

The lexical access of multiple words during a single fixation: overlapping access processes?

Laura J. Wakeford
Wayne S. Murray

This is an Accepted Manuscript of an article published by Taylor & Francis in Language, Cognition and Neuroscience on 16 July 2020, available online:

<http://www.tandfonline.com/10.1080/23273798.2020.1792952>

The lexical access of multiple words during a single fixation: Overlapping access processes?

Laura J. Wakeford¹ & Wayne S. Murray²

¹Abertay University, Dundee, Scotland

²Psychology, University of Dundee, Dundee, Scotland

Address for Correspondence:

Laura J. Wakeford

Division of Psychology

Abertay University

Dundee,

DD1 1HG

Scotland

Tel: +44 1382 308591

Email: l.wakeford@abertay.ac.uk

This work was supported by an Economic and Social Research Council (ESRC) studentship awarded to the first author while at the University of Dundee, under grant ES/H018824/1.

The lexical access of multiple words during a single fixation: Overlapping access processes?

Since it has become increasingly difficult to tease apart the predictions of serial and parallel models of eye movement control during reading, we return to the underlying theoretical question of whether parallel lexical processing of two words is, at the very least, psychologically plausible. Two horizontally aligned letter strings were presented simultaneously on a screen, the task being to decide whether they were physically identical or not. Even with presentation durations short enough to prohibit serial inspection of each word the results show clear lexical effects: high frequency word pairs were responded to faster and with fewer errors than low frequency words. Effects of lexicality, orthography and scanning direction were also found. The results suggest that two words can be processed at a lexical level in an overlapping fashion.

Keywords: word recognition; models of eye movement control; same-different matching; parallel lexical processing; attention.

Introduction

Along with development in the nineties of computational models of eye movement control during reading, there was a sustained increase in research aiming to determine whether multiple words are lexically processed in an overlapping fashion¹. According to serial models such as E-Z Reader (Reichle, Rayner & Pollatsek, 2003), attention can only move to the next word in text when the currently-attended word has been fully identified. In contrast, parallel models such as SWIFT (Engbert, Nuthmann, Richter & Kliegl, 2005) allow overlapping lexical processing of multiple words falling within the perceptual span (McConkie & Rayner, 1976).

Historically, distinguishing these perspectives largely focussed on identifying effects considered incompatible with the serial perspective. In recent years, however, proponents of the serial perspective have suggested that many seemingly parallel effects can be accounted for within the architecture of E-Z Reader, potentially rendering them less diagnostic than previously thought. For example, it has been suggested that so-called *parafoveal-on-foveal effects*, in which lexical or post-lexical properties of a parafoveal word are reflected in foveal inspection time (e.g., Hyönä & Bertram, 2004; Murray & Rowan, 1998), could be a consequence of calibration (e.g. Reichle & Drieghe, 2015), or saccadic targeting errors, with processing of the upcoming word, although the eye has landed short of this target (Drieghe, Rayner & Pollatsek, 2008)². Although whether it is possible, within the architecture of E-Z-Reader, to integrate such a “stay-and-process” response to a saccadic undershoot without compromising its ability to account for other benchmark effects, such as the inverted optimal viewing position effect that assumes a quick error correcting saccade in response to saccadic error (e.g., Vitu, McConkie, Kerr & O’Regan 2001), is yet to be determined. At first approximation, these two responses to oculomotor error appear to be mutually exclusive (Engbert & Kliegl, 2011), and this appears likely to pose a significant modelling challenge.

Simulations with E-Z Reader have, however, suggested that it is compatible with advanced lexical processing of an upcoming word ($n+1$), including the extraction of its semantic properties, and the extraction of orthographic properties from word $n+2$ (Schotter, Reichle & Rayner, 2014). As attention shifts and eye movements are decoupled in this model, so called *semantic preview benefit* (e.g., Hohenstein, Laubrock & Kliegl, 2010; Schotter, 2013; Veldre & Andrews, 2016) and *word $n+2$ preview benefit* effects (e.g., Angele & Rayner, 2011) can arise when lexical identification of words n and $n+1$ complete in fast succession, allowing (covert)

attention to shift to the following word before an eye movement (with associated saccadic latency) catches up. It should be noted, however, that when the materials used in one of these simulations were similar to those used in an experiment that did report a semantic preview effect (Schotter, 2013), word $n+1$ only reached advanced lexical processing on 8% of occasions.

Proponents of the serial perspective have, by these means, provided at least partial explanations for what was initially interpreted as parallel processing effects, but the degree to which these explanations hold across material sets, and indeed whether models such as E-Z Reader can adequately be adapted to successfully account for these effects, is yet to be determined.

Nevertheless, with such alignment in the effects deemed compatible with both serial and parallel models it becomes increasingly difficult to distinguish them on the basis of eye movement data. But we can still return to the original question: is it psychologically plausible to assume that two words can be lexically processed in a parallel fashion? Reichle, Liversedge, Pollatsek and Rayner (2009) stated “we know of no compelling empirical evidence supporting the claim that two or more words can be simultaneously processed” (pp116). In contrast, Snell and Grainger (2019a) recently suggested there is emerging evidence from other paradigms that is more profitably viewed from a parallel than from a serial processing perspective. While Schotter and Payne (2019) were not convinced by the evidence, they do make a case for synthesising data across experimental paradigms when addressing the question. In that vein, this study investigates whether evidence compatible with parallel processing can be obtained using the same-different matching task.

In same-different matching participants are typically asked to decide whether two letter strings are visually identical or not. Despite requiring a match based only on physical similarity, many studies have shown that “same” RTs, at least, are influenced by *orthography* (legal nonwords faster than illegal), *lexicality* (words matched faster than legal nonwords), and *frequency* (high frequency words matched faster than low frequency). With “different” response times not always showing these effects, Chambers and Forster (1975) proposed an account involving a multi-level ‘race’ model in which comparisons between the two letter strings proceed simultaneously at letter, letter cluster and lexical levels of analysis. Whichever analysis returns an answer first wins the ‘race’ and determines RT. There is a tradeoff between increased processing time for higher-levels of analysis, but fewer comparisons required, and, assuming that the comparison process is relatively time-consuming, higher level analysis will often win the race – at least for “same” decisions, which require full stimulus processing. The model has been extended and successfully applied to sentence matching (e.g. Murray, 1982; Clahsen, Hong & Sonnenstuhl-Henning, 1995) and shown to be a useful tool for investigating early lexical processes (e.g. Murray & McKague, 2011). In general, of critical relevance here, it is clear that the speed advantage from the comparison of higher level representations results in this task providing a seamless unobtrusive way to reflect early ‘automatic’ word identification processes. Matching time reflects these processes not because the task demands it, since matching could equally be based on letter or featural identity, but simply because the comparison process is fastest that way.

It appears, therefore, that this task provides an ideal vehicle for investigating parallel lexical processing. Normally, of course, participants are free to move their eyes while making a matching decision, but will lexical effects still occur if presentation duration prevents direct

fixation of both words? If so, this suggests the words are likely to have been processed in an overlapping fashion. Varying word length also allows an insight into whether ease of processing influences the likelihood of demonstrating lexical effects. While two short words might reasonably be processed sequentially within a single fixation, this becomes less likely as word length is increased.

To simulate the way words appear in text, participants were presented with horizontally aligned word pairs that were always separated by a single character space. Pairs comprised either 4- or 6-letter words, of high or low frequency, or legal or illegal nonwords. Pairs were either identical to one another or differed by just one character.

Presentation times either allowed the potential fixation of both words, with an eye movement (514ms), or prevented one (216ms). The short presentation duration was based on the commonly-reported average fixation duration during reading of 225ms to 250ms (Rayner, 2009) and the eye-to-brain lag of approximately 50ms (Foxye & Simpson, 2002). The shortest average fixation duration combined with the time require to propagate information to the visual cortex should be around 275ms - 20% longer than this duration.

In order to ensure that participants were initially fixating the centre of the screen and not directing overt or covert attention elsewhere, one word was always presented aligned with a fixation mark at the centre of the screen, while the peripheral word was presented either to its left or right with equal frequency. This procedure also has the advantage of preventing anticipatory saccades, which in turn should increase saccadic latency. Additionally, it allows assessment of the extent and direction of the perceptual span in this task. In natural reading in English this has

been shown to extend further to the right than to the left (15 vs 3-4 characters; McConkie and Rayner, 1976).

Differences in matching time between legal and illegal nonwords implicate the engagement of sub-lexical orthographic processing; differences between legal nonwords and words suggest that at least some aspect of lexical processing was involved; while differences in matching time with word frequency suggest full lexical identification has taken place, with both words processed in parallel.

Method

Participants

Sixty-four native English speakers with normal or corrected to normal vision were tested for course credit or £5 payment.

Materials and Design

Four item conditions were tested with an equal number in each: high frequency words (mean: 232.45/million; SD: 170.66), low frequency (mean 5.60/million, SD 3.33), orthographically legal nonwords, and illegal nonwords (random letter strings). The 128 exemplars in each category comprised an equal number of 4- and 6-letter items and were tested both as identical pairs (requiring a “same” decision) or matched with a comparison item that differed by one character (requiring a “different” response). This character occurred with roughly equal frequency in each letter position. To prevent discrimination between “same” and “different” items based on word status, 50% of the differing items also formed words. Each item pair was displayed for either a

short (216ms) or long (514ms) duration and was aligned either to the left or right of the point of fixation. This resulted in an eight file counterbalanced design, with position, duration and response type for each item rotated across the item files.

Procedure

Items were presented on a VDU screen using DmDx (Forster & Forster, 1997). They appeared in lower case white 'Courier New' text on a black background. There were around 2.25 characters per degree at the average viewing distance of 550mm. Responses were recorded using left- and right-hand buttons on an attached button box.

Participants were told that two horizontally aligned letter strings would be displayed on the screen in front of them and these might or might not form words. It was made clear that word length, word position and presentation duration would all vary from trial to trial. However their task was simply to decide as rapidly and as accurately as possible whether the two items were physically identical or not. They were informed that if a pair differed, it would be by just one letter. A practice session with eight item pairs preceded the 512 experimental items.

Commencing each trial, a fixation cross appeared in the centre of the screen. After 1.7 seconds, the cross disappeared and the two letter strings were displayed. The central letter string was horizontally and vertically aligned with the fixation cross; the peripheral word appeared either to its left or right separated by a single character space. In the 'different' condition, the central string was always the base word or nonword, with a differing letter string in the periphery. Participants responded by pressing either a *same* (right) or *different* (left) button. After responding, they pressed a request button to trigger the next trial.

Results

As a result of the brief presentation and side-by-side display, error rates were understandably high, but well below chance (50%) in the vast majority of cases. However, to ensure that participants were treating the task seriously, rather than guessing, any with overall errors of 34% or higher were replaced. This occurred on only three occasions. RT cutoffs were set at 2.5 standard deviations from mean RT, calculated individually for each participant. Any RT falling outside this range was replaced by the relevant cutoff value. Cutoffs were applied to only 1.8% of the data.

Response times and error rates were analysed for 'same' and 'different' responses separately. In each case, a 4x2x2x2 ANOVA was carried out, treating subjects (F1) and items (F2) as random variables and including file as a between-groups dummy variable. The four factors were (a) item type (high and low frequency words; legal and illegal nonwords), (b) word length (4- vs. 6-letters), (c) presentation duration (short vs. long) and (d) presentation position (left vs. right).

An effect of item type was followed up with planned contrasts to determine whether it was driven by an effect of frequency (high vs. low frequency words), lexicality (low frequency words vs. legal nonwords), or orthography (legal vs. illegal nonwords), or a combination thereof. The Greenhouse-Geisser correction is reported where there was any violation of sphericity.

'Different' Responses

Mean response times and error rates for 'different' items are presented in Figure 1. An effect of item type was apparent in both RTs ($F_1(1.64, 91.54) = 11.50, MSE = 10822.07, p < .001$; $F_2(3,$

448) = 7.38, MSE = 9713.70, $p < .001$) and errors ($F_1(1.99, 111.57) = 93.68$, MSE = 482.84, $p < .001$; $F_2(3, 448) = 36.01$, MSE = 834.59, $p < .001$). The RT analysis shows that while there was little variation between the high and low frequency words and legal nonwords (all $F_s < 1$), there was a clear orthographic effect, with legal nonwords responded to faster than illegal nonwords ($F_1(1, 56) = 19.48$, MSE = 7965.71, $p < .001$; $F_2(1, 224) = 14.38$, MSE = 11616.46, $p < .001$). This pattern in 'different' responses is similar to that reported by Chambers and Forster (1975, experiment 2), with an orthographic but not a lexical effect when word pairs differed by a single character.

(Figure 1 about here)

Error rates showed an effect of frequency by-subjects ($F_1(1, 56) = 6.27$, MSE = 173.50, $p < .05$) but not by-items ($F_2(1, 224) = 1.75$, MSE = 633.60, $p = .19$), and an effect of lexicality by-subjects ($F_1(1, 56) = 6.97$, MSE = 144.76, $p < .05$) but not by-items ($F_2(1, 224) = 1.66$, MSE = 613.45, $p = .20$), suggesting this trend was driven by a response bias where participants were more likely to incorrectly respond 'same' to more familiar or 'word-like' letter strings (see also the reverse trend in 'same' items below). A clear effect of orthography was, however, present, with double the error rate for illegal than for legal letter strings ($F_1(1, 56) = 200.00$, MSE = 369.12, $p < .001$; $F_2(1, 224) = 71.26$, MSE = 1035.58, $p < .001$). We will defer consideration of how a Race Model might account for these effects until the Discussion.

The remainder of this section focuses on 'same' responses, where lexical effects would be anticipated.

'Same' Responses

Effects of Item Type: As can be seen in Figure 2, both mean RT and error rate clearly show effects of item type ($F(1, 110.44) = 209.91, MSE = 9211.97, p < .001$; $F(3, 448) = 215.18, MSE = 6421.52, p < .001$; and $F(1, 102.42) = 115.33, MSE = 613.96, p < .001$; $F(3, 448) = 136.19, MSE = 316.91, p < .001$, respectively). Pairwise comparisons show that high frequency words were responded to faster ($F(1, 56) = 76.57, MSE = 2205.42, p < .001$; $F(1, 224) = 37.17, MSE = 4686.97, p < .001$) and received fewer errors ($F(1, 56) = 44.85, MSE = 150.02, p < .001$; $F(1, 224) = 40.62, MSE = 165.90, p < .001$) than low frequency. These, in turn, received faster responses ($F(1, 56) = 174.45, MSE = 4096.26, p < .001$; $F(1, 224) = 109.45, MSE = 7216.42, p < .001$) and fewer errors ($F(1, 56) = 175.60, MSE = 159.98, p < .001$; $F(1, 224) = 79.22, MSE = 352.58, p < .001$) than legal nonwords. Finally, legal nonwords were responded to faster ($F(1, 56) = 48.30, MSE = 5236.32, p < .001$; $F(1, 224) = 33.68, MSE = 8156.08, p < .001$) and with fewer errors ($F(1, 56) = 10.53, MSE = 457.74, p < .01$; $F(1, 224) = 10.27, MSE = 467.92, p < .01$) than illegal nonwords. Unlike ‘different’ items, it is clear that effects of frequency, lexicality and orthography are readily apparent in ‘same’ responses.

(Figure 2 about here)

Since the central word was identical for ‘same’ and ‘different’ items (with a change of the peripheral word for ‘different’), the effects of frequency and lexicality cannot be attributed to foveal processing alone, or the same pattern should be seen with ‘different’ items. It is therefore clear that characteristics of the parafoveal word must have contributed to the pattern of effects. We can determine whether this implicates parallel lexical processing by examining how item type interacts with presentation position, word length and presentation duration.

Effects of Position: Right-sided word pairs were responded to significantly faster (651ms vs. 671ms; $F_1(1, 56) = 32.37$, $MSE = 6817.10$, $p < .001$; $F_2(1, 448) = 56.82$, $MSE = 3704.71$, $p < .001$) and with fewer errors (17% vs. 20%; $F_1(1, 56) = 15.39$, $MSE = 270.80$, $p < .001$; $F_2(1, 448) = 26.42$, $MSE = 157.57$, $p < .001$) than left. As Figure 3 indicates, position and item type did not interact in the error rate analysis ($F_1(3, 168) = 1.66$, $MSE = 124.76$, $p = .18$; $F_2(3, 448) = 1.32$, $MSE = 157.72$, $p = .27$), but they did in RT ($F_1(2.66, 148.77) = 3.06$, $MSE = 3751.12$; $p < .05$; $F_2(3, 448) = 3.75$, $MSE = 3704.71$; $p < .05$). Follow-up analyses revealed that this RT interaction was not driven by position interacting with either lexicality (both $F_s < 1$) or orthography ($F_1(1, 56) = 1.59$, $MSE = 4702.50$; $p = .21$; $F_2(1, 224) = 1.53$, $MSE = 4634.21$; $p = .22$). Rather, it was an interaction between position and frequency ($F_1(1, 56) = 4.95$, $MSE = 2180.76$; $p < .05$; $F_2(1, 224) = 3.90$, $MSE = 2775.20$; $p = .05$), with a larger frequency effect when the peripheral word appeared on the right, although the frequency effect was reliable on both the left ($F_1(1, 56) = 22.24$, $MSE = 2118.89$; $p < .001$; $F_2(1, 224) = 12.97$, $MSE = 3786.41$; $p < .001$) and the right ($F_1(1, 56) = 58.46$, $MSE = 2267.29$; $p < .001$; $F_2(1, 224) = 36.98$, $MSE = 3675.76$; $p < .001$).

(Figure 3 about here)

The results here suggests that same-different matching engages a similar asymmetrical distribution of attention as found in natural reading (e.g., McConkie & Rayner, 1976), with this and a variety of other tasks all showing a word identification advantage with scanning to the right (e.g., Rayner, 1978; Balota & Rayner, 1983; Simola, Holmqvist & Lindgren, 2009).

Effects of Word Length: There was a main effect of word length, with 4-letter responded to faster than 6-letter words (641ms vs 681ms: $F_1(1, 56) = 160.07$, $MSE = 5125.78$; $p < .001$; F_2

(1, 448) = 129.66, MSE = 6421.52; $p < .001$) and generating fewer errors (16% vs 22%: F1 (1, 56) = 64.00, MSE = 284.28; $p < .001$; F2 (1, 448) = 57.43, MSE = 316.91; $p < .001$). As can be seen in Figure 4, word length did not affect the item type effect on RTs (F1 (3, 168) = 1.20, MSE = 3830.22; $p = .31$; F2 (3, 448) = 1.41, MSE = 6421.52; $p = .24$), but there was a marginally significant interaction in error rate by-subjects (F1 (3, 168) = 2.69, MSE = 181.95; $p = .05$; F2 (3, 448) = 1.54, MSE = 316.91; $p = .20$). The frequency effect persisted for both 4- (F1 (1, 56) = 32.08, MSE = 58.09; $p < .001$; F2 (1, 112) = 18.91, MSE = 98.56; $p < .001$) and 6-letter words (F1 (1, 56) = 26.13, MSE = 203.08; $p < .001$; F2 (1, 112) = 22.80, MSE = 233.24; $p < .001$), but there appears to be a ceiling effect on error rate in nonwords. The orthographic effect was significant for 4-letter words (F1 (1, 56) = 20.73, MSE = 247.81 $p < .001$; F2 (1, 112) = 10.18, MSE = 502.14 $p < .01$) but not for 6-letter words (F1 (1, 56) = 1.70, MSE = 412.95, $p = .20$; F2 (1, 112) = 1.62, MSE = 433.70, $p = .21$). Thus, while the effect size for frequency was slightly attenuated for the shorter words, suggesting that the 6% error rate was the floor error rate for 4-letter high frequency words, it appears that lexical representations were accessed and utilised regardless of word length.

(Figure 4 about here)

Effects of Duration: As can be seen in Figure 5, there was a speed-accuracy trade-off in which short presentation durations resulted in faster responses (650ms vs. 672ms; F1 (1, 56) = 18.93, MSE = 13708.84, $p < .001$; F2 (1, 448) = 68.44, MSE = 3870.89, $p < .001$) but were more error prone (23% vs. 14%; F1 (1, 56) = 97.68, MSE = 390.88, $p < .001$; F2 (1, 448) = 209.74, MSE = 181.97, $p < .001$).

(Figure 5 about here)

Duration interacted with item type for both RT ($F_1(3, 168) = 9.11, MSE = 3212.95, p < .001$; $F_2(3, 448) = 7.68, MSE = 3870.89, p < .001$) and error rate ($F_1(2.67, 149.41) = 7.20, MSE = 224.48, p < .001$; $F_2(3, 448) = 7.91, MSE = 181.97, p < .001$). Neither interaction was driven by duration influencing the word frequency effect ($F_1(1, 56) = 2.52, MSE = 2788.39, p = .12$; $F_2(1, 224) = 2.42, MSE = 2554.79, p = .12$; and $F_1(1, 56) = 1.88, MSE = 117.23, p = .18$; $F_2(1, 224) = 2.06, MSE = 107.82, p = .15$, respectively); with clear effects of frequency for both durations in RT (short duration: $F_1(1, 56) = 26.32, MSE = 2032.79, p < .001$; $F_2(1, 224) = 14.81, MSE = 3874.67, p < .001$; long duration: $F_1(1, 56) = 41.33, MSE = 2961.12, p < .001$; $F_2(1, 224) = 36.53, MSE = 3367.09, p < .001$) and error rate (short duration: $F_1(1, 56) = 27.43, MSE = 171.07, p < .001$; $F_2(1, 224) = 27.98, MSE = 168.13, p < .001$; long duration: $F_1(1, 56) = 23.47, MSE = 96.17, p < .001$; $F_2(1, 224) = 21.38, MSE = 105.59, p < .001$), clearly suggesting that even at shorter durations matching was performed at a lexical level.

As is apparent in Figure 5A, the interaction in RT was primarily driven by an attenuated effect of orthography when presentation duration was short ($F_2(1, 56) = 7.29, MSE = 3992.37, p < .01$; $F_2(1, 224) = 7.54, MSE = 5186.99, p < .01$), although the effect remained significant for both short ($F_1(1, 56) = 15.15, MSE = 3645.64, p < .001$; $F_2(1, 224) = 8.54, MSE = 6238.23, p < .01$) and long durations ($F_1(1, 56) = 40.62, MSE = 5583.05, p < .001$; $F_2(1, 224) = 36.68, MSE = 7104.83, p < .001$). It appears that when the most time-consuming process is necessary to complete the task - letter-by-letter comparisons – extended presentation time is utilised when it is available.

The interaction between item type and duration in error rate was principally driven by an interaction between lexicality and presentation duration ($F_1(1, 56) = 5.53, MSE = 222.62, p <$

.05; $F_2(1, 224) = 6.40$, $MSE = 192.40$, $p < .05$); strong effects of lexicality were however present for both short ($F_1(1, 56) = 83.63$, $MSE = 244.53$, $p < .001$; $F_2(1, 224) = 60.37$, $MSE = 338.72$, $p < .001$) and long durations ($F_1(1, 56) = 63.13$, $MSE = 138.07$, $p < .001$; $F_2(1, 224) = 42.26$, $MSE = 206.26$, $p < .001$). Thus the pattern reported here suggests the ability to match at a lexical level has provided an advantage when presentation duration was short.

Since the frequency, lexicality and orthographic effects remain significant for both presentation durations, it appears that even when a second fixation was precluded, information from both words was still being utilised. Critically, the lexical effects are undiminished at the shorter presentation duration. In addition, the clear absence of three-way interactions between item type, word length and presentation duration for both RT ($F_1(3, 168) = .60$, $MSE = 3542.70$, $p = .62$; $F_2(3, 448) = 2.11$, $MSE = 3870.89$, $p = .10$) and errors ($F_1(3, 168) = .30$, $MSE = 209.66$, $p = .83$; $F_2(3, 448) = .35$, $MSE = 181.97$, $p = .79$), suggests a generality of the effects across display duration and word length.

Discussion

This study aimed to investigate whether parallel lexical processing is possible. In the task employed here, it is clear that the comparisons were based on automatically-derived higher-level lexical representations. Since these effects were apparent with presentation durations designed to prevent an effective second fixation taking the eye to the parafoveal word, this suggests that the two words were indeed processed lexically in an overlapping fashion.

Taken as a whole, the results are relatively straightforward: Despite the very brief display and use of the side-by-side format, the pattern for the ‘same’ decisions replicates that found in

classic same-different word matching studies. Like earlier studies (e.g., Chambers & Forster, 1975), there was clear evidence for a frequency effect, a lexicality effect, and an effect of orthography. As outlined above, Chambers and Forster (1975) proposed a ‘race’ model to account for this pattern of effects. This suggests that the comparison progresses at three levels of analysis simultaneously: the letter, letter-cluster and lexical levels, and whichever completes fastest wins the ‘race’ to determine the response. If a comparison can be made at a lexical level, this will likely complete fastest because, despite that representation taking longest to derive, only one comparison will be required, and this appears to be reflected in the expressions of effects of frequency and lexicality in the present set of data. Equally, since there are fewer letter-cluster than letter comparisons, this model also explains the effects of orthography seen here.

In line with Chambers and Forster (1975), it appears that ‘different’ decisions were not usually made on the basis of higher level lexical representations and it appears feasible to attribute this, as they did, to differences in the level of representation most likely to produce a ‘no’ decision most quickly. The race model accounts for this pattern of effects with a trade-off between the time required to extract a representation at each level of analysis, and the number of comparisons required to find the difference. If a mismatch is present, fewer comparisons will be required at lower levels of analysis, assuming the process operates on a self-terminating basis. This, combined with the reduced time required to obtain the representations at these levels, is likely to prevent the lexical level from winning the ‘race’, resulting in an absence of lexical effects. The presence of an orthographic effect for ‘different’ decisions implies that the trade-off, nonetheless, still favours the letter cluster level, presumably because the time required to identify letter clusters and make the comparisons is still faster than performing multiple letter-level comparisons until the mismatch is found.

The differences between ‘same’ and ‘different’ item types were anticipated; indeed, they nicely replicate those obtained by Chambers and Forster (1975), and they fit well with the conceptions of the race model. But since the ‘different’ items do not exhibit lexical effects, and the focus here is on the possibility of overlapping lexical processes, the remainder of this discussion will focus on the ‘same’ items, in which robust lexical effects were clearly observed.

This study differed from classic same-different word matching studies in that the words were horizontally and not vertically aligned. This adjustment allowed the two words to be presented in a format akin to natural reading. The same patterns of apparently automatic higher-level effects were, however, obtained, with clear effects of frequency, lexicality and orthography all surviving the change in format. It is also apparent that the horizontal alignment encouraged a similar asymmetrical span of attention to that found in natural reading (e.g., McConkie & Rayner, 1976). This replicates the findings from other isolated word and word pair identification tasks (e.g., Rayner, 1978; Balota & Rayner, 1983; Simola et al, 2009) with a processing advantage associated with scanning parafoveal words to the right. Thus, it appears that many elements of reading were automatically engaged with the task, making it an ideal vehicle to determine whether two words can in fact be lexically processed in parallel.

The results of this study, where parafoveal processing was clearly encouraged, also align well with recent research by Snell and colleagues, who provided evidence that simultaneous lexical processing advances on parafoveal words even when those words are task-irrelevant. In one study, Snell, Declerck & Grainger (2018) engaged French-English bilinguals in a semantic categorization task (natural vs. artefactual), also using brief (170 ms) presentation durations. They found that participants were faster to respond when the central (English) word was flanked

by (non-cognate) translations of the word, compared with when it was flanked by an unrelated French word. Given the brief presentation durations, Snell et al (2018) argue that their results support the assumption that both foveal and parafoveal processing must have proceeded simultaneously (see also Snell, Meeter & Grainger, 2017 for similar conclusions drawn from a syntactic categorization task).

The only ‘natural reading’ study with some similarity to the one reported here was conducted by Inhoff, Starr and Schindler (2000). They included a word repetition condition (e.g., ‘mother’s mother’); they also had an associated condition (e.g., ‘mother’s father’) and an unassociated condition (e.g., ‘mother’s garden’). These words were embedded within sentences. They found that gaze duration was shorter on the foveal word when the parafoveal word was related or associated than when it was unassociated. It might therefore be suggested that word repetition here facilitated the occurrence of parallel lexical processing, as it appears to have done in the Inhoff et al study. Thus, there is evidence that the results reported here translate, at least loosely, to a more natural reading task. It must be acknowledged, however, that the word repetition employed here and perhaps also in Inhoff et al’s study is not typical of natural text, and it is not clear the extent to which parallel processing might depend on this. Thus, while no assumption can be made regarding the generalizability of these results across a variety of reading situations, they nevertheless provide evidence suggesting that it is feasible to consider the possibility of overlapping processing of multiple words at a lexical level. This appears to be the evidence that Reichle et al (2009) claimed was missing.

The present experiment cannot adjudicate between serial and parallel models of eye movement control during reading. These models were built to simulate data from the reading of

sentences and not to make predictions about tasks such as this one, where reading happens to be automatically engaged. Indeed, that was not the purpose of the study. Its aim was to determine whether it is psychologically plausible to propose a model in which two words can be lexically processed in an overlapping manner. It did not aim to show that such a process is necessarily engaged during natural reading – just that it *could* be. At the very least, this study has clearly demonstrated that two words can be lexically processed within the shortest fixation duration.

Naturally, there is an argument that participants might be employing task dependent strategies in studies that do not involve ‘natural reading’ (see e.g., Rayner, White, Kambe, Miller & Liversedge, 2003; Schotter & Payne, 2019). A frequently-cited example of how non-reading tasks fail to capture the processes underlying natural reading is the finding that the frequency effect disappears during a search task (Rayner & Raney, 1996). But while the automaticity of reading is not apparent in every task where a word acts as the primary stimulus (e.g. visual word search), this appears to be the exception rather than the rule. Same-different matching is clearly sensitive to higher level linguistic processes, not because of a task-dependent strategy, but simply because it appears to be fastest to make the decision on the basis of (automatically generated) higher level representations. Certainly, for same-different matching tasks, such as the one reported here, the patterns of results do closely mirror many of those from reading studies, suggesting that reading appears to be the default process across many tasks.

Additionally, all experimental tasks involving reading include a set of compromises in terms of validity. For example, participants may adjust their reading behaviour in eye tracking experiments of sentence reading due to task instructions (e.g., being told to ‘read for comprehension’ and answer comprehension questions) and the environment these tasks

necessitate (e.g., being monitored, biting on a bar, keeping head still, etc.). Thus, the utility of tasks such as this one should not be dismissed simply because participants are not required to ‘read sentences’. The generality of these sorts of results across a variety of reading situations does need to be tested, but then so does the generality of a whole host of ‘reading effects’ that might be potentiated or influenced by a range of situational or task demands (see also the arguments in Murray and Rowan, 1998, regarding the influence of task demands in reading studies). We are lucky to have a variety of paradigms, old and new, that can be invoked to tackle different aspects of such an undeniably complex process as attentional distribution during reading. (see Snell & Grainger, 2019a & 2019b). But without an independent definition of what constitutes “natural reading processes”, we will always run the risk that proponents of a particular position may define some tasks as ‘unnatural’ on the basis of them either showing or not showing a particular effect.

The use of newer technologies, such as the fMRI study by White, Palmer, Boynton and Yeatman (2019), might facilitate better titration of the point or points where the (early) broadly parallel processes contributing to word recognition appear to start taking on a serial perspective, but in the absence of a clear definition of exactly what constitutes ‘word recognition’ how can we necessarily say that this point is either ‘pre-’ or ‘post-access’? This problem, and the way in which definitions become highly theory-specific, was highlighted by Alan Garnham (1985) and it is perhaps fair to say that we have not made a lot of progress on its resolution in the intervening 35 years. Do we conclude that White et al’s serial effects in VWFA-2 represent a characteristic of the access/identification process or do they reflect the process of categorisation after the word has been identified?

But, these issues aside, can we necessarily assume that a duration of 216ms is short enough to preclude an overt shift of attention, with an eye movement to the peripheral word? Research by Bolata & Rayner (1983) certainly suggests so. They presented participants with a central string of letters (either a real word or string of Xs) together with a parafoveal word displayed either to its left or to its right, the task being to pronounce the parafoveal word. They found that saccadic latency to a parafoveal word was influenced by the lexical status of the foveal word, with longer latencies with real words (287ms) compared to nonwords (255ms). Thus even when a foveal word is unrelated to the task, saccadic latencies were far beyond the 216ms presentation duration used in the current study. When the task requires foveal processing, as it did here, saccadic latencies would be expected to further increase. It appears that the lexical representations of both words were simultaneously available for comparison within that brief fixation duration.

If serial lexical processing of two adjacent words, to the point of complete recognition, can routinely be completed within one brief fixation, and word recognition is the ‘engine’ driving eye movements (Reichle et al, 2003), then why is there not a far higher incidence of word skipping? It seems more parsimonious to suggest that there can be overlapping lexical processing of two adjacent words during a single fixation and this is capable of providing sufficiently-developed representations of both words to underpin the sorts of (lexically-based) identity and difference decisions made here. But that is not of course to say that full, on-going, syntactic and semantic sentence processing could necessarily be completed in that time, and consequently it might not be efficient for the reader to move on, past the subsequent word, any more often than we know that they do.

Open Practices

All data and materials created during this experiment are openly available from the University of Dundee Institutional Repository, Discovery at DOI: 10.15132/10000151. This experiment was not preregistered.

Disclosure of interest

The authors report no conflict of interest.

References

- Angele, B., & Rayner, K. (2011). Parafoveal processing of word $n + 2$ during reading: do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1210–20.
- Balota, D. A., & Rayner, K. (1983). Parafoveal visual information and semantic contextual constraints. *Journal of Experimental Psychology: Human Perception and Performance*, 9(5), 726–38.
- Chambers, S. M., & Forster, K. I. (1975). Evidence for lexical access in a simultaneous matching task. *Memory & Cognition*, 3(5), 549–59.
- Clahsen, Hong & Sonnenstuhl-Henning (1995). Grammatical constraints in syntactic processing: Sentence matching experiments in German. *The Linguistic Review*, 12(1), 5-34.

- Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology*, 61(8), 1239–49.
- Engbert R, Kliegl R. Parallel graded attention models of reading. In: Liversedge SP, Gilchrist ID, Everling S (Eds). *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press; 2011. p. 787–800.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813.
- Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116-124.
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. *Experimental Brain Research*, 142(1), 139–50.
- Henderson, L. A. (1974) Word superiority effect without orthographic assistance. *Quarterly Journal of Experimental Psychology*, 26, 301–311.
- Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1150.

- Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology*, 16(1-2), 104–127.
- Inhoff, A. W., Starr, M., & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception & Psychophysics*, 62(7), 1474–84.
- McConkie, G., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, 8, 365–368. McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17, 578–586.
- Murray, W.S. (1982). *Sentence matching: The influence of meaning and structure. Unpublished doctoral dissertation*. Monash University, Victoria, Australia.
- Murray, W. S., Fischer, M. H., & Tatler, B. W. (2013). Serial and parallel processes in eye movement control : Current controversies and future directions. *The Quarterly Journal of Experimental Psychology*, 66(3), 417–428.
- Murray, W.S. & McKague M. (2011). Word learning (and the lack of it). Paper presented to the 52nd annual meeting of the Psychonomics Society, Seattle. US, 2011, November.
- Murray, W. S., & Rowan, M. (1998). Early, Mandatory, Pragmatic Processing. *Journal of Psycholinguistic Research*, 27(1), 1–22.
- Rayner, K. (1978). Foveal and parafoveal cues in reading. In J. Requin (Ed.), *Foveal and parafoveal cues in reading* (pp. 149–162). Hillsdale: Erlbaum.

- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–506.
- Rayner, K., White, S.J., Kambe, G., Miller, B., & Liversedge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 213–234). Amsterdam: Elsevier.
- Reichle, E. D., & Drieghe, D. (2015). Using E-Z Reader to Examine the Consequences of Fixation-Location Measurement Error, 41(1), 262–270.
- Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, 13(3), 115–119.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences*, 26(4), 445–476.
- Schotter, E. R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language*, 69(4), 619–633.
- Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading. *Trends in cognitive sciences*, 23(10), 811-812.
- Schotter, E. R., Reichle, E. D., & Rayner, K. (2014). Rethinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N+ 2 preview effects. *Visual Cognition*, 22(3-4), 309-333.

- Simola, J., Holmqvist, K., & Lindgren, M. (2009). Right visual field advantage in parafoveal processing: evidence from eye-fixation-related potentials. *Brain and Language*, 111(2), 101–113.
- Snell, J., Declerck, M., & Grainger, J. (2018). Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition and Neuroscience*, 33(5), 563-574.
- Snell, J., & Grainger, J. (2019a). Readers are parallel processors. *Trends in cognitive sciences*. 23(7), 537-546.
- Snell, J., & Grainger, J. (2019b) Consciousness is not key in the serial versus parallel processing debate. *Trends in Cognitive Science*, 23(1), 814-815.
- Snell, J., Meeter, M., & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PloS one*, 12(3). 1–17. DOI: 10.1371/journal.pone.0173720
- Veldre, A., & Andrews, S. (2016). Semantic preview benefit in English: Individual differences in the extraction and use of parafoveal semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(6), 837.
- Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision research*, 41(25-26), 3513-3533.

White, A., Boynton, G., & Yeatman, J. (2019). You can't recognize two words simultaneously. *Trends in Cognitive Sciences*, 23(10), 812-814.

White, A., Palmer, J., Boynton, G., & Yeatman, J. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences*, 116(20), 10087–10096.

Footnotes

1. An analysis by Murray, Fischer and Tatler (2013) showed that the number of published papers referring to serial and parallel processing in the context of reading had risen steadily from just 8 per year in 1997 to over 100 per year in 2012.
2. Of course, such effects only provide explanatory value if they fall in the same direction as those that would otherwise have been produced in the targeted word.
3. Henderson (1974) also used horizontal alignment, but his items were restricted to orthographically illegal letter strings that were either meaningful or not meaningful (FBI vs IBF), for which he obtained a 'word' superiority effect for the former class of words.

Figure Captions

Figure 1: Mean (A) Response Time (ms) and (B) Error Rate (%) for “Different” Items as a Function of Item Type.

Figure 2: Mean (A) Response Time (ms) and (B) Error Rate (%) for “Same” Items as a Function of Item Type.

Figure 3: Mean (A) Response Time (ms) and (B) Error Rate (%) for “Same” Items for Each Presentation Position and as a Function of Item Type.

Figure 4: Mean (A) Response Time (ms) and (B) Error Rate (%) for “Same” Items for Each Word Length and as a Function of Item Type.

Figure 5: Mean (A) Response Time (ms) and (B) Error Rate (%) for “Same” Items for Each Presentation Duration and as a Function of Item Type.

Figure 1

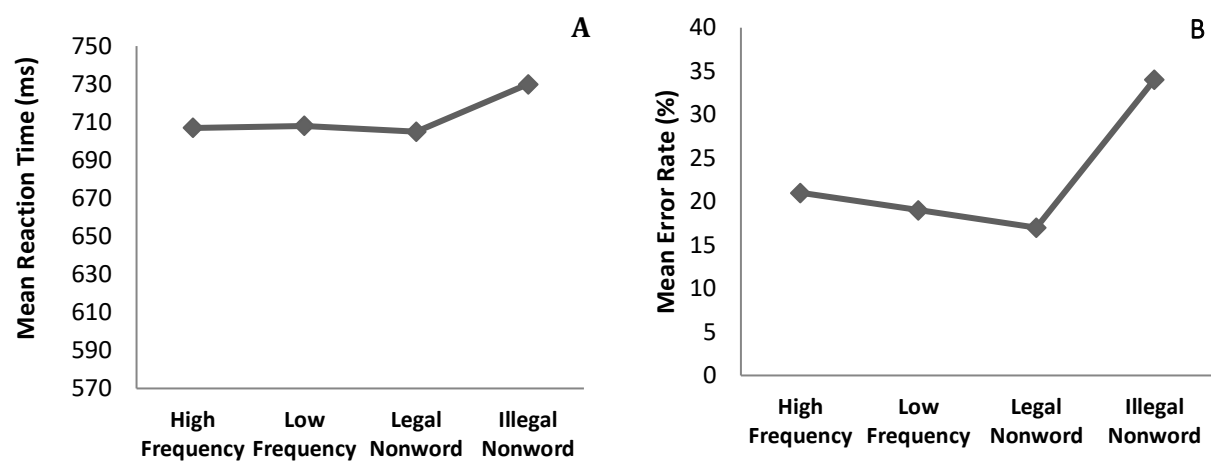


Figure 2

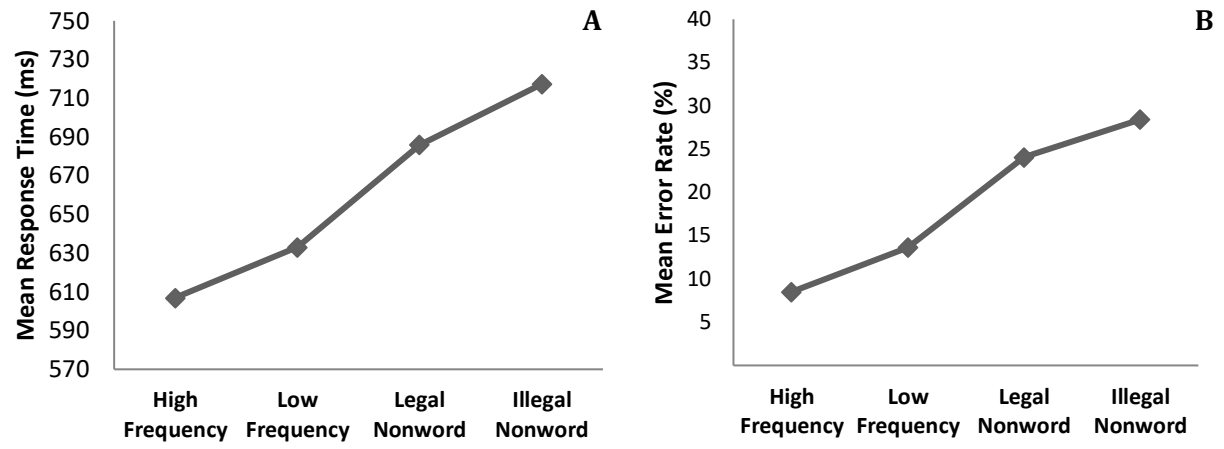


Figure 3

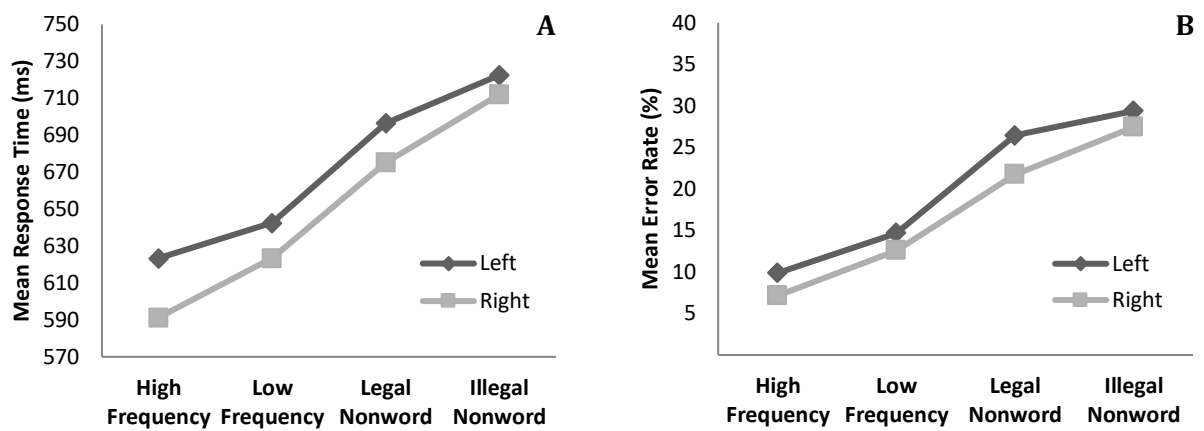


Figure 4

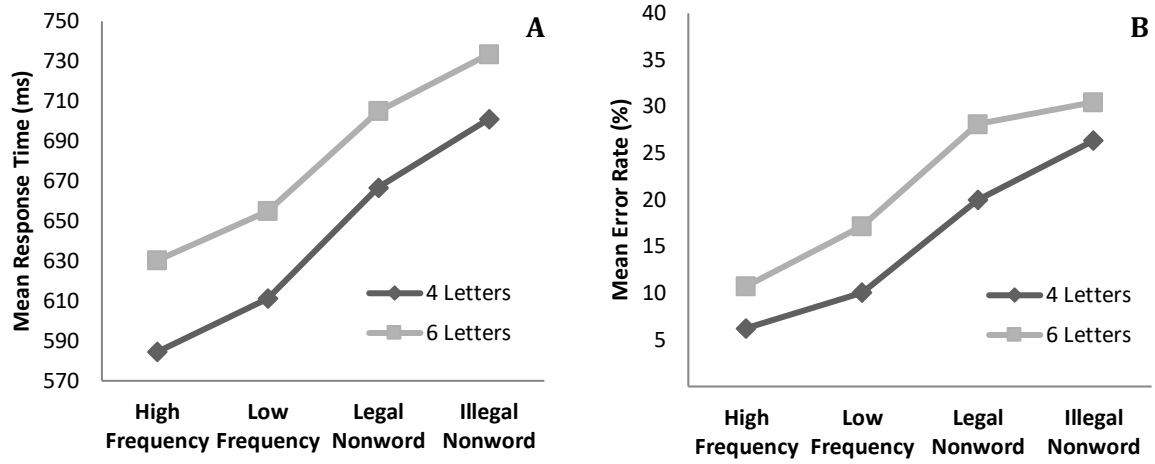


Figure 5

