Language and Linguistics Compass 9/8 (2015): 311–327, 10.1111/lnc3.12151

The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation

Adrian Staub*
Department of Psychological and Brain Sciences, University of Massachusetts

Abstract

A word's predictability in its context has a reliable influence on eye movements in reading. This article reviews the extensive literature that has investigated this influence, focusing on several specific empirical issues. These include assessment of cloze probability as the critical measure of predictability, the form of the relationship between predictability and reading time, the distributional effects of predictability, the interaction between predictability and word frequency, and the interaction between predictability and parafoveal preview. On the basis of this review, two theoretical conclusions are proposed. First, predictability effects in reading result from graded activation of potentially many words, as opposed to discrete prediction of a specific word. Second, this activation has the effect of facilitating either very early stages of lexical processing or pre-lexical processing of visual features or letters.

One of the most important influences on how a word is read is the word's *predictability*. Ehrlich and Rayner (1981) first demonstrated that when a word can be predicted from the preceding context, the eyes are relatively likely to skip over the word rather than directly fixate it, and that when the eyes do fixate a predictable word, they spend relatively little time on it. These effects have been confirmed in dozens of subsequent studies. Computational models of eye movements in reading such as E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) treat predictability as one of the two main linguistic variables affecting decisions of when and where to move the eyes; the other is context-independent word frequency. These effects have also provided empirical support for language processing models according to which processing difficulty is a function of a word's probability in its context (Hale, 2001; Levy, 2008). More generally, they have been a source of interest in recent psycholinguistic work emphasizing the role of predictive or anticipatory processes in language comprehension.

Understanding cloze in autism tells us about their lexical prediction abilities

Dozens of eye movement experiments have now investigated the details of predictability effects on eye movements in reading, the specific circumstances under which they arise, and their interaction with other experimental manipulations. The goal of this review is to clarify the state of our current knowledge and to explore its implications. Two theoretical questions are prominent. The first is whether predictability effects are due to genuinely predictive or anticipatory processes, and if so, what kind. One possibility is that a reader will sometimes maintain a specific, discrete prediction about what word is likely to come next. Another possibility is that in the course of language comprehension, words that are related to the current context are broadly activated; multiple words may be activated at once, and this activation may be graded rather than all-or-none. Finally, yet a third possibility is that facilitation for predictable words is not actually due to anticipatory processes, but rather to easier integration of predictable words into the syntactic, semantic, and discourse contexts in which they appear. The second question is about the stage or stages of processing that are influenced by predictability. Does a word's predictability facilitate lexical access itself? Or does it facilitate either a later, post-lexical stage, or an earlier, pre-lexical stage, such as the processing of a word's visual and orthographic features?

This review is organized as follows. The next section reviews the basic effects of predictability on eye movements in reading, which have been replicated many times. The following sections take up a range of specific empirical issues. First, we discuss whether predictability effects are fully captured by *cloze probability* (Taylor, 1953), the measure with which predictability is almost always identified. We then address the form of the relationship between cloze probability and reading time, and the effect of predictability on distributions of reading times. Two sections then review studies involving joint manipulations of predictability and other variables. The first investigates the relationship between predictability effects and effects of word frequency. The second investigates the interaction between predictability and manipulations of the visibility of a word before it is directly fixated, i.e., *parafoveal preview* manipulations. Finally, we return to the theoretical questions raised above.

Two preliminary notes are in order. First, this review focuses mainly on traditional experiments in which researchers address specific hypotheses by having subjects read individual sentences or short passages in which variables of interest have been manipulated and other variables have been carefully controlled. A number of studies have also investigated predictability effects in the context of continuous reading of more natural texts (e.g., Kliegl, Nuthmann, & Engbert, 2006; Kennedy, Pynte, Murray, & Paul, 2013; Whitford & Titone, 2014), where the effect of any one variable is assessed by means of a regression model that accounts statistically for the influence of other variables. These studies are not the present focus, in part due to space limitations, and in part because their conclusions are mostly consistent with the experimentally controlled studies. A few exceptions will be mentioned, however.[1] Second, it is very well known that predictability also has clear effects in another important paradigm for studying on-line language comprehension, the event-related potential (ERP) method. There are several recent reviews of predictability effects in ERPs (e.g., DeLong, Troyer, & Kutas, 2014; Van Petten & Luka, 2012), and this material will not be covered here.

1. *The Basic Phenomena*

In almost all experiments investigating the influence of predictability, this construct is operationalized in terms of a word's cloze probability (Taylor, 1953; see Staub, Grant, Astheimer, & Cohen, 2015, for discussion of the cloze task itself). A group of subjects is asked to guess the next word of the sentence, given the first several words. A word's cloze probability is simply the proportion of subjects who provide that word. In the subsequent eye movement experiment, predictability is manipulated in one of two ways: the target word is held constant and the preceding context is varied so that the target has high or low cloze probability, or the context is held constant and two different target words are used, one with high and one with low cloze probability in that context.

To illustrate the basic effects of predictability on eye movements, we describe a relatively recent study in detail. Rayner, Slattery, Drieghe, and Liversedge (2011) manipulated the predictability of a target word along with its length. There were three levels of word length: short (4–6 letters), medium (7–9 letters), and long (10–12 letters). In this study, the target word was held constant, while the preceding context was manipulated by means of a separate, preceding sentence, as in (1); in this and subsequent examples the target word is in italics.

(1) a Britney always drives way too recklessly. She just got into an *accident* that wrecked her car.

b Britney is in a really bad mood today. She just got into an *accident* that wrecked her car.

Based on a cloze study with 93 subjects, the predictability of the target words used in the eye movement study averaged over .7 in the predictable condition (1a), and around .15 in the unpredictable condition (1b). These values did not significantly interact with the word length manipulation. The predictability of the target in the unpredictable condition is higher in this study than is typical; the mean in the low-predictability condition is often near .01, and is usually below .05.

Twenty-eight subjects participated in the eye movement experiment, with each subject reading nine sentences in each of the six conditions (i.e., 2 (predictability) × 3 (word length)). As is usually the case (e.g., Brysbaert & Vitu, 1998), word length strongly influenced the probability that the target word was skipped rather than directly fixated on first pass reading. However, a highly predictable target was also more likely than a less predictable target to be skipped, whether the target was short (36% vs. 28%, respectively), medium (22% vs. 18%), or long (18% vs. 10%). Highly predictable targets also elicited shorter reading times on the majority of trials when the word was directly fixated. On the first fixation duration measure (the duration of the very first eye fixation on the word), this facilitation was present for short words (means of 205 vs. 209 ms), medium words (193 vs. 204 ms), and long words (195 vs. 211 ms). This was also the case for gaze duration (the sum of all fixation durations on the word before the eyes leave the word for the first time; on many trials, this measure is identical to first fixation duration), for short words (211 vs. 222 ms), medium words (213 vs. 231 ms), and long words (222 vs. 246 ms). Effects of the interaction between length and predictability did not reach statistical significance.

These effects of predictability on skipping rate and the early reading time measures of first fixation duration and gaze duration are usually in evidence. There are, of course, exceptions. Sometimes, skipping rates in all conditions are very low (i.e., less than 10%) which makes effects on skipping difficult to detect statistically. There are also experiments in which the effect of predictability on one of the early reading time measures fails to reach significance, which is unsurprising given the number of studies and the expected Type II error rate. But across the literature as a whole, what stands out is the consistency of these effects, despite variation in how predictability is manipulated, in the strength of the manipulation and in experimental procedures.

Relatively late stages of processing are thought to be reflected in measures such as the probability of making a leftward saccade from the word, or *regression*, and a reading time measure that includes time spent re-reading earlier material after such a regression, known as *go-past time* (see, e.g., Clifton, Staub, & Rayner, 2007). Sentence level manipulations, such as manipulations of syntactic ambiguity, frequently influence these measures (e.g., Clifton et al., 2007; Staub, 2011b). An effect on the reading time of subsequent material, or a *spillover* effect, is also often attributed to relatively late processes of integration of a word into its context and often appears in experiments that manipulate factors such as a word's plausibility (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004b), which we briefly discuss below. Does predictability influence these 'late' measures? Unfortunately, reporting of these measures has been extremely inconsistent in the predictability literature. For example, the Rayner et al. (2011) paper just discussed but did not report the rate of regressions, go-past times, or spillover reading times. There are individual studies, however, that do report these measures. For example, Staub (2011a) obtained a significant effect of predictability on the rate of regressions. On the other hand, Balota, Pollatsek, and Rayner (1985) and White, Rayner, and Liversedge (2005) both found that predictability did not influence spillover fixation durations with normal presentation of text (see below for further discussion of the manipulations in these experiments). The interpretation of any spillover effects is also complicated by the fact that material downstream of the critical word is often not well controlled. Spillover effects of predictability have also been investigated in corpus studies, with contradictory results (Kliegl et al., 2006; cf. Kennedy, Pynte, Paul, & Murray, 2013). In sum, given the inconsistency of reporting, and given that most experiments have not been designed to investigate these effects, any general conclusion about predictability effects on these 'late' measures would be unwarranted.

Finally, it is notable that predictability does not influence where the eyes initially land in a word (Lavigne, Vitu, & d'Ydewalle, 2000; Rayner, Binder, Ashby, & Pollatsek, 2001). This null finding is consistent with a large body of research indicating that while skipping probability is influenced by linguistic factors in addition to word length, the eyes' landing position in a word, when it is fixated, is determined almost entirely by word length.

2. Cloze Probability and Other Measures of Predictability

Cloze probability is a subjective measure of a word's predictability, based on aggregating responses from many subjects. A corpus of text can also be used to compute an objective measure of the probability of a word's occurrence, given the preceding word or words. McDonald and Shillcock (2003) investigated the effect of the forward transitional probability (TP) between two words – the probability that the second word will follow the first – on reading time on the second word. They focused on verb–noun sequences in which TP in a large corpus of text was either clearly non-zero (approximately .01; e.g., *accept defeat*) or was very close to zero (approximately .0004; e.g., *accept losses*). Their critical finding was that first fixation duration was significantly shorter on the noun in higher TP sequences, though neither the effect on gaze duration nor the effect on word skipping was significant.

However, a follow-up study by Frisson, Rayner, and Pickering (2005) demonstrated that apparent effects of transitional probability are likely to be cloze probability effects in disguise. Though cloze probability was generally low for the target nouns in the McDonald and Shillcock study, it was 10 times higher for the high TP nouns than for the low TP nouns (.08 vs. .008). Frisson et al. (2005) replicated the McDonald and Shillcock (2003) experiment, also introducing a contextual manipulation designed to make either the high TP or low TP noun more predictable. Frisson et al. found an effect of TP on gaze duration, though not on other measures. As usual, they found effects of cloze probability on both skipping and early reading time measures. Most importantly, they found that the effect of TP was restricted to items in which cloze probability differed substantially between the low TP and high TP nouns. In a second experiment with better-controlled materials, they found only effects of cloze probability, with no effects of TP.

This is not to say, however, that predictability effects cannot be due to the word immediately preceding the target. Fitzsimmons and Drieghe (2013) conducted an experiment with items like (2):

(2) a Bill has always been a fearful person. He screamed when he saw the hairy *spider* in the bath.

b Bill has always been a fearful person. He screamed when he saw the *spider* in the bath.

c Bill is scared of eight-legged creatures. He screamed when he saw the *spider* in the bath.

The addition of a single adjective (*hairy*) results in a dramatic increase in cloze probability in (a) compared to (b). The critical result was that this way of manipulating cloze probability had just the usual effects on eye movements: Word skipping was more likely in (a), and fixation durations were shorter. Moreover, eye movements on the critical word in condition (a) did not differ from condition (c), which had similar cloze probability to (a), but where the target word's predictability was due to a manipulation of the preceding sentence. It appears that what matters is a word's cloze probability itself, whether this value is attributable to a single preceding word or to the more remote context. Importantly, this experiment again casts doubt on a role for transitional probability. Though the probability of *spider* after *hairy* is relatively high (certainly much higher than after *the*, as in (b) and (c); other adjective–noun combinations included *sharp knife*, *muddy boots*, and *slippery floor*), this statistical regularity apparently did not have any effect on reading times over and above the effect of cloze probability.

A regression analysis of eye movement data from the Potsdam sentence corpus (Boston, Hale, Kliegl, Patil, & Vasishth, 2008) has reported a significant effect of TP (or a closely related measure, bigram frequency) when cloze probability is taken into account. These data are a record of eye movements of a large number of subjects reading 144 sentences of various types, where linguistic variables such as predictability and TP were not explicitly manipulated (Kliegl et al., 2006). Boston et al. (2008) found significant effects of both cloze probability and bigram frequency on a range of eye movement measures. Given the inherent difficulty involved in making causal inferences when there is a complex correlational structure among predictors, this result may be regarded as suggestive, but the lack of any reliable effects of TP in controlled experiments is arguably more compelling.[2]

Predictability effects may also be thought to involve semantic priming, which has been shown to exert an influence on fixation durations (Carroll & Slowiaczek, 1986; Morris, 1994; Traxler, Foss, Seely, Kaup, & Morris, 2000). In typical predictability experiments, the context preceding a high-cloze target may contain semantic associates of the target (e.g., *drives* and *recklessly* in 1a, and *eight-legged* in 2c). Perhaps simple priming is responsible for some or all of the apparent effect of cloze probability? This question has actually received little direct investigation, though one relevant study is by Calvo and Meseguer (2002). In this study, lexical associates of the target word were always present in the context, whether the target word had high or low cloze probability. Effects of cloze probability did not appear in the usual early measures (skipping probability, first fixation, and gaze duration) but only in later measures such as the probability of a regressive eye movement. It appears that equating the level of lexical priming from the context may reduce, if not eliminate, the usual predictability effects. Clearly, this issue should be explored in further experiments.

Finally, we consider the question of whether in addition to effects of a word's predictability, there are also effects of contextual constraint. Constraint is a property of a word's context, not the word itself. A high constraint context is one in which a single word is highly predictable, while a low constraint context is one that elicits many different kinds of continuation. Constraint may be measured simply by the cloze probability of the modal response or by an entropy-based measure; in practice, these two measures will tend to be correlated (Staub et al., 2015). In those experiments that hold the context constant and display either a high-predictability or low-predictability target word, the context will be constraining, simply because there is indeed a high cloze probability continuation. But in experiments in which the target is held constant and the context is varied, one context will tend to be highly constraining, while the other is not. Does constraint itself have an effect? If a word has low cloze probability, does it matter whether some other word has high cloze probability (i.e., the context is constraining) or if no other word is predictable?

explanation of high and low constraint

Recently, Frisson, Harvey, Drieghe, and Staub (2015) addressed this question. They manipulated the context's degree of constraint, as well as the predictability (in the high constraint context) of the target word, as in (3):

(3) a The doctor told Fred that his drinking would damage his *liver* very quickly.

b The doctor told Fred that his drinking would damage his *heart* very quickly.

c Yesterday Fred told his friend that they will look at his *liver* very thoroughly.

d Yesterday Fred told his friend that they will look at his *heart* very thoroughly.

As usual, there were skipping and reading time benefits when a word was predictable, compared to when the same word was unpredictable (*liver* in (a) vs. (c)). However, reading of an unpredictable word was not influenced by whether the context was constraining; there were no significant differences in skipping or fixation duration measures on *heart* when *liver* was a predictable alternative (b) compared to when no word was predictable (d). A second experiment replicated these results, and also found no effect on the usual early measures (though there was a small effect on the probability of a regression) of whether the low-cloze word was semantically related to the alternative high-cloze word (e.g., *heart/liver*) or unrelated (e.g., *skill/liver*).

In sum, cloze probability is the only predictability-related measure that has a clear influence on eye movements in reading. The weight of the evidence supports the conclusion that statistically based measures such as TP do not have independent effects, and that contextual constraint itself does not influence processing of an unpredictable word. These conclusions raise the question of what it is about cloze probability that makes it such a useful 'all-in' operationalization of predictability (see Staub et al., 2015, for discussion).

3. The Function Relating Predictability and Reading Time

The Frisson et al. (2005) study shows that even subtle predictability differences can give rise to significant effects in the eye movement record. One interpretation suggests that the small differences in predictability investigated by Frisson et al. have a particularly pronounced effect because they occur at the low end of the scale. Smith and Levy (2013) constructed a statistical model of gaze durations on all words in the English portion of the Dundee corpus (Kennedy et al., 2013). Using a trigram-based measure of predictability (i.e., an objective measure like TP), they found a logarithmic effect of predictability on gaze duration, across a range from predictability of 1 to .000001. A logarithmic effect means that a very small difference in predictability at the low end of the scale has an effect on reading time that is orders of magnitude larger than a predictability difference of the same absolute size at the higher end of the scale. Looked at the other way around, a logarithmic effect implies that a predictability difference of .01 vs. .001 has the same size reading time effect as a predictability difference of 1 vs. .1.

There is reason to doubt, however, that the relationship between predictability and reading time can be literally logarithmic. A logarithmic effect would imply that encountering a truly unlikely word, compared to a merely unpredictable one, should slow down reading very dramatically. This prediction has been tested in several experiments examining the processing of a word that is implausible in its context (e.g., Rayner et al., 2004a; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Warren & McConnell, 2007; see Matsuki et al., 2011, for discussion of the relationship between plausibility and predictability). Consider (4) and (5), used by Rayner et al. (2004a) and Warren and McConnell (2007), respectively:

(4) John used an axe to chop the large *carrots* for dinner last night.

(5) The man used a blow-dryer to dry the thin *spaghetti* yesterday evening.

Neither of these studies found increased first fixation or gaze duration on the target word, compared to when the same word was plausible in its context. Instead, they found effects only on later measures that take into account regressive re-reading or reading of subsequent material. Early effects of plausibility have been reported with more extreme semantic violations (e.g., in the 'anomalous' condition of Rayner et al., 2004a; Staub et al., 2007), but these effects are certainly not larger than the usual predictability effects. Thus, it is not the case that an extremely unpredictable word generates a reliable cost, let alone an exceptionally large one, on the early reading time measures that are affected by the usual predictability manipulations in the range from about .01 to 1. It is also worth noting that these studies have generally not found plausibility effects on word skipping.

Given the usual manipulations of predictability in the range from about .01 to near 1, is the effect concentrated at the higher or lower end of the cloze probability scale? Rayner and Well (1996) first addressed these questions systematically. In this study, the pre-target context was held constant; in a third of the items, the target word was high (ranging from .73 to 1) or low (.03–.08) in cloze probability, in a third it was medium (.13–.68) or low in cloze probability, and in a third the target was high or medium cloze.[3] Rayner and Well found that reading times for high-cloze and medium-cloze words did not differ significantly from each other, while reading times for low cloze words were longer. On the other hand, the skipping rate was lower for high-cloze continuations than for the other classes, which did not differ from each other.

Follow-up studies have muddied this picture. In a French study, Lavigne et al. (2000) again found that on reading time measures, high-predictability and medium-predictability words differed from low-predictability words, but not from each other. They did not find skipping effects, due to a very low rate of skipping in general. Rayner, Li, Juhasz, and Yan (2005) presented Chinese readers with translations of the Rayner and Well (1996) items, and again found the same pattern in reading times, but found that low-predictability targets also differed from both high and medium on the skipping measure, in contrast to Rayner and Well (1996). Most notably, however, Rayner, Reichle, Stroud, and Williams (2006), again using the Rayner and Well (1996) materials, found high-predictability words behaving differently from medium and low on all measures, with no statistically significant differences between the latter two classes. It is important to note, however, that across this literature, the means for medium-predictability targets fall between the high and low targets, on essentially all measures. These studies had limited power to detect a three-way distinction and given the consistency in the numerical pattern, the failures to obtain significant differences between adjacent categories may reflect Type II errors. It appears that predictability has a graded effect on both skipping and reading time.

4. *Distributional Effects of Predictability*

Predictability influences the mean time that readers spend on a word. By itself, this fact does not tell us much about what aspect or stage of processing is affected by a word's predictability. But examination of the effect on the *distribution* of fixation durations may help to answer this question. As is the case with most temporal measures, distributions of fixation durations for individual subjects tend to be unimodal, but with some degree of right skew. Consider two alternatives. One possibility is that predictability reduces both short and long fixations by about the same amount, resulting in an overall shift to the left of the distribution of fixation durations. This would tell us that predictability is influencing some aspect of processing that occurs on most or all trials. Another possibility, however, is that predictability has most or all of its effect on trials when processing would otherwise be difficult, resulting in a less pronounced right tail of the distribution.

One way to determine which of these patterns actually obtains is to fit to the data a distributional model such as the ex-Gaussian (Balota & Yap, 2011). The ex-Gaussian is a convolution of a Normal distribution (with parameters μ and σ) and an exponential distribution (with parameter τ). An effect on the location of a distribution will appear as an effect on μ, while an effect on the weight of the right tail of the distribution will appear as an effect on τ. Several studies have fit the ex-Gaussian to distributions of fixation durations in reading, which requires many more observations per subject than are usually collected in eye movement studies (e.g., Reingold, Reichle, Glaholt, & Sheridan, 2012; Staub, White, Drieghe, Hollway, & Rayner, 2010; White & Staub, 2012). Staub et al. (2010) and Reingold et al. (2012) found that a word frequency manipulation influences both the μ and τ parameters, not only shifting the distribution but also showing a particularly pronounced effect on slow trials. Reingold et al. (2012) also found that the presence of valid parafoveal preview (discussed further below) influences both parameters. White and Staub (2012), on the other hand, found that a manipulation of stimulus quality (i.e., visual contrast) influences only the location of the fixation distribution, having no effect on distributional shape.

Three experiments have focused on predictability. In Staub (2011a), each subject reads 50 target words twice, once when it was highly predictable, once when it was not. Staub and Benatar (2013) reported a replication of Staub (2011a) with both more items (85 target words read in each type of context) and more subjects. In Sheridan and Reingold (2012), there were 40 target words in each type of context. The conclusion of all three of these studies is the same: The distribution of fixation durations for high-predictability words is simply shifted to the left. The effect of predictability on the μ parameter is significant and accounts for essentially all of the effect on the mean; the effect on τ was not significant in any of these studies, as shown in Table 1. This pattern is also evident in visualizations of distributional shape. Figure 1 shows the .1, .3, .5, .7, and .9 quantiles of the first fixation duration distribution for high-predictability and low-predictability words in the Staub and Benatar (2013) data. It is evident that the effect of predictability is approximately constant across the entire distribution.

These results argue that predictability is influencing an aspect of processing that occurs on most or all trials. Predictability does not have any particular effect when processing is slow, as it reduces the duration of short fixations about as much as long ones. Thus, predictability has a distributional effect that is similar to that of stimulus quality but different from that of word frequency and parafoveal preview. Below, we return to the implications of these results.

5. Predictability and Word Frequency

As noted above, context-independent word frequency also has a reliable influence on both skipping and early reading time measures (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Staub et al., 2010). Researchers have been keenly interested in the relationship between the frequency and predictability effects. A common intuition is that the cost of processing a low-frequency word should be largely or entirely eliminated when that word is made very

Table 1. Predictability effect on mean first fixation duration, in ms, and effects on ex-Gaussian parameters.

	Mean	μ	σ	τ
Staub (2011)	15.5 ($p < .001$)	14.1 ($p < .001$)	4.6 (NS)	3.0 (NS)
Staub and Benatar (2013)	15.1 ($p < .001$)	15.6 ($p < .001$)	1.1 (NS)	−.04 (NS)
Sheridan and Reingold (2012)	8 ($p < .01$)	8 ms ($p < .05$)	−1 (NS)	1 (NS)

NS, not significant.

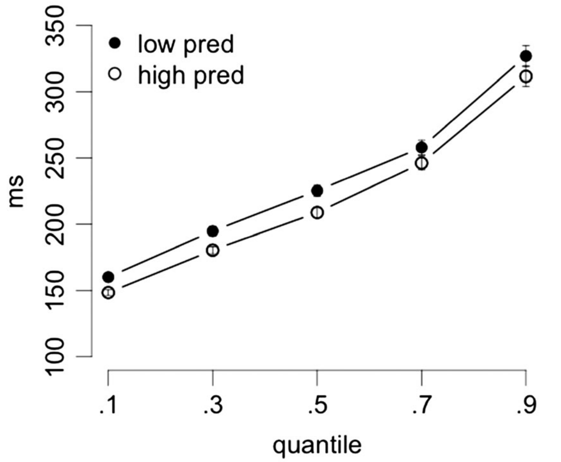

Fig. 1. Mean quantiles of individual subjects' first fixation duration distribution and standard error, for high-predictability and low-predictability words in data reported by Staub and Benatar (2013).

predictable; equivalently, the predictability effect should be larger for low-frequency than high-frequency words. This intuition is formalized in probabilistic models such as the Bayesian Reader (Norris, 2006) and Surprisal (Levy, 2008), which propose that the frequency effect is actually a form of predictability effect. If the context provides no information as to what specific word is coming up, a given high-frequency word will be much more expected than a given low-frequency word. But if low-frequency and high-frequency words are equally high in predictability, the frequency effect may disappear.

A large number of studies, summarized in Tables 2 and 3, have now factorially manipulated the frequency and predictability of target words in order to test for interaction effects. These include studies with special populations (deaf readers, Bélanger & Rayner, 2013; bilingual readers, Altarriba, Kroll, Sholl, & Rayner, 1996; Gollan et al., 2011; highly skilled readers, Ashby, Rayner, & Clifton, 2005) and two studies in French (Lavigne et al., 2000; Miellet, Sparrow, & Sereno, 2007). The conclusions are remarkably consistent regarding reading time measures: There is no interaction. Indeed, the numerical trend is not consistently in the direction predicted by the account sketched above, as approximately as many studies show numerically larger predictability effects for high-frequency words as for low-frequency words. Figure 2 illustrates the almost perfectly additive effects of predictability and frequency on gaze duration in the Staub and Benatar (2013) data. This experiment had an unusually high degree of power to detect an interaction, as each subject reads 85 target words twice, in contexts that rendered the word either highly predictable or not, with word frequency varying from very low (e.g., *synonym*) to very high (e.g., *baby*).

The situation for word skipping is somewhat less clear. Three significant interactions have been reported, but critically, the direction of the effect is not consistent. One study (Gollan et al., 2011) found a larger predictability effect for low-frequency words. But two studies (Rayner, Ashby, Pollatsek, & Reichle, 2004a; Hand, Miellet, O'Donnell, & Sereno, 2010) found significantly larger predictability effects for high-frequency words, i.e., the opposite interaction from the one expected based on a probabilistic account of the frequency effect. Given these conflicting results together with the number of null effects, there is no justification for concluding that there is an interaction effect on word skipping in either direction.

Findings of invariance are critically important in building psychological theories (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009). The finding that the

Table 2. Frequency and predictability effects on gaze duration, in ms, and the interaction trend, in studies factorially manipulating the two variables.

	Freq	Pred	Interaction
Altarriba et al. (1996)[a]	35	18	NS+
Lavigne et al. (2000, E1)	34	24	NR
Rayner et al. (2001, E2)	17	15	NS
Rayner et al. (2004a)	17	19	NS+
Ashby et al. (2005, E2, skilled readers)[b]	25	23	NS+
Ashby et al. (2005, E2, average readers)	5 (NS)	24	NS−
Miellet et al. (2007)[c]	84	30 (NS)	NS−
Hand et al. (2010)	33	12	NS−
Gollan et al. (2011)[d]	10	12	NS−
Belanger & Rayner (2013)[d]	30	20	NS+
Kretzschmar, Schlesewsky, and Staub (2015)	29	33	NS−

NR, not reported; NS, not significant, direction of effect not reported; NS+, not significant, numerically larger predictability effect for low-frequency words; NS−, not significant, numerically larger predictability effect for high-frequency words.
All frequency and predictability main effects are significant at the .05 level except as noted.
Notes:
[a]This paper included a condition with Spanish target words in English sentences, not reported here.
[b]Groups of high-skill and average-skill readers are reported separately.
[c]Corpus study including 20 critical target words in which the critical variables were orthogonally manipulated.
[d]Each of these studies had three separate subject groups that are combined here, as there were no relevant interactions.

Table 3. Frequency and predictability effects on skipping proportion, and the interaction effect, in those studies reported in Table 2 that also reported word skipping.

	Freq	Pred	Interaction
Altarriba et al. (1996)	.03 (NS)	.05	NR
Rayner et al. (2001, E2)	NS	.10	NS; numerical trend not reported.
Rayner et al. (2004a)	.04 (NS)	.04 (NS)	$p < .05$; predictability effect for HF, not for LF
Miellet et al. (2007)	.01 (NS)	0 (NS)	NS; predictability effect in expected direction only for HF, but very low skipping rate overall.
Hand et al. (2010)	.03	.01 (NS)	$p < .01$ by subjects, $p = .064$ by items; predictability effect for HF, not for LF
Gollan et al. (2011)	.02 (NS by items)	.03	$p < .05$; larger predictability effect for LF
Belanger & Rayner (2013)	.04	.06	NS; near perfect additivity
Kretzschmar et al. (2015)	.05	.03	NS; numerical trend towards larger pred effect for HF

NR, not reported; NS, not significant; HF, high frequency; LF, low frequency.

frequency effect on eye movements is not modulated by a predictive sentence context should inform our understanding of both frequency effects and predictability effects. The

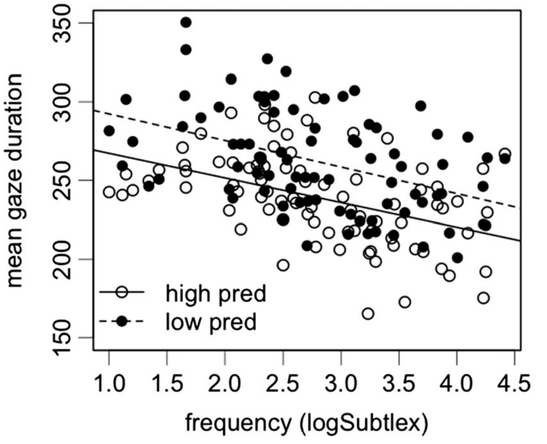

Fig. 2. Plot of by-item mean gaze duration in data reported by Staub and Benatar (2013); each target word is represented twice, with filled circles representing low-predictability means and open circles representing high-predictability means. Dashed and solid lines represent lines of best fit to the low-predictability and high-predictability means, respectively. Word frequency is based on the Subtlex corpus (Brysbaert & New, 2009).

contextual invariance of frequency effects supports the assumption made by traditional word recognition models (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 2004) that word frequency influences lexical access for reasons other than the difference in the probability of encountering a high-frequency vs. low-frequency word; in most models, the effect is attributed to differences in memory strength. On the other hand, the additive pattern suggests that predictability may *not* operate during lexical access itself. While it is widely acknowledged that additive effects do not prove that two manipulations are influencing distinct processing stages (e.g., McClelland, 1979), such a pattern is often best accounted for by a serial, staged model (Roberts & Sternberg, 1993). The next section provides suggestive evidence about the specific processing stage at which predictability may operate.

6. Predictability and Parafoveal Preview

A number of studies, beginning with Balota et al. (1985), have investigated the interaction of predictability and parafoveal preview. These experiments have used the *boundary paradigm* (Rayner, 1975) to manipulate the material that is visible in the location of a target word before that word is directly fixated. The stimulus presentation software changes the text that is displayed in this location during the reader's saccade into the target word. Though readers are usually not consciously aware of this change, incorrect parafoveal preview reliably increases reading time on the target word (see Schotter, Angele, & Rayner, 2012, for a review).

Balota et al. (1985) manipulated a target word's predictability together with the validity of parafoveal preview. In this experiment, the context was held constant and the target was a high-cloze or low-cloze word. There were five parafoveal preview conditions: identity (the target word itself, e.g., *cake*), an orthographically similar non-word (*cahc*), the other possible target word (*pies*), a non-word that was orthographically similar to this other target (*picz*), or an unrelated word (*bomb*). The first major finding was that predictability of the target increased skipping only when the reader had either valid preview (i.e., the identity preview condition) or preview of a visually similar non-word. This suggests that predictability-based skipping requires that parafoveal visual input roughly match the form of the predictable word. More surprising was a second finding: The predictability benefit on first fixation and gaze duration was also present only if the parafoveal preview was at least visually similar to the target. Put differently, lack of valid parafoveal preview lengthened reading times on a predictable word much more than on an unpredictable word; preview of an unrelated word increased gaze duration on a predictable word from 232 to 292 ms but increased gaze duration on an unpredictable word from 264 to 290 ms.

These patterns have been replicated in subsequent experiments. White et al. (2005) used an invalid preview condition in which the letter *s* was inserted between the target word and the subsequent word, so that parafoveal word length information was incorrect. Again, they found that predictability effects on both skipping and fixation duration measures were eliminated with invalid preview. Juhasz, White, Liversedge, and Rayner (2008, Experiment 3) varied the context while keeping the target constant. The target was either a compound word (e.g., *backhand*) or a two-word sequence generated by deleting one letter from the compound (*back and*), and was either predictable or unpredictable in the context. The preview was either correct or was the alternate target (e.g., *back and* as preview for *backhand*). While skipping effects were not reliable in this experiment, the neutralization of the predictability effect on first fixation duration by invalid preview was observed once again. Gaze duration did pattern differently here, showing additive effects of predictability and preview validity. Finally, Schotter, Lee, Reiderman, and Rayner (2015, Experiment 2) again found that predictability effects on fixation durations were eliminated when readers had parafoveal preview of an unrelated word; they did find a predictability effect when the preview was of a closely semantically related word.

Though there are a few inconsistencies across these experiments, a relatively clear picture emerges. The processing facilitation for a predictable word that is seen in both skipping and reading time measures depends on the reader obtaining valid parafoveal preview of the word. While the skipping result is arguably not surprising, the reading time result is less expected. This result suggests that the effects of predictability may result from facilitation of the processing that takes place while a word is still being viewed parafoveally, before it is directly fixated. Interestingly, there is a suggestion in some of the experiments mentioned above (e.g., White et al., 2005) that in conditions with invalid parafoveal preview, the predictability effect is pushed downstream, emerging on reading time measures on the following word. Also interesting is that while the word frequency effect on fixation durations is similarly attenuated when parafoveal preview is absent (e.g., Inhoff & Rayner, 1986; Kennison & Clifton, 1995; Reingold et al., 2015), some frequency effect on reading times appears to survive. While no experiment has directly assessed this apparent discrepancy, the frequency effect may be somewhat less dependent on parafoveal preview than is the predictability effect.

7. Towards a Theoretical Account of Predictability Effects in Reading

The first of the two theoretical questions that framed this review was whether predictability effects are due to genuinely predictive or anticipatory processing, and if so, what kind. We suggested three alternatives: readers may maintain specific, discrete predictions of upcoming words; sentence contexts may broadly activate words that are likely to appear in the input; or predictable words may simply be more easily integrated into developing syntactic, semantic, and discourse representations.

Several findings argue against discrete, all-or-none prediction of specific words. One is the finding that predictability effects appear to be present even at the low end of the cloze probability range; it seems unlikely that readers would specifically predict a word that has relatively low cloze probability. Another is the lack of any effect of contextual constraint in the recent experiments by Frisson et al. (2015). If a constraining context causes readers to predict a specific word, encountering a different word in its place would be expected to cause some disruption, which is not in evidence. There is also a clear argument against the view that there is no prediction or pre-activation at all, but only easier integration for predictable words. Predictability influences very early eye movement measures, including the earliest possible measure, word skipping. This influence is at least as early as the influence of factors that are thought to be either pre-lexical or lexical in their locus, such as stimulus quality and word frequency. The distributional findings also argue against this view, as predictability does not specifically affect the long fixations that are plausibly equated with integration difficulty. While the existing data do not show that predictability has no influence on integration difficulty (as noted above, the literature has not consistently assessed effects on the relevant 'late' eye movement measures), the reliable effects on early measures make clear that the effect of predictability is unlikely to be primarily, or even largely, an integration effect.

The view that is left standing, then, is that predictability effects are due to diffuse pre-activation of likely upcoming words, with this activation varying as a function of a word's cloze probability. This view accounts for the fact that predictability effects arise early in processing, and that they are present across the entire range of cloze probabilities. This view also can accommodate the distributional findings and the finding that an unpredictable word is not inhibited when there is a predictable alternative.

The second of our theoretical questions focused on the processing stage at which predictability operates. Does predictability influence lexical processing itself, or does it influence a pre-lexical or post-lexical stage? Perhaps it is already clear that a primarily post-lexical locus is not a viable option. In deciding between a lexical and pre-lexical locus, one relevant finding is that in distributional terms, the effect of predictability is similar to the effect of stimulus quality (i.e., visual contrast) and qualitatively distinct from the effect of word frequency. Another relevant finding is that frequency and predictability effects are consistently additive, not interactive. Yet a third is that lack of valid parafoveal preview eliminates the usual predictability effects in most studies. All of these findings point towards a locus either very early in lexical processing or at a pre-lexical stage. Pre-activation of a potential upcoming word may facilitate the extraction of the visual features and letters in that word, and identification of a word's orthographic form, rather than facilitating later stages of lexical access such as identification of a word's meaning. The E-Z Reader model posits two distinct stages of lexical processing, L1 and L2, and proposes that while word frequency and many other factors influence the duration of both stages, stimulus quality influences only the duration of L1 (Reingold & Rayner, 2006); within this framework, an intriguing possibility is that predictability also influences only L1.

8. Conclusions

Putting these tentative theoretical conclusions together, the overall picture that emerges is as follows. Reading a sentence results in activation of potential upcoming words. This does not seem to involve prediction of a specific word, but rather graded activation of words approximately in proportion to their cloze probability. Processing a word that is pre-activated in this way is then easier than it otherwise would be. The benefit of this pre-activation seems to accrue primarily in the very earliest stages of lexical processing or even during pre-lexical (i.e., feature and letter) processing.

This review has raised several theoretically important empirical questions that are as yet unanswered. First, how much of the predictability effect is attributable to semantic priming? Second, what function relating predictability to reading time would emerge from controlled experiments? Third, does invalid parafoveal preview diminish the effect of predictability more than it does other effects, such as the effect of word frequency? Fourth, does predictability have reliable effects on late processing measures, such as the probability of a regression or spillover reading time?

In concluding, we repeat an interpretive point made recently by Staub et al. (2015). If cloze probability is taken to *define* the notion of predictability, then the effects that we have discussed throughout this paper are unambiguously effects of predictability. But if a word's cloze probability is seen as a measure of its activation by context – and Staub et al. (2015) argue that it should be – it is no longer obvious that the effects we have been discussing are best viewed as predictability effects. Rather, notions of predictability and prediction may drop out of the explanatory picture, and the empirical effects of cloze probability may be viewed as directly reflecting pre-activation by context.

Notes

* Correspondence address: Adrian Staub, Department of Psychological and Brain Sciences, University of Massachusetts, 430 Tobin Hall, 135 Hicks Way, Amherst, MA 01003, USA. E-mail: astaub@psych.umass.edu

[1] It is a matter of debate whether statistical control in corpus studies is a substitute for traditional experimental control or even has advantages over traditional controlled experiments (Kliegl, Grabner, Rolfs, & Engbert, 2004; cf. Angele et al., 2015). In the interest of full disclosure, the present author's bias is to regard results from corpus studies as a useful source of hypotheses that then require testing in controlled experimental settings.

[2] Boston et al. (2008) also obtained significant effects of a word's Surprisal (Hale, 2001; Levy, 2008) on many eye movement measures. Like TP, Surprisal is a measure of conditional probability, but where the word's probability is conditioned on a more complex representation of the preceding sentence context. However, Boston et al. used a 'non-lexicalized' version of Surprisal, essentially measuring the conditional probability of a word's part of speech. Demberg and Keller (2008) also found significant effects of non-lexicalized Surprisal in the Dundee corpus, a record of 10 participants' eye movements while reading extended passages of newspaper text. However, they did not find any effects of a lexicalized version of the Surprisal metric, which computes the conditional probability of each specific word.

[3] Cloze probabilities in this study were based on a modified version of the task, in which subjects provided up to three continuations for every fragment. As a result, the combined cloze probabilities for two words in a given context could sum to more than 1; e.g., *oven* had cloze .98 and *stove* had cloze .23 after the fragment *The woman took the warm cake out of the*____.

Works Cited

Altarriba, J., J. F. Kroll, A. Sholl, and K. Rayner. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Memory & Cognition* 24. 477–492.

Angele, B., E. R. Schotter, T. J. Slattery, T. L. Tenenbaum, K. Bicknell, and K. Rayner. 2015. Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *Journal of Memory and Language* 79. 76–96.

Ashby, J., K. Rayner, and C. Clifton Jr. 2005. Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology* 58A. 1065–1086.

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

Balota, D. A., and M. J. Yap. 2011. Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science*, 20(3). 160–166.

Bélanger, N. N., and K. Rayner. 2013. Effects of frequency and predictability in skilled and less skilled deaf readers. *Visual Cognition* 21. 477–497.

Boston, M. F., J. Hale, R. Kliegl, U. Patil, and S. Vasishth. 2008. Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2. 1–19.

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

Brysbaert, M., and F. Vitu. 1998. Word skipping: implications for theories of eye movement control in reading. In *Eye guidance in reading and scene perception*, ed. G. Underwood, 125–148. Oxford: Elsevier.

Calvo, M. G., and E. Meseguer. 2002. Eye movements and processing stages in reading: relative contribution of visual, lexical, and contextual factors. *The Spanish Journal of Psychology* 5(01). 66–77.

Carroll, P., and M. L. Slowiaczek. 1986. Constraints on semantic priming in reading: a fixation time analysis. *Memory & Cognition* 14(6). 509–522.

Clifton, C., Jr., A. Staub, and K. Rayner. 2007. Eye movements in reading words and sentences. In *Eye movements: a window on mind and brain*, ed. R. van Gompel, 341–371. Amsterdam, the Netherlands: Elsevier.

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

DeLong, K. A., M. Troyer, and M. Kutas. 2014. Pre-processing in sentence comprehension: sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass* 8(12). 631–645.

Demberg, V., and F. Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210.

Ehrlich, S. F., and K. Rayner. 1981. Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior* 20.6. 641–655.

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

Fitzsimmons, G., and D. Drieghe. 2013. How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(4). 1054–1063.

Frisson, S., Harvey, D., Drieghe, D., and Staub, A.. 2015. Eye movements in reading unpredictable words: is there a cost? *Manuscript in preparation*.

Frisson, S., K. Rayner, and M. J. Pickering. 2005. Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31. 862–877.

Gallistel, C. R. 2009. The importance of proving the null. *Psychological Review* 116. 439–453.

Gollan, T. H., T. J. Slattery, D. Goldenberg, E. Van Assche, W. Duyck, and K. Rayner. 2011. Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology: General* 140. 189–209.

Hale, J. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of NAACL* 2. 159–166.

Harm, M. W., and M. S. Seidenberg. 2004. Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review* 111. 662–720.

Hand, C. J., S. Miellet, P. J. O'Donnell, and S.C. Sereno. 2010. The frequency-predictability interaction in reading: it depends on where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance* 36. 1294–1313.

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

Juhasz, B. J., S. J. White, S. P. Liversedge, and K. Rayner. 2008. Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance* 34(6). 1560–1579.

Kennedy, A., J. Pynte, W.S. Murray, and S.-A. Paul. 2013. Frequency and predictability effects in the Dundee Corpus: an eye movement analysis. *The Quarterly Journal of Experimental Psychology* 66. 601–618.

Kennison, S. M., and C. Clifton. 1995. Determinants of parafoveal preview benefit in high and low working memory capacity readers: implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(1). 68–81.

Kliegl, R., E. Grabner, M. Rolfs, and R. Engbert. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16. 262–284.

Kliegl, R., A. Nuthmann, and R. Engbert. 2006. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General* 135.12–35.

Kretzschmar, F., M. Schlesewsky, and A. Staub. 2015. Dissociating word frequency and predictability effects in reading: evidence from co-registration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Lavigne, F., F. Vitu, and G. d'Ydewalle. 2000. The influence of semantic context on initial eye landing sites in words. *Acta Psychologica* 104(2). 191–214.

Levy, R. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.

Matsuki, K., T. Chow, M. Hare, J. L. Elman, C. Scheepers, and K. McRae. 2011. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(4). 913–934.

McClelland, J. L. 1979. On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review,* 86, 287–330.

McDonald, S. A., and R. C. Shillcock. 2003. Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research* 43. 1735–1751.

Miellet, S., L. Sparrow, and S. C. Sereno. 2007. Word frequency and predictability effects in reading French: an evaluation of the E-Z Reader model. *Psychonomic Bulletin & Review* 14. 762–769.

Morris, R. K. 1994. Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20. 92–103.

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

Rayner, K. 1975. Parafoveal identification during a fixation in reading. *Acta Psychologica* 39(4). 271–281.

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

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

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

Rayner, K., K. S. Binder, J. Ashby, and A. Pollatsek. 2001. Eye movement control in reading: word predictability has little influence on initial landing positions in words. *Vision Research* 41(7). 943–954.

Rayner, K., X. Li, B. J. Juhasz, and G. Yan. 2005. The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review* 12(6). 1089–1093.

Rayner, K., E. D. Reichle, M. J. Stroud, and C. C. Williams. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging* 21. 448–465.

Rayner, K., Slattery, T. J., Drieghe, D., and Liversedge, S. P.. 2011. Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance* 37(2). 514–528.

Rayner, K., T. Warren, B. J. Juhasz, and S. P. Liversedge. 2004b. The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30. 1290–1301.

Reichle, E. D., A. Pollatsek, D. L. Fisher, and K. Rayner. 1998. Toward a model of eye movement control in reading. *Psychological Review* 105(1). 125–157.

Reichle, E. D., K. Rayner, and A. Pollatsek. 2003. The E-Z Reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences* 26. 445–476.

Reingold, E. M., and K. Rayner. 2006. Examining the word identification stages hypothesized by the EZ Reader model. *Psychological Science* 17(9). 742–746.

Reingold, E. M., E. D. Reichle, M. G. Glaholt, and H. Sheridan. 2012. Direct lexical control of eye movements in reading: evidence from a survival analysis of fixation durations. *Cognitive Psychology* 65(2). 177–206.

Roberts, S., and S. Sternberg. 1993. The meaning of additive reaction-time effects: tests of three alternatives. In *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, eds. D. E. Meyer and S. Kornblum, 611–653. Cambridge, MA: MIT Press.

Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson. 2009. Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review* 16. 225–237.

Schotter, E. R., B. Angele, and K. Rayner. 2012. Parafoveal processing in reading. *Attention, Perception, & Psychophysics* 74(1). 5–35.

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

Sheridan, H., and E. M. Reingold. 2012. The time course of predictability effects in reading: evidence from a survival analysis of fixation durations. *Visual Cognition* 20.733–745.

Smith, N. J., and R. Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3). 302–319.

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

—— 2011b. Word recognition and syntactic attachment in reading: evidence for a staged architecture. *Journal of Experimental Psychology: General* 140. 407–433.

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

Staub, A., Grant, M., Astheimer, L., and Cohen, A.. 2015. The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language* 82. 1–17.

Staub, A., K. Rayner, A. Pollatsek, J. Hyönä, and H. Majewski. 2007. The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33. 1162–1169.

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

Taylor, W. L. 1953. "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly* 30. 415–433.
Traxler, M. J., D. J. Foss, R. E. Seely, B. Kaup, and R. K. Morris. 2000. Priming in sentence processing: intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research* 29(6). 581–595.
Van Petten, C., and B. J. Luka. 2012. Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology* 83.176–190.
Warren, T., and K. McConnell. 2007. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review* 14. 770–775.
White, S. J., and A. Staub. 2012. The distribution of fixation durations during reading: effects of stimulus quality. *Journal of Experimental Psychology: Human Perception & Performance* 38. 603–617.
White, S. J., K. Rayner, and S. P. Liversedge. 2005. The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review* 12. 466–471.
Whitford, V., and D. Titone. 2014. The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *Quarterly Journal of Experimental Psychology* 67. 1151–1165.