically smaller than in the uniform sentences of three letter words. Given that we observed word length effects between the uniform conditions, and that the words in the non-uniform condition were on average the same length as in the uniform sentences of four letter words, this finding suggests a cost of sentence uniformity on reading times.

Table A1

Mean Reading Time per Word for each Sentence Type. Standard Deviations are Presented in Parentheses.

	U3	U4	U5	NU
TPW	250(73)	269(79)	329(100)	247(66)

Note. TPW= reading time per word. U3, U4, and U5 refer to uniform sentences of three, four, and five letter words respectively, and NU to non-uniform sentences.

Table A2

Linear Mixed Effects Model Output for Reading Time per Word.

	TPW		
Model	*b*	*SE*	*t*
Intercept	5.56	0.05	117.59
U3	**-0.07**	**0.03**	**-2.79**
U5	**0.20**	**0.03**	**6.62**
NU	**-0.08**	**0.03**	**-2.49**

Note. Significant factors are presented in bold. TPW= reading time per word. U3, U4, and U5 refer to uniform sentences of three, four, and five letter words respectively, and NU to non-uniform sentences.

As well as examining global reading times, we calculated several measures of fixation times on individual words within sentences, as a function of word length and sentence uniformity. The fixation time measures were first fixation duration (FFD; the duration of the first fixation on a word), gaze duration (GD; the sum of all fixations on a word from the first fixation until a saccade is made to another word), single fixation duration (SFD; the duration of a fixation when it is the only one made on a word) and total time (TT; the sum of all fixations on a word). These measures were log transformed in order to increase normality and were analysed using linear mixed effects models. Initially, these models included an interactive effect of sentence uniformity and word length as well as the main effects of these variables. However, in the models reported here the interaction term was removed, due to it failing to significantly improve the fit of any of our models. In addition, log word frequency was treated as a fixed factor in order to account for any differences between the average word frequencies in each sentence type. Participants and items were treated as random factors, and a full random structure was used, although random slopes were sometimes removed in cases when the full model failed to converge.

The means and standard deviations for these measures are displayed in Table A3, and the LME output in Table A4. There was a clear tendency across all measures for longer words to receive longer fixations. This effect was significant in both gaze durations and total times, and marginal in single fixation durations. Furthermore, words in the uniform sentences received longer fixations than those in the non-uniform sentences, with a significant effect of uniformity appearing in all four measures. The tendency for individual words within the uniform sentences to receive longer fixations than words of the same length in non-uniform sentences is in line with the finding that the uniform sentences were generally read more slowly than non-uniform sentences.

Table A3

Fixation Landing Position, Local Reading Time Measures, and Skipping Probabilities as a Function of Word Length, Sentence Uniformity, and the Difference (D) Between each Word Length in Uniform and Non-uniform Sentences.

	Three letter words			Four letter words			Five letter words		
	U	NU	D	U	NU	D	U	NU	D
FFD	224(72)	217(69)	7	229(75)	222(76)	7	234(74)	220(67)	14
GD	238(89)	227(82)	11	247(95)	238(93)	9	259(98)	244(93)	15
SFD	226(72)	219(71)	7	231(77)	222(75)	9	238(74)	222(68)	16
TT	275(129)	254(120)	21	282(134)	265(128)	17	320(161)	280(137)	40

Note. U= uniform; NU= non-uniform; D= difference; FFD = first fixation duration; GD= gaze duration; TT= total time; SFD= single fixation duration.

While these effects could be seen to suggest that there was some sort of inhibitory effect of within-sentence word length uniformity on lexical processing, we believe there is a simple explanation for these findings. As mentioned in the main body of the manuscript our uniform sentences were rated as being significantly less natural sounding than our non-uniform sentences. We maintain that this confound could not have made a significant contribution to our observed pattern of effects for skipping probabilities and saccadic targeting. However, in the case of our reading measures it seems plausible that reduced naturalness would have led to the trend towards increased reading times for all three word lengths in uniform relative to non-uniform sentences.

Table A4

Linear Mixed Model Analyses for Local Measures of Reading.

	FFD			GD			TT			SFD		
Model	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Intercept	5.35	0.05	114.88	5.37	0.05	114.97	5.37	0.07	76.80	5.34	0.05	113.51
SU	**-0.03**	**0.01**	**-2.94**	**-0.04**	**0.01**	**-3.39**	**-0.08**	**0.02**	**-4.05**	**0.02**	**0.01**	**3.10**
WL	**0.01**	**0.01**	**2.35**	**0.03**	**0.00**	**4.43**	**0.06**	**0.01**	**4.99**	**-0.04**	**0.01**	**-3.24**
Frequency	-0.00	0.00	-1.59	**-0.01**	**0.00**	**-3.56**	**-0.01**	**0.00**	**-2.00**	-0.00	0.00	-1.95

Note. Significant terms are presented in bold. FFD = first fixation duration; GD= gaze duration; TT= total time; SFD= single fixation duration; SU= effect of moving from uniform to non-uniform sentences; WL= word length.