Visual Cognition

The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations

Heather Sheridan & Eyal M. Reingold

Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada


To cite this article: Heather Sheridan & Eyal M. Reingold (2012): The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations, Visual Cognition, 20:7, 733-745

To link to this article: http://dx.doi.org/10.1080/13506285.2012.693548
whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations

Heather Sheridan and Eyal M. Reingold

Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada

To investigate the time course of predictability effects in reading, the present study examined distributions of first-fixation durations on target words in a low predictability versus a high predictability prior context. In a replication of Staub (2011), ex-Gaussian fitting demonstrated that the low predictability distribution was significantly shifted to the right of the high predictability distribution in the absence of any contextual differences in the degree of skew. Extending this finding, the present study used a survival analysis technique to demonstrate a significant influence of predictability on fixation duration as early as 140 ms from the start of fixation, which is similar to prior results obtained with the word frequency variable. These results provide convergent evidence that lexical variables have a fast acting influence on fixation durations during reading. Implications for models of eye-movement control are discussed.

Keywords: Distributional analysis; Eye movements; Predictability; Reading; Survival analysis.

Fixation times during reading are influenced by a variety of lexical and linguistic variables (for reviews, see Rayner, 1998, 2009). For example, extensive research concerning the word frequency and predictability (i.e., contextual constraints) variables has shown that fixation times are longer on low frequency than on high frequency words (see White, 2008, for a review), and longer on low predictability than on high predictability words (e.g., Ehrlich & Rayner, 1981). These two lexical effects are widely considered to be benchmark findings that must be accounted for by models of eye movement control during reading (Rayner 1998, 2009).
However, competing models differ considerably in their assumptions concerning the time course of lexical effects. Specifically, one class of models assumes that fixation times are primarily driven by visual/oculomotor factors and that lexical variables can only impact a small subset of long fixations (e.g., Deubel, O’Regan, & Radach, 2000; Feng, 2006; McConkie & Yang, 2003; Yang, 2006; Yang & McConkie, 2001), whereas a competing class of models assumes that lexical and linguistic processes play a nontrivial role in controlling fixation times in reading (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Just & Carpenter, 1980, 1987; Kliegl & Engbert, 2003; McDonald, Carpenter, & Shillcock, 2005; Morrison, 1984; Pollatsek, Reichle, & Rayner, 2006; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2006; Reichle, Rayner, & Pollatsek, 1999, 2003; Reilly & Radach, 2003, 2006; Richter, Engbert, & Kliegl, 2006; Thibadeau, Just, & Carpenter, 1982).

To help distinguish between the competing time course assumptions of the various models, recent work has employed distributional analyses of fixation times during reading (e.g., Reingold, Reichle, Glaholt, & Sheridan, 2012; Sheridan & Reingold, 2012; Staub, 2011; Staub, White, Drieghe, Hollway, & Rayner, 2010; White & Staub, 2011; White, Staub, Drieghe, & Liversedge, 2011; White, Warren, Staub, & Reichle, 2011). In particular, Staub et al. (2010) introduced ex-Gaussian fitting as a technique for modelling individual participants’ distributions of fixation times during reading. A key advantage of ex-Gaussian fitting is that it clarifies whether a variable's impact on mean fixation times is due to a shift in the location of the distribution and/or a change in the degree of skew. As explained by Staub et al., a shift effect indicates that the variable is having an early acting influence on the majority of fixation durations, whereas a skew effect primarily stems from an influence on long fixation durations. In support of an early time course of lexical effects, both the word frequency variable (Reingold et al., 2012; Staub et al., 2010) and the predictability variable (Staub, 2011) have been shown to produce a shift effect.

In addition to ex-Gaussian fitting, another distributional analysis approach has examined survival curves of first-fixation durations during reading (Reingold et al., 2012; see also Sheridan & Reingold, 2012). Specifically, for a given time $t$, the percentage of first fixations with a duration greater than $t$ are referred to as the percentage survival at time $t$. Thus, when $t$ equals zero, survival is at one hundred per cent, but then declines as $t$ increases and approaches zero per cent as $t$ approaches the duration of the longest observed first fixation. Reingold et al. (2012) examined the time course of word frequency effects during reading by calculating separate survival curves for first-fixation durations on low frequency and high frequency target words. They then examined the earliest point in time at which the high and low frequency survival curves began to
significantly diverge (henceforth referred to as the divergence point). Importantly, Reingold et al. argued that the divergence point provided an estimate of the earliest significant influence of the word frequency variable. Based on their survival curve analyses, they concluded that there is a significant influence of word frequency on fixation duration in normal reading as early as 145 ms from the start of fixation (for other applications of survival curves, see Feng, Miller, Shu, & Zhang, 2001; Glaholt & Reingold, 2012; see also Reingold & Sheridan, 2011, for a review of the use of survival analysis in the study of visual expertise).

The main goal of the present study is to employ the survival analysis technique (Reingold et al., 2012) to study the time course of the influence of predictability on first-fixation duration. Specifically, the same target words were read once in a low predictability context and once in a high predictability context. Using the technique introduced by Reingold et al. (2012), we examined survival curves for the distributions of first-fixation durations on target words in the low and high predictability contexts, and we then calculated the divergence point for the low and high predictability curves in order to provide an estimate of the earliest significant influence of the predictability manipulation. Given that ex-Gaussian analyses have demonstrated a shift effect for predictability (Staub, 2011), we predicted that the present study’s predictability manipulation might produce a divergence point that was equally as early as the word frequency divergence point that was reported by Reingold et al. Such a finding would provide convergent evidence for an early time course of lexical variables during reading.

**METHOD**

**Subjects**

All 60 undergraduate students at the University of Toronto who participated were native English speakers and were given either course credit or $10 (Canadian) per hour. All had normal or corrected-to-normal vision.

**Materials**

Forty sentence pairs were created such that a target word (e.g., “muscle”) was either highly predictable (e.g., “The athlete pulled a muscle in his leg

---

1 For the distributional analyses, we used the first-fixation duration measure because it provides more data than the single-fixation measure. First-fixation duration is the duration of the first forward fixation on a word regardless of the number of subsequent fixations on the word, whereas single-fixation duration is the first-fixation value for the subset of trials in which there was only one first-pass fixation on the word.
during the competition’’) or unpredictable (e.g., “Peter says that a muscle in his leg was bothering him during soccer practice’’). To obtain cloze norms for these sentence pairs, an independent sample of 20 undergraduates were given the sentence frames up to but not including the target words, and they were asked to produce the word that seemed most likely to come next in the sentence. The high predictability sentences had a mean cloze probability of .78 (range = .5 to 1) and the low predictability sentences had a mean cloze probability of .01 (range = 0 to .08). In addition, to ensure that the local context surrounding the target word was similar in complexity and difficulty across the two sentence versions, an average of 1.4 words immediately before the target and 2.8 words after the target remained constant across the two sentence versions. The mean word length of the target words was 6.25 (range = 4 to 9 letters), and the mean SUBTLex word frequency (Brysbaert & New, 2009) was 66.76 occurrences per million.

Participants read a total of 255 sentences, including five practice trials, 80 experimental trials, and 170 nonexperimental filler sentences that were designed to mask the experiment’s purpose. To provide sufficient power for the distributional analyses, the participants were shown both the high predictability and low predictability versions for all 40 of the sentence pairs. The order of trials was randomized with the constraint that each target word was read once in each half of the experiment, with 75 nonexperimental filler trials separating the two halves of the experiment. For each participant, half of the target words appeared in the high predictability sentence first, and half of the target words appeared in the low predictability sentence first, and the order of the two versions of the sentence pairs was counterbalanced across participants. All sentences were displayed on a single line, and the target words were located near the middle of the sentences.

**Apparatus and procedure**

Eye movements were measured with an SR Research EyeLink 1000 system with high spatial resolution and a sampling rate of 1000 Hz. Following calibration, gaze-position error was less than 0.5°. The sentences were displayed on a 21-inch ViewSonic monitor. All letters were lowercase (except when capitals were appropriate) and in a monospaced Courier font. The text was presented in black (4.7 cd/m²) on a white background (56 cd/m²). Participants were seated 60 cm from the monitor, and 2.4 characters equalled approximately 1 degree of visual angle. Participants were instructed to read the sentences for comprehension. After reading each sentence, they pressed a button to end the trial and proceed to the next sentence. To ensure that participants were reading for comprehension, about 20% of the sentences were followed by multiple-choice comprehension questions. The average accuracy rate was 97%.
RESULTS AND DISCUSSION

The present study’s main focus was on using a survival analysis technique (Reingold et al., 2012) to study the time course of the predictability variable. However, prior to reporting the survival analysis results, we first confirm that the present study’s predictability manipulation can replicate past findings by examining standard eye movements measures, as well as the ex-Gaussian parameters that were reported by Staub (2011).

For the analyses reported here, we did not use any outlier rejection cutoffs. However, 14.0% of trials were removed due to skipping of the target words. For the remaining trials, the following measures were used to compare mean fixation times on the target words in the high predictability context versus the low predictability context conditions: (1) First-fixation duration (i.e., the duration of the first forward fixation on the target, regardless of the number of subsequent fixations on the target); (2) gaze duration (i.e., the sum of all the consecutive first-pass fixations on the target, prior to a saccade to another word); (3) single-fixation duration (i.e., the first-fixation value for the subset of trials in which there was only one first-pass fixation on the target); (4) go-past time (i.e., the sum of all fixations from the first fixation on the target up to and including the fixation prior to the reader moving past the target to a later part of the sentence); (5) the probability of skipping (i.e., trials in which there was no first-pass fixation on the target regardless of whether or not the target was fixated later in the trial); and (6) the probability of a single first-pass fixation. For all of these measures, planned comparisons by participants ($t_1$) and by items ($t_2$) were performed across the two context conditions. Table 1 presents the means and standard errors of the different measures and the corresponding $t$-test results. As shown in Table 1, fixation times on the target words were longer for the low predictability context condition relative to the high predictability context condition, and this effect was significant for first-fixation, gaze duration, single-fixation, and go-past time. In addition, the target words were more likely to be skipped in the high predictability condition than in the low predictability condition, but the probability of a single fixation did not differ across conditions. Overall, this pattern of results replicates a large body of prior work concerning the predictability variable (e.g., Ehrlich & Rayner, 1981; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Well, 1996).

Next, we fitted the ex-Gaussian distribution to the first-fixation duration data using a similar procedure as Staub et al. (2010). The ex-Gaussian distribution (Ratcliff, 1979) is the convolution of the Gaussian normal distribution and an exponential distribution, and can be specified with the following three parameters: $\mu$ (the mean of the Gaussian component), $\sigma$ (the standard deviation of the Gaussian component), and $\tau$ (the mean
and the standard deviation of the exponential component). Following Staub et al., the first-fixation duration data for each participant in each condition were fitted separately, using an algorithm known as quantile maximum likelihood estimation (QMPE; Cousineau, Brown, & Heathcote, 2004; Heathcote, Brown, & Mewhort, 2002). There were an average of 34 usable observations per cell, and all fits successfully converged. Table 2 displays the mean number of usable observations per cell, the means of the μ, σ, and τ parameters, and the magnitude and significance of the low versus high predictability context effects, and Figure 1 displays the distributions of first-fixation duration (Panel a) and the density functions generated from the best-fitting ex-Gaussian parameters (Panel b). As shown in Table 2 and Figure 1, the low predictability distribution was shifted to the right of the high predictability distribution, resulting in a significant μ effect. There were no significant differences for σ and τ. This pattern of ex-Gaussian results replicates Staub’s (2011) predictability findings, and stands in contrast to prior findings that word frequency affects both μ and τ (Reingold et al., 2012; Staub et al., 2010). Thus, predictability effects on mean fixation times primarily reflect a shift in the location of the distributions, whereas word frequency effects reflect both a shift effect and a change in the degree of skew.

**Table 1**

Average fixation time measures (in milliseconds) and the probability (proportion) of skipping and single fixation by context condition

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low predictability</th>
<th>High predictability</th>
<th>Difference</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-fixation (all trials)</td>
<td>216 4.3</td>
<td>208 3.6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 3.34, p &lt; .01</td>
<td>t2 = 2.59, p &lt; .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze duration (all trials)</td>
<td>243 6.3</td>
<td>230 5.2</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 3.46, p &lt; .01</td>
<td>t2 = 2.63, p &lt; .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-fixation</td>
<td>218 4.5</td>
<td>208 3.8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 3.12, p &lt; .01</td>
<td>t2 = 2.67, p &lt; .05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Go-past time</td>
<td>302 9.3</td>
<td>274 8.3</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 5.01, p &lt; .001</td>
<td>t2 = 3.57, p &lt; .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of skipping</td>
<td>.12 .01</td>
<td>.16 .01</td>
<td>−.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 3.61, p &lt; .01</td>
<td>t2 = 3.16, p &lt; .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of single fixation</td>
<td>.75 .01</td>
<td>.74 .01</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t1 = 1.00, p = .319</td>
<td>t2 = 1.00, p = .323</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the t-tests shown above, df for t1 = 59, and df for t2 = 39. The means and standard errors shown are based on the by-participants analyses.
To extend these ex-Gaussian findings, we next computed survival curves for first-fixation durations in the low and high predictability conditions, using the same procedure as Reingold et al. (2012). Specifically, for each 1 ms time bin \( t \) (\( t \) was varied from 0 to 600 ms), the percentage of first fixations with a duration greater than \( t \) constituted the percentage of survival at time \( t \). The survival curve was computed separately for each condition and for each participant, and then averaged across participants. As shown in Figure 1c, the low and high predictability survival curves appear to diverge. As previously argued by Reingold et al., this divergence point corresponds by definition to the shortest first-fixation duration value at which the contextual manipulation had a significant impact. To estimate the divergence point, we employed a bootstrap resampling procedure (Efron & Tibshirani, 1994). Specifically, on each iteration of this procedure, the set of observations (first-fixation durations) for each participant in each condition was randomly resampled with replacement. For each iteration of the bootstrap procedure, an individual participant’s survival curves were then computed and averaged. Next, the value for each 1 ms bin in the high predictability survival curve was subtracted from the corresponding value in the low predictability survival curve. This procedure was repeated 10,000 times, and the obtained differences for each bin were then sorted in order of magnitude. The range between the fifth and the 9995th value was then defined as the confidence interval of the difference for each bin (given the multiple comparisons we performed, we used this conservative confidence interval in order to protect against making a Type I error). To compute the divergence point between the low and high predictability survival curves, we identified the time bins for which the low predictability survival rate was significantly greater than the high predictability survival rate (i.e., for which the lower bound of the confidence interval of the difference between the low and high predictability curves was greater than zero). The divergence point was then defined as the earliest significant difference point that was part of a run of five consecutive significant difference points (significant differences

### Table 2

Number of observation per cell and ex-Gaussian parameters by condition, with standard errors shown in parentheses

<table>
<thead>
<tr>
<th></th>
<th>( n )</th>
<th>( \mu ) (( \mu ))</th>
<th>( \sigma ) (( \sigma ))</th>
<th>( \tau ) (( \tau ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low predictability</td>
<td>35 (0.4)</td>
<td>165 (3.4)</td>
<td>37 (2.4)</td>
<td>52 (4.6)</td>
</tr>
<tr>
<td>High predictability</td>
<td>34 (0.5)</td>
<td>157 (2.9)</td>
<td>38 (2.2)</td>
<td>51 (2.6)</td>
</tr>
<tr>
<td>Difference</td>
<td>1</td>
<td>8</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Significance</td>
<td>( t = 3.61, p &lt; .001 )</td>
<td>( t = 2.48, p &lt; .05 )</td>
<td>( t &lt; 1 )</td>
<td>( t &lt; 1 )</td>
</tr>
</tbody>
</table>

For the \( t \)-test results shown above, \( df = 59 \).
Figure 1. (a) Distributions of first-fixation duration on target words in the low and high predictability conditions, (b) ex-Gaussian density functions, and (c) survival curves. The row of asterisks at the top of Panel (c) indicates the time bins with a significant difference between the low and high predictability curves. See text for further details.
between the low and high predictability curves are shown in Figure 1c as a row of asterisks above the survival curves).

As can be seen in Figure 1c, the low and high predictability survival curves significantly diverged at a duration of 140 ms, which is similar to the 145 ms divergence point that was previously obtained for the word frequency variable under normal reading conditions (see Reingold et al., 2012). Furthermore, the divergence point defines the percentage of first fixations with durations that were too short to exhibit an influence of predictability. In the present study, only approximately 10% of first fixations had durations that were shorter than the divergence point, which is once again similar to the prior word frequency findings (Reingold et al., 2012). Thus, the predictability and word frequency variables produce equally fast-acting influences on fixation duration, which is consistent with the finding that both of these variables produce a $\mu$ effect. However, as pointed out by Staub (2011), even though the two effects show a similar early time course, there appears to be some later emerging differences between them, as indicated by the finding of a $\tau$ effect for word frequency but not for predictability.

**GENERAL DISCUSSION**

The present study's results support an early temporal onset of predictability effects during reading. Specifically, target words were read faster when they were encountered in a predictable context, relative to an unpredictable context, and this effect emerged as early as the first fixation on the target word. In replication of Staub (2011), the distribution of first-fixation durations in the low predictability context was shifted to the right of the high predictability distribution, in the absence of any contextual differences in the degree of skew. Moreover, extending these findings, the present study used a survival analysis (see Reingold et al., 2012) to demonstrate a rapid effect of predictability on fixation times that emerged as early as 140 ms from the start of fixation.

The present study’s findings converge with prior results (e.g., Rayner, Liversedge, White, & Vergilino-Perez, 2003; Reingold, Yang, & Rayner, 2010; Sereno, Brewer, & O’Donnell, 2003; Sereno, Rayner, & Posner, 1998; Staub, 2011; Staub et al., 2010) in supporting an early time course of lexical influences during reading. In particular, the finding that lexical variables can produce a shift in the distributions indicates that both short and long fixations were impacted (Staub et al., 2010), and the present study’s survival analyses coincide with prior work (Reingold et al., 2012; Sheridan & Reingold, 2012) in demonstrating rapid lexical influences on fixation durations. Taken together, such findings are consistent with models of eye-movement control that postulate fast-acting lexical influences on the
majority of fixation durations during reading, and are inconsistent with a competing class of models that instead assumes that most fixation durations are exclusively driven by visual/oculomotor factors (see Rayner, 1998, 2009 for reviews).

An additional implication of the present study’s results concerns the temporal relationship between the predictability and word frequency variables. Specifically, the two variables appear to be temporally overlapping during an early stage of processing, as evidenced by the present study’s predictability divergence point of 140 ms, which is comparable to the word frequency divergence point of 145 ms that was previously reported by Reingold et al. (2012). In addition, due to the temporal constraints inherent to saccadic programming in reading (Sereno & Rayner, 2003), these lexical variables begin to exert their influence even earlier than these divergence points (Reingold et al., 2012). This pattern of survival analysis results is consistent with ERP work that demonstrated both word frequency and context effects during the N1 component, from 132–192 ms poststimulus (Sereno et al., 2003). Moreover, with regards to the ex-Gaussian findings, the survival analysis results are generally consistent with the finding that both of the variables can produce a shift in the location of the distributions. However, as discussed by Staub (2011), the finding that there is a skew effect for word frequency but not for predictability indicates that there are some temporal differences between the two variables with respect to long fixations.

Given that word frequency effects are considered to be a marker of lexical access (Rayner, 1998), the similar temporal onset of word frequency and predictability effects suggests that contextual predictability influences are rapid enough to impact early lexical processes (see Sereno et al., 2003, for a related discussion). However, temporal overlap is not sufficient evidence that both variables are impacting the same stage or process. In fact, previous studies that manipulated both word frequency and predictability have often shown that the two variables do not interact with each other (e.g., Rayner et al., 2004; Slattery, Staub, & Rayner, 2012). Thus, although distributional analyses have revealed an early time course for several prominent reading variables, further work is needed to clarify the exact processes and mechanisms that underlie these early effects.

REFERENCES


*Manuscript received March 2012*

*Manuscript accepted April 2012*

*First published online July 2012*