Abstract

Return-sweeps are an essential eye-movement during reading that take fixation from the end of one line to the start of the next. Despite their regular occurrence during normal reading, the eye movement literature is dominated by single line reading experiments devoid of return-sweeps. Here we present an eye-movement experiment of return-sweeps designed to explore what readers are targeting with their return-sweeps. Participant’s read two short stories of Oz by Frank L. Baum while their eye movements were monitored. One story was presented normally while the other was manipulated to highlight the first word on each line of text by bolding it. This bolding manipulation had a significant impact on the return-sweeps—resulting in these saccades landing closer to the left margin of the new line compared to the control condition. However, this effect of bolding had no influence on the duration of line initial fixations or the standard reading time measures of the line initial words. Moreover, the impact of bolding didn’t interact with the length of the line initial word. We discuss the implication of these findings for the targeting of return-sweeps and eye movement control during reading.

*Keywords*: eye-movements, reading, return sweep, bold text, saccade targeting

Word count: XX words

**Introduction**

The scientific study of reading has benefitted immensely from the examination of people’s eye movements while they read (see Rayner, 2009 for a review). While there is a growing number of eye movement studies examining the return-sweep in reading, the vast majority of eye movement reading studies involve only single lines of text and are therefore devoid of return-sweeps. Furthermore, our understanding of return-sweep saccades is still largely limited to a description of their characteristics. For instance, we know that return-sweeps are normally launched from around 5 characters from the end of a line of text (add refs) and these saccades land around 6 characters from the leftmost character of the next line (add refs), however landing positions are strongly influenced by the length of the lines of text. With longer lines of text, the landing position shifts rightward further into the line (Hofmiester, Heller, & Radach, 1997). This rightward shift with increasing line length is the result of saccadic range error; the finding that saccades to distant targets tend to undershoot them, while saccades to nearby targets tend to overshoot them (add refs). However, what is still an unresolved question is what readers are targeting with their return-sweeps.

* With intra-line reading saccades, there is considerable evidence to suggest that readers are targeting word centres for their saccades (McConkie & Zola, 1984; Rayner, 1979).
  + detail the Rayner 79 results comparing regression landing sites within words to forward fixation landing sites within words.
  + describe the preferred viewing position, the optimal viewing position, and the IOVP effect.
  + discuss mislocated fixations and the assumption of overlapping Gaussian landing positions based on word units (e.g. Drieghe, Rayner, & Pollatsek, 2008; Nuthmann, Engbert, & Kliegl, 2005).
  + if word count permits, possibly add a discussion of word grouping findings (e.g. Drieghe, Pollatsek, Staub, & Rayner, 2008).
* Describe return-sweep planning and execution (Hofmiester, Heller, & Radach, 1997 Exp 2.) and detail how it may be different from intra-line saccade planning.
  + longer movements, and therefore more error
    - unlike long intra-line regressions, planned to areas that haven’t been visually processed yet so there should be no spatial memory of word locations.
  + line initial words may be too far in periphery to be accurately segmented—readers may lack the word length information needed to target a word centre.
  + instead it may be possible that readers are targeting an area relative to the leftmost character on the next line.
* Indicate that visual saliency influences saccade programming (work of Tatler and colleagues).
  + make argument that bolding the first word on each line will make them more salient and should strengthen the length information of these words thereby allowing word centres to be determined.
* Predictions: if readers are attempting to target their return-sweeps to line initial words, then the landing site distributions of these saccades should shift to the right as line initial words get longer. Such a pattern, if it were to exist, should be more apparent when line initial word length information is made more salient with bolding.

**Method**

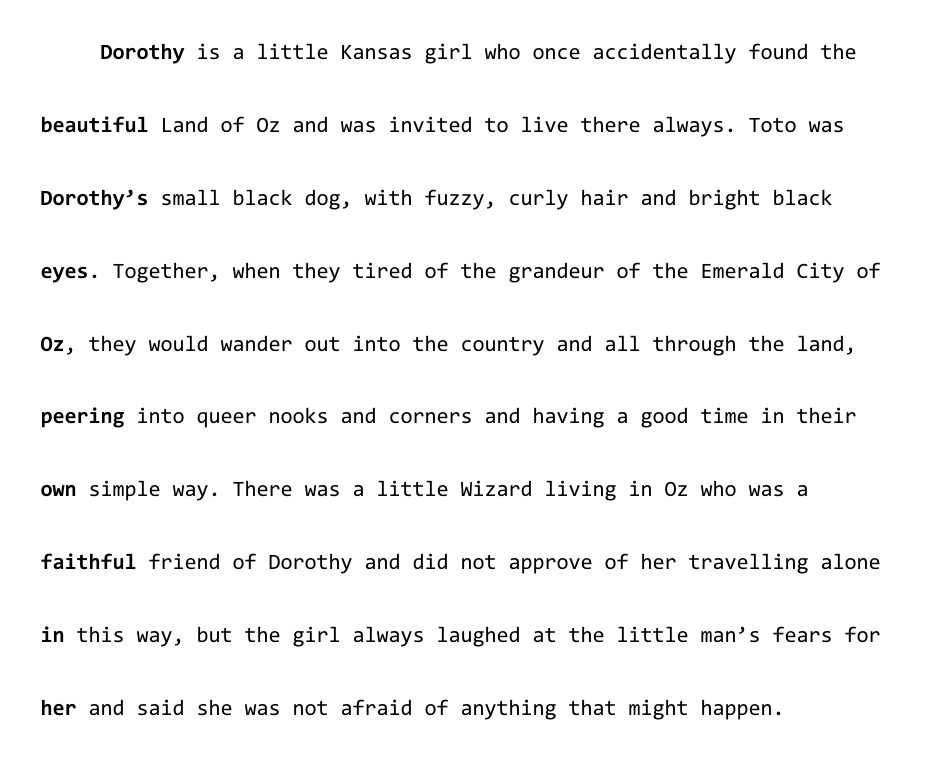
**Participants**

Thirty-two members of the Bournemouth community (20 female) participated for a payment of £10. Their mean age was 28 years (*SD*= 12.1 years; range: 19-63 years). Participants were speakers of English who reported normal or corrected-to-normal vision and no prior history of reading disorders. Participants were naïve as to the purpose of the experiment. The study was approved by the Bournemouth University Research Ethics Committee (protocol No. 16769).

**Materials and Design**

The reading stimuli consisted of two illustrated short stories from the book “Little wizard stories of Oz” by L. Frank Baum (Baum, 2008/1914). The stories were “Little Dorothy and Toto” and “Tiktok and the Nome King”. The text was divided into 17 text screens for the first story and 18 text screens for the second story. Each text screen was considered as a separate item in the statistical analyses. The illustrations were presented on separate screens at the point in which they occurred in the original stories and were not accompanied by any text from the story.

There were two experimental conditions: 1) *Bold type* condition in which the first word on each line was formatted in bold typeface (see Figure 1); and 2) a *Normal type* control condition in which the first word on the line was formatted normally (i.e., not in bold). Each story was assigned to one of the two experimental conditions. In the Bold type condition, the bolding on each line remained present for the duration of the whole story. The assignment of conditions and the order of the two stories were counterbalanced with a full Latin square design.



*Figure 1*. An example page from the story “Little Dorothy and Toto” in the Bold experimental condition. In the Normal condition, each line-initial word was formatted normally (i.e., not in bold).

**Apparatus**

Participants’ eye-movements were recorded with an EyeLink 1000 Tower Mount at a sampling frequency of 1000 Hz. Viewing was binocular, but only the right eye was recorded[[1]](#footnote-1). Participants rested their head on a chin-and-forehead rest in order to reduce head-movement artefacts. The experiment was programed in Python 2.7 by using the PsychoPy (Peirce, 2007) and PyGaze (Dalmaijer, Mathôt, & Van der Stigchel, 2014) libraries.

The text stimuli were presented on a Cambridge Research Systems LCD++ monitor. The screen resolution was 1920 x 1080 pixels and the refresh rate was 100 Hz. The text was formatted in a monospaced Consolas 11 pt. font and appeared as black text over white background at the centre of the screen. The text was double spaced and aligned to the left. The width of each letter was 12 pixels. Participants sat 80 cm away from the monitor and at this distance each letter subtended approximately 0.30° per visual angle. The experiment was run on a PC in a Windows 7 environment.

**Procedure**

Participants were tested in a session that lasted for approximately 35-45 minutes. Before the experiment, participants were calibrated on a 9-point calibration grid. Additionally, a drift check was presented before each trial and participants were re-calibrated whenever that was necessary. The calibration was kept at < .40 ° across the experiment. Trials started with a black gaze box that was centred at the location of the first letter on the first line. Once a stable fixation on the gaze box was detected, the box disappeared and the text was presented on the screen.

Participants were instructed to read the stories as they would typically read a book. As we felt the bolding manipulation would be obvious to most if not all of the participants, they were informed that one of the stories would have a typographic manipulation and to just read this story normally. Participants were free to read the stories as they wished, and could move back or forward through the pages by pressing the right and left arrow keys on the keyboard. While they had the opportunity to go back a page, none of the participants actually did so. The reading of each story was terminated when participants reached the last page of text and pressed the button to move forward. After each story, participants answered five multiple-choice comprehension questions about the contents of the story by pressing a button on the keyboard to indicate their answer.

**Data Analysis**

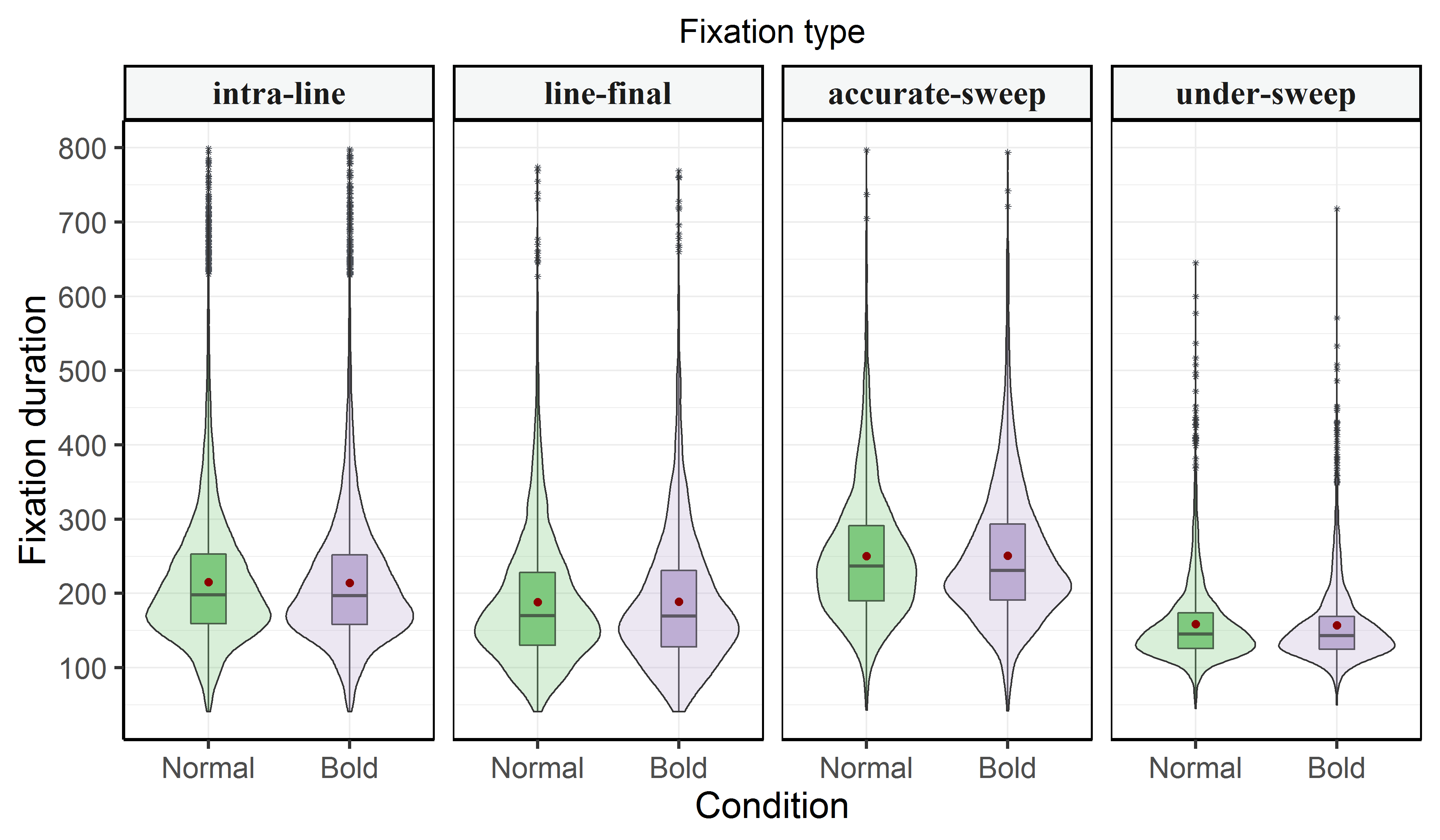
The experiment had one within-subject factor with two levels: the first word on a line was either formatted in bold type (“Bold” condition) or in a normal type (“Normal” condition). A few analyses were carried out in this experiment. The first analysis tested whether the Bold text condition affected the four types of fixations in the experiment: 1) *intra-line fixations* (i.e., fixations occurring within a line that are not immediately preceded or followed by a return sweep), *line-final fixations* (i.e., the last fixation on a line, immediately before readers execute a return sweep to the start of the next line), *under-sweep fixations* (i.e., fixations that land short of the start of a new line after a return sweep and are then followed by a corrective saccade to the left), and *accurate-sweep* *fixations* (i.e., fixations after a return sweep that land close enough to the intended location on a new line that they are followed by a saccade to the right). Line-final, under-sweep and accurate-sweep fixations were coded as such only during the first pass-reading of the text. Fixations occurring during the second-pass reading of the text were coded as intra-line fixations. Two additional analyses tested the landing position on a new line and the probability of making a return sweep as a function of the experimental condition.

The data were analysed with (Generalised) Linear Mixed Models ((G)LMMs) by using the “lme4” package v. 1.1-12 (Bates, Machler, Bolker, & Walker, 2014) in the R statistical software v.3.5.1 (R Core Team, 2018). Fixation durations were log-transformed in all models. Participants and items were added as random intercepts in the models (Baayen, Davidson, & Bates, 2008). Additionally, random intercepts for experimental condition were also added for both participants and items (Barr, Levy, Scheepers, & Tily, 2013)[[2]](#footnote-2). Treatment contrast coding was used for the text type condition, where Bold was the baseline. The fixation type variable was also coded with a treatment contrast, where intra-line fixations was the baseline. The results were considered as statistically significant if the |*t*| or |*z*| values were ≤ 1.96.

**Results**

The fixation data were manually processed with the EyeDoctor software (Stracuzzi & Kinsey, 2009) to re-align the vertical position of fixations when necessary. Fixations shorter than 80 ms that occurred within one character space of another fixation were combined with that fixation. Fixations longer than 800 ms were removed as outliers in all analyses (0.12 % of all observations). All participants had comprehension accuracy greater than 80 %, thus indicating that they understood the stories. There was no significant difference in comprehension accuracy between the Bold (*M*= 96.2 %, *SD*= 19 %) and Normal (*M*= 97.5 %, *SD*= 15.6 %) text condition, *b*= 0.69, *SE*= 1.51, *z*= 0.46.

The distribution of fixation durations for the four fixation types is illustrated in Figure 2. Line-final (*b*= -0.15, *SE*= 0.006, *t*= -26.46) and under-sweep fixations (*b*= -0.28, *SE*= 0.008, *t*= -36.72) were significantly shorter than intra-line fixations. In contrast, accurate-sweep fixations were significantly longer than intra-line fixations (*b*= 0.18, *SE*= 0.009, *t*= 19.36). There was no main effect of experimental condition (*b*= -0.005, *SE*= 0.006, *t*= -0.81) or any interactions between experimental condition and fixation type (all |*t|*s ≤ 0.47). Therefore, participants spent less time fixating the end of the line immediately before a return sweep and when they undershoot the start of the next line. In contrast, participants fixated longer at the beginning of a new line after making an accurate return sweep. Critically, however, the bold text condition had no influence on the duration of any of the four fixation types. A separate model was fit predicting the launch position of return sweeps as a function of the experimental condition. The results indicated no difference in return sweep launch position between the Normal and Bold conditions (b= -0.28, SE= 0.41, t= -0.68), thus suggesting that the bolding manipulation did not influence the position from which participants launched their return sweeps.



*Figure 2*. Probability density and box plots of the four fixation types as a function of experimental condition (Normal vs. Bold). Horizontal black lines on the boxplots indicate the median and dark red circles indicate the mean.

The results from the landing position analysis of return sweep saccades are presented in Table 1. There was a main effect of experimental condition, which indicated that participants landed, on average, 0.52 characters closer to the beginning of the new line in the Bold compared to the Normal condition. Additionally, there was a main effect of launch site, which was due to participants landing closer to the beginning of a new line when they also launched from closer to the beginning of the previous line (i.e. when the saccade didn’t have as far to go to reach the start of the line). Additionally, there was a significant interaction between launch site and the length of the first word on the new line (see Figure 3 for an illustration). This was due to participants landing further away from the beginning of the new line with increasing length of the first word, but only when the launch position was distant (approx. > 50 characters); when the launch site was closer the to the beginning of the new line, the reverse trend was observed and the saccade landed closer to the start of the new line with increasing length of the first word. No other effects were statistically significant.

Table 1

*LMM Results for Landing Position Relative to the Start of the New Line as a Function of Experimental Condition, Incoming Saccade Length, and Length of the First Two Words on the Line*

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | b | SE | t |
| Intercept | 7.187 | 0.42 | **17.111** |
| Cond | -0.535 | 0.139 | **-3.852** |
| Launch | 0.534 | 0.06 | **8.837** |
| LenW1 | 0.026 | 0.059 | 0.429 |
| LenW2 | 0.096 | 0.06 | 1.605 |
| Cond x Launch | -0.115 | 0.084 | -1.377 |
| Cond x LenW1 | 0.063 | 0.083 | 0.763 |
| Launch x LenW1 | 0.149 | 0.07 | **2.136** |
| Cond x LenW2 | -0.072 | 0.083 | -0.863 |
| Launch x LenW2 | 0.023 | 0.061 | 0.374 |
| LenW1 x LenW2 | 0.047 | 0.067 | 0.704 |
| Cond x Launch x LenW1 | -0.017 | 0.096 | -0.177 |
| Cond x Launch x LenW2 | 0.094 | 0.087 | 1.085 |
| Cond x LenW1 x LenW2 | -0.007 | 0.094 | -0.071 |
| Launch x LenW1 x LenW2 | 0.121 | 0.075 | 1.62 |
| Cond x Launch x LenW1 x LenW2 | 0.105 | 0.109 | 0.968 |

*Note*: Statistically significant *t*-values are formatted in bold. Cond: Experimental condition (Bold vs. Normal). Launch: launch position of the return sweep saccade (centred at 0). LenW1: length of the first word on a line in characters (centred at 0). LenW2: length of the second word on a line in characters (centred at 0).

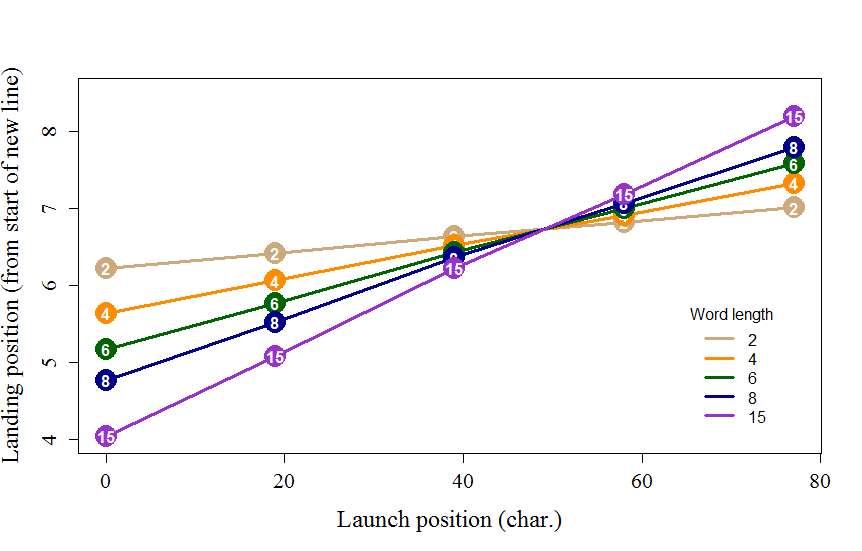
*Figure 3*. An illustration of the significant interaction between launch position and word length from the landing position model in Table 2.

Table 2

*GLMM Results of Under-sweep probability in the Experiment*

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | b | SE | z |
| (Intercept) | 0.535 | 0.174 | **3.081** |
| Condition | -0.249 | 0.062 | **-4.012** |
| Launch | 0.673 | 0.037 | **18.129** |
| Condition x Launch | 0.019 | 0.052 | 0.366 |

*Note*: Statistically significant *z*-values are formatted in bold. Launch: launch position of the return sweep saccade from the beginning of the previous line (centred at 0).

The results from the GLMM analysis of under-sweep probability are presented in Table 2. There was a main effect of experimental condition, indicating that the probability of making an under-sweep fixation was smaller in the Bold compared to the Normal condition. Additionally, there was also a main effect of launch position, which was due to greater probability of making an under-sweep fixation with increasing distance between the launch site and the beginning of the new line. Interestingly, while the main effect of experimental condition was significant, the interaction with launch site was not. This suggests that the Bold condition reduced the probability of making an under-sweep fixation regardless of where participants launched their return sweep saccade.

The results for word-level fixation duration measures are presented in Table 3. The standard lexical frequency and word length effects were significant for both gaze duration and total viewing time, indicating that fixation durations became longer with decreasing lexical frequency and increasing word length. However, there was no main effect of experimental condition or any interactions with lexical frequency or word length. This suggests that the experimental manipulation had no influence on the lexical processing of line-initial words.

Table 3

*LMM Results for Fixation Duration Measures on Line-initial Words in the Text*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Effect | Gaze duration | | | Total Viewing Time | | |
| b | SE | t | b | SE | t |
| Intercept | 5.753 | 0.047 | **122.07** | 5.839 | 0.05 | **117.55** |
| Condition | -0.031 | 0.052 | -0.586 | -0.012 | 0.058 | -0.214 |
| Freq | -0.025 | 0.003 | **-7.689** | -0.029 | 0.004 | **-8.101** |
| WordLen | 0.093 | 0.026 | **3.57** | 0.1 | 0.029 | **3.491** |
| Condition x Freq | 0.001 | 0.005 | 0.301 | 0.001 | 0.005 | 0.197 |
| Condition:WordLen | 0.054 | 0.036 | 1.489 | 0.053 | 0.04 | 1.328 |
| Freq: WordLen | -0.004 | 0.003 | -1.565 | -0.005 | 0.003 | -1.805 |
| Condition: Freq: WordLen | -0.004 | 0.004 | -1.184 | -0.004 | 0.004 | -0.911 |

*Note*: Statistically significant *t*-values are formatted in bold. Freq: lexical frequency (log-transformed). WordLen: word length (centred at 0). Because of the literary style of the stories, 10 % of the line-initial words did not have lexical frequency entries in the SUBTLEX-UK database (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). These words were excluded from the analyses.

**Discussion**

**-** potentially discuss in terms of a “when” and “where” dissociation- fixation durations (when) are not affected by the bolding but saccade targetting (where) is.

**Acknowledgments**

Martin Vasilev was supported by a PhD studentship from Bournemouth University. We thank Milla Räsänen, Bradley Barnard, Grace Kempton, and Mia Pang for their help with the data collection and pre-processing.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Machler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, L. F. (2008). *Little wizard stories of Oz*. Urbana, Illinois: Project Gutenberg. Retrieved from http://www.gutenberg.org/ebooks/55 (Original work published 1914)

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Stracuzzi, D. J., & Kinsey, J. D. (2009). EyeDoctor (Version 0.6.5) [Computer Software]. Retrieved from http://blogs.umass.edu/eyelab

Team, R. C. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

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

1. The left eye was recorded for four participants as a result of tracking problems with their right eye due to glasses or contact lenses. [↑](#footnote-ref-1)
2. The random slope for items was removed for the comprehension accuracy and under-sweep probability models due to convergence failure. [↑](#footnote-ref-2)