Article

Examining Semantic Parafoveal-on-Foveal Effects Using a Stroop Boundary Paradigm

Zang, Chuanli, Zhang, Zhichao, Zhang, Manman, Degno, Federica and Liversedge, Simon Paul

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Examining semantic parafoveal-on-foveal effects using a Stroop boundary paradigm

Chuanli Zang [a,b,1,*], Zhichao Zhang [b], Manman Zhang [b], Federica Degno [c], Simon P. Liversedge [a,2]

[a] School of Psychology and Computer Science, University of Central Lancashire, United Kindom
[b] Key Research Base of Humanities and Social Sciences of the Ministry of Education, Faculty of Psychology, Tianjin Normal University, China
[c] Department of Psychology, Bournemouth University, United Kindom

ARTICLE INFO

ABSTRACT

The issue of whether lexical processing occurs serially or in parallel has been a central and contentious issue in respect of models of eye movement control in reading for well over a decade. A critical question in this regard concerns whether lexical parafoveal-on-foveal effects exist in reading. Because Chinese is an unspaced and densely packed language, readers may process parafoveal words to a greater extent than they do in spaced alphabetic languages. In two experiments using a novel Stroop boundary paradigm (Rayner, 1975), participants read sentences containing a single-character color-word whose preview was manipulated (identity or pseudocharacter, printed in black [no-color], or in a color congruent or incongruent with the character meaning). Two boundaries were used, one positioned two characters before the target and one immediately to the left of the target. The previews changed from black to color and then back to black as the eyes crossed the first and then the second boundary respectively. In Experiment 1 four color-words (red, green, yellow and blue) were used and in Experiment 2 only red and green color-words were used as targets. Both experiments showed very similar patterns such that reading times were increased for colored compared to no-color previews indicating a parafoveal visual interference effect. Most importantly, however, there were no robust interactive effects. Preview effects were comparable for congruent and incongruent color previews at the pretarget region when the data were combined from both experiments. These results favour serial processing accounts and indicate that even under very favourable experimental conditions, lexical semantic parafoveal-on-foveal effects are minimal.

It is well established that when a reader is fixating a word, they not only process that word itself, but also the word to its right, that is, the parafoveal word (Rayner, 1998, 2009). If useful information from the parafoveal word is available, then fixation durations on that word when it is subsequently fixated decrease significantly. This facilitation from preview of the parafoveal word is known as preview benefit (Rayner, 1975; Schotter, Angele, & Rayner, 2012 for a review). Typically, it is measured by the use of a boundary paradigm (Rayner, 1975). In this paradigm, in languages read from left to right, a preview string is initially presented at a target location, and an invisible boundary is positioned to the left of the target. Once the reader's eyes cross the boundary, the preview quickly changes to the target word. Generally, the reader is unaware of the change, because it occurs during a saccade

when vision is suppressed. Using this paradigm, a considerable amount of research has shown that readers access low-level visual and orthographic characteristics of the parafoveal word, however, there is some controversy concerning whether higher-level, semantic information about the parafoveal word is accessed and integrated with sentence context prior to its direct fixation. Further, in respect of our primary theoretical concern here, it is controversial as to whether such information from the parafoveal word is extracted very early during processing, that is, early enough to influence processing of the currently fixated, foveal, word. Such effects have been termed lexical parafoveal-on-foveal (PoF) effects, and studies demonstrating any such effects are generally assumed to show that two words may be processed in parallel during natural sentence reading (Reichle, Liversedge, Pollatsek, &

Rayner, 2009).

Strong empirical studies that investigate these questions are important in informing the current theoretical debate between serial versus parallel word identification during natural sentence reading. Currently, there are three influential computational models of eye movement control during reading of alphabetic languages: the E-Z Reader model (Reichle, 2011; Reichle, Pollatsek, Fisher, & Rayner, 1998), the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert & Kliegl, 2011; Schad & Engbert, 2012) and OB1-Reader (Snell, van Leipsig, Grainger, & Meeter, 2018a). More recently, the newly proposed Chinese Reading Model (CRM) of word segmentation, identification and eye movement control during reading of Chinese (Li & Pollatsek, 2020) has also been proposed.[3] E-Z Reader (e.g., Reichle et al., 1998) adopts a serial, sequential lexical identification approach and posits that attention is allocated in a strictly serial manner and shifts from one word to the next such that words are lexically processed sequentially and saccade targeting decisions are made on a word-by-word basis. According to E-Z Reader, parafoveal processing of the upcoming word occurs after the currently fixated word has been fully identified but before the eyes actually move to it, meaning that attention is allocated to the upcoming word for a brief time prior to its direct fixation and prior to its full identification. This implies that parafoveal processing is limited to low-level processing such as visual or orthographic rather than higher-level semantic processing, though, simulations of the E-Z Reader may possibly explain a modest-sized semantic preview effect that has been reported in the literature (Schotter, Reichle, & Rayner, 2014). Similarly, in the context of the serial processing framework, lexical processing of the upcoming word should not influence processing of the currently fixated word, that is, there should generally be no lexical PoF effect (Brothers, Hoversten, & Traxler, 2017; see Drieghe, 2011 for a review).

In contrast to the E-Z Reader model, the SWIFT model (e.g., Engbert et al., 2002) operates according to a parallel processing approach and posits that attention is spatially distributed as a gradient (typically distributed over 3–4 words). Within the gradient multiple words can be lexically identified non-sequentially and in parallel. This implies that lexical processing of the parafoveal word can occur even when the eyes are on the foveal word, and therefore, semantic preview effects and lexical influences of a parafoveal word on fixations on the foveal word should be observed. In other words, lexical PoF effects should be prevalent. Similar to the SWIFT model, the more recent OB1 Reader (Snell et al., 2018a) assumes that multiple words within an attentional visual input are lexically processed simultaneously. However, word identification is implemented according to an open-bigram coding mechanism, in which a word activates all the relative position open bigrams it contains, and inhibits other word units that share similar bigrams across the visual field. In addition, OB1 proposes a spatiotopic sentence-level representation that provides a reference frame for representing the position of words in a sentence, thereby providing a mechanism to explain how comprehension difficulty might not necessarily arise when words are identified out-of-order. According to OB1, a word is recognised when it can be mapped onto a plausible location in a sentence-level representation, on the basis of its approximate length indicated by spaces and expected syntactic structure. Semantic information can be accessed via lexical identification across multiple words in parallel, but those meanings are only integrated postlexically at the sentence level (Snell, Declerck, & Grainger, 2018b). Presumably, however, effects associated with an individual word's meaning, as opposed to effects arising through the integration of different words across (portions of) a sentence, should presumably appear in the eye movement record at the point in time that

the word's semantic meaning is accessed. Overall, then, it seems that even though OB1 provides an account of processing under the parallel gradient framework, it predicts an absence of semantic PoF effects that derive from combined lexical influences across words. Interestingly, Snell et al. (2018a, see p.981) specify that OB1 does predict reversed lexical successor effects such that a highly frequent word $n + 1$ inhibits the word node associated with word n, thus leading to increased fixation times on word n. It is not clear to us why, under the stipulations of OB1, the lexical characteristic of frequency does exert a PoF influence across words, but lexical semantic characteristics do not. All of this said, regardless of whether a positive or reversed lexical (frequency based) PoF effect is predicted, parallel processing models theoretically stipulate that such effects should occur.

The question of whether lexical PoF effects occur during reading of alphabetic languages is contentious. Empirical evidence for lexical PoF effects has been mainly reported in corpus studies where participants are required to read multiple sentences or longer passages, and fixation durations on the fixated word n are shown to be influenced by lexical properties (e.g., frequency and/or predictability) of the upcoming word $n + 1$ (e.g., Angele et al., 2015; Kennedy & Pynte, 2005; Kliegl, Nuthmann, & Engbert, 2006). Corpus studies of this type have shown such effects, though these have been inconsistent in their directions across studies (see Drieghe, 2011 for a review). In addition, in a clever study, Angele et al. (2015) reported reliable PoF effects of frequency and trigram predictability (conditional probability of an upcoming word occurring given the two preceding words), even when the parafoveal word $n + 1$ was masked and, therefore, visually unavailable prior to its direct fixation. In this situation any pre-processing of the target word itself was completely prevented, implying that PoF effects in this situation are not caused by the direct availability of information from the parafovea. Instead, it seems likely that the context preceding the target provided cues that facilitated anticipatory processing with respect to the upcoming word. Another possibility is that the target may have been cued by stored correlations that exist between adjacent words, or even lexical features of adjacent words. To be clear, these effects deriving from corpus studies may actually be successor effects. Successor effects are effects that occur at a word due to expectations about upcoming parafoveal words rather than being caused by the actual characteristics of the words themselves. To this extent, successor effects are not, strictly speaking, PoF effects (Kliegl et al., 2006). Indeed, when Angele et al.'s analyses were restricted to the target word and its frequency was manipulated experimentally, no effect of parafoveal $n + 1$ frequency was obtained on the preceding word n. Relatedly, Brothers et al. (2017) orthogonally manipulated frequency of the parafoveal word $n + 1$ such that it was high or low. Frequency is a lexical characteristic that consistently produces very robust experimental effects. Brothers et al. investigated whether reduced reading times were observed on the preceding word n when $n + 1$ was high relative to low frequency. No such pattern emerged in four experiments (with a total of 244 participants). A Bayesian meta-analysis showed a 4 to 5 ms non-significant effect in favour of the null hypothesis. Brothers et al. argued that their experiment represented a very strong test of whether lexical parafoveal-on-foveal effects occur during English reading and they concluded that such effects do not exist. It seems that lexical PoF effects are elusive in carefully controlled experimental studies.

Unlike English, the Chinese writing system is character based, unspaced, and densely packed with most words being comprised of one or two characters. Thus, upcoming parafoveal words are located relatively close to the point of fixation (or foveal vision) during reading. Second, written Chinese was developed from pictographs, and the connection between orthography and semantics is more transparent in Chinese compared to English. Also, up to 90 % of Chinese characters are comprised of semantic (and phonetic) radicals, conveying meaning information about the characters (Hoosain, 1991; Zang, Liversedge, Bai, & Yan, 2011). Third, word boundary ambiguity is prevalent in Chinese (He et al., 2021), which may require readers to allocate more attention to the

---

[3] The CRM adopts a number of the architectural features of the longstanding Interactive Activation Framework (McClelland & Rumelhart, 1981), but does not take a clear stance in respect of either a parallel or serial lexical processing approach. For this reason, we will defer consideration of this account until the General Discussion.

parafoveal characters for successful word segmentation and optimised saccadic targeting. All these characteristics may lead to aspects of semantic meaning being accessed efficiently from the parafovea (e.g., Yan, Richter, Shu, & Kliegl, 2009), which in turn might influence processing of the foveal word and result in possible lexical PoF effects for Chinese readers.

Given these characteristics, apparently, the Chinese writing system provides an excellent candidate language in which we might expect to observe lexical PoF effects. In a correlational study on Chinese reading, Li et al. (2014) showed PoF effects of predictability (or, potentially, successor predictability effects), such that increased predictability of word $n + 1$ resulted in shorter gaze durations on the preceding word n. Li et al. (2014) also found that word n was more likely to be fixated when word $n + 1$ was frequent compared with infrequent (an inverse PoF effect on fixation probability), but no PoF effects appeared in gaze durations on word n (i.e., they obtained a null PoF effect of frequency on fixation times). Again, the reasons for the inconsistent effects are not understood at present. There are additional studies that have reported results pertaining to this issue when analysing reading times on pretarget words in the context of research questions associated with parafoveal processing, and again, results are mixed. For example, Yan et al. (2009) found shorter gaze durations on pretarget words followed by semantically related than unrelated preview characters during Chinese reading (see also Pan, Laubrock, & Yan, 2016). Note that the related preview was more plausible based on the preceding context relative to the unrelated preview, and that when previews were not plausible, Yang, Wang, Tong and Rayner (2012) failed to replicate this effect. Also, Yan and Sommer (2015) reported an emotial PoF effect with increased fixation durations (rather than decreased) on the pretarget word when the target was emotionally positive relative to neutral. Again, it is hard to unambiguously attribute this effect to parafoveal influences per se (rather than successor effects) since these studies did not utilise the boundary paradigm. It seems reasonable to summarise the current state of the literature as being at a point where questions still remain with respect to the existence and nature of lexical PoF effects in Chinese (and other language) reading.

In the present study, given the criticality for distinguishing between serial accounts of reading and parallel accounts that clearly specify lexical PoF effects, we aimed to provide a very strong test of semantic PoF effects. For the reasons outlined earlier, therefore, we examined this issue in Chinese reading. Furthermore, we employed a novel Stroop boundary paradigm in which we combined the boundary paradigm (Rayner, 1975) with the classical Stroop paradigm (MacLeod, 1991; Stroop, 1935). It is well documented that in a Stroop task, reaction times are increased when the color in which a word appears is incongruent with the name of the word (e.g., the word green written in red). The Stroop effect is, arguably, the strongest experimental demonstration of an influence of a word's visual characteristic (its color) on semantic processing. In the Stroop boundary paradigm that we developed, participants were required to read sentences each including a color word (the target region) located roughly in the middle of the sentence. The preview of the color word was manipulated to be a pseudocharacter or an identity preview, printed in black (e.g., the word 绿 meaning *green* was printed in black 绿), a congruent color (e.g., 绿 meaning *green* was presented in green 绿) or an incongruent color (e.g., 绿 meaning *green* was presented in red 绿). In order to minimize any potential disruption that a non-uniform color character might cause, we included two invisible boundaries with one positioned two characters before the target word (immediately before the pretarget region) and the other one immediately to the left of the target. Prior to the participant making an eye movement to cross the first boundary, a black pseudocharacter was presented in place of the target character, thereby ensuring that during reading up to the first boundary, no visual information regarding the identity of the target character was available. Immediately after the eyes crossed the first boundary, the pseudocharacter preview of the target

was replaced by either an identity, or the same pseudocharacter preview, printed in black, a color congruent with the character meaning, or a color incongruent with the character meaning. Following this change, once the eyes crossed the second boundary, a second change occurred. In this case, the recently changed target preview was replaced by the identity target word in black (see Fig. 1 for an example stimulus). Thus, in both our experiments, participants always fixated words that were presented in black, with any color previews being exclusively presented prior to direct fixation. In addition to adopting very strong experimental manipulations, we were keen to ensure that our experiments were powerful enough to detect any effects associated with those manipulations. For this reason, we used large numbers of stimuli to collect large data sets (102 participants and 72 stimuli in Experiment 1, 96 participants and 72 stimuli in Experiment 2; we also undertook combined analyses for both expeirments in which there were 198 participants and 108 stimuli including only trials with red and green target color manipulations. This amounts to 1764 observations per condition which is significantly higher than the guidance by Brysbaert and Stevens, (2018; see also the power analyses section). These large sets of participants and stimuli enabled us to achieve substantial statistical power in our analyses which also maximized the possibility of obtaining lexical PoF effects.

Our predictions were straightforward: For the target word, we should obtain the Stroop effect for both identity and pseudocharacter previews. When the colored preview changes to the non-colored target and this is subsequently fixated, assuming that the color of the preview is perceived in the parafovea, then the incongruent color preview (relative to the congruent color preview) that remains in memory from the parafovea will produce a Stroop effect in relation to the meaning of the fixated non-colored target word. This will result in longer fixation durations under incongruent relative to congruent conditions. To be clear, a target word Stroop effect that is driven by the color of the preview should occur. Whilst this prediction is important, and such results would evidence the efficacy of our Stroop boundary paradigm, it is not particularly critical in respect of our primary theoretical concern, namely, discriminating between serial accounts and parallel accounts according to which semantic PoF effects should occur. To engage with this issue, it is important that we consider predictions for the pretarget region. Critically, for the pretarget region, there are two possible results. If words are lexically identified in parallel, and the semantic characteristics of upcoming parafoveal words are available prior to direct fixation, then a PoF Stroop effect should occur for identity word previews at the pretarget word, but such an effect should not occur for pseudocharacter previews. That is to say, on the pretarget region there should be shorter fixations and reduced reading times for congruent than incongruent target word previews. On the other hand, if words are lexically identified serially, then there should be no sensitivity to the semantic meaning of the target word during pretarget fixations, and therefore, there should be no pretarget Stroop effect.

Experiment 1

Method

Participants

One hundred and two students (Mean Age = 21 years, SD = 6; Male = 10, female = 92) from Tianjin Normal University, participated in the experiment for monetary compensation. They were all native Chinese speakers, had normal color vision and normal or corrected to normal vision, without any history of reading impairments. They were all naïve regarding the purpose of the experiment.

Apparatus

Participants' eye movements were recorded by an SR Research Eyelink 1000 system at a sampling rate of 1000 Hz. A 19-in DELL CRT monitor with a resolution of 1024 × 768 pixels and a refresh rate of 150

Target Preview	Color Preview	Sentence
Identity	Black	所有申请者的证件照片|均以|绿背景作为统一标准。
Identity	Congruent	所有申请者的证件照片|均以|绿背景作为统一标准。
Identity	Incongruent	所有申请者的证件照片|均以|绿背景作为统一标准。
Pseudocharacter	Black	所有申请者的证件照片|均以|桬背景作为统一标准。
Pseudocharacter	Congruent	所有申请者的证件照片|均以|桬背景作为统一标准。
Pseudocharacter	Incongruent	所有申请者的证件照片|均以|桬背景作为统一标准。

Fig. 1. An example set of sentences under the six conditions in the experiment. The vertical lines represent the positions of the two invisible boundaries. When the eyes fixated the characters before the first boundary, a pseudocharacter was presented in black at the target region. As readers' eyes crossed the first boundary, the identity or pseudocharacter preview was presented in black, a congruent or an incongruent color relative to the meaning of the target word. Once the eyes crossed the second boundary, the preview was replaced by the target word in black. The English translation for the sentence is "All applicants' ID photos must have a green background to meet the standard". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Hz, was used to display sentences. Stimuli were presented in Song font on a white background, with one Chinese character equal to approximately 1.1° of visual angle at the viewing distance of 65 cm.

Materials and design

Four Chinese single-character color-words (红 meaning *red*, 绿 meaning *green*, 黄 meaning *yellow* and 蓝 meaning *blue*) were selected as target words that could be presented in red (RGB coordinates = 255, 0, 0), green (RGB coordinates = 0, 255, 0), yellow (RGB coordinates = 255, 255, 0), or blue (RGB coordinates = 0, 0, 255). Each color word was embedded in the middle of 18 sentences that were 16–26 characters in length (*M* = 21 characters), and 72 experimental sentence frames were thus constructed. The sentence naturalness was rated on a 5 point scale by 19 participants who did not take part in the eye-tracking experiment, and the mean was 4.26 (*SD* = 0.32, 1 = very unnatural, 5 = very natural). The color words were unpredictable given the preceding sentence context, the mean predictability was 0.009 (*SD* = 0.03) rated by 17 participants who did not take part in the naturalness norming task or eye-tracking experiment.

We used a Stroop boundary paradigm (Rayner, 1975) and manipulated the preview of the color word to be either an identity preview or a pseudocharacter preview printed in black (no-color), a congruent color (e.g., the word 绿 meaning *green* was presented in green 绿) or an incongruent color (e.g., the word 绿 meaning *green* was presented in red 绿, yellow 绿 or blue 绿, and colors were counterbalanced in the experiment). As mentioned, two invisible boundaries were included within each sentence, with the first one positioned before the pretarget region that triggered the target word to change to a color (or remain black in the non-color condition). This manipulation reduced potential disruption that might occur due to the persistent presence of a colored word within the sentence. The second boundary triggered the target word display change, that is, replacement of the preview by the non-coloured target word (see Fig. 1). In order to increase the probability of participants making saccades to cross the first boundary and make a direct fixation on the pretarget region, the first boundary was positioned two characters prior to the pretarget boundary which appeared immediately before the target word. The two characters between the two boundaries always formed at least one word. The "pseudocharacters" were characters with extremely low frequency that appeared very rarely and were categorized as pseudocharacters in a prescreen test by 25 participants who did not take part in the formal eye-tracking study (see Zang et al., 2020). The number of strokes was matched between the identical (*M* = 10, *SD* = 3) and pseudocharacter previews (*M* = 10, *SD* = 2, *F* < 1).

The experimental design was a 2 (Target Preview: Identity vs Pseudocharacter) × 3 (Color Preview: Black, Congruent vs Incongruent) within-participant repeated measures design. We constructed six files with each file containing 72 experimental sentences (12 in each condition). Conditions were rotated across files according to a Latin square. There were eight practice sentences presented prior to the experimental sentences, and 144 filler sentences without any changes. Sentences were presented randomly and 34 % of the sentences were followed by a yes/no comprehension question. According to the related study by Yan and Sommer (2015) in which an emotional PoF effect was reported for fixation durations on pretarget words, the average Cohen's *dz* was 0.42. Based on Westfall (2015), the power of our sample size in Experiment 1 (102 participants and 72 sets of stimuli) is estimated to be 0.90, which is greater than the recommended level of 0.80.

Procedure

Each participant was tested individually in a quiet laboratory environment. They were required to read sentences carefully and try their best to understand them, then press a space key to terminate the sentence display and press the "F" (yes) or the "J" (no) key to answer a comprehension question where necessary. Prior to presenting the sentences, participants were required to complete a 3-point horizontal calibration procedure, the average error of each participant's calibration was below 0.20°. After each trial, a drift correction procedure was carried out to trigger the onset of the following sentence. The whole experiment lasted about 40 min. At the end of the experiment, participants were required to indicate whether they had perceived any color or character changes during reading of the sentences, and if so, to estimate the percentage of trials including changes that they had noticed.

Results and discussion

Participants' mean comprehension accuracy was 88 % suggesting that they read and fully understood the sentences. Also, 95 % of participants reported that they were aware of display changes. Among these, they estimated that on average there were 21 % trials (SD = 17 %) including character changes, 24 % (SD = 15 %) including color changes, and 11 % (SD = 13 %) including both changes. These estimates are roughly similar to our real changes (16 % character changes, 21 % color changes, and 11 % both changes, among all the sentences including filler sentences that participants read). Note that although participants were aware of these changes, the color of the target word/pseudocharacter string was only available when the eyes were fixating the pretarget region, and the black target word replaced the preview after they crossed the second boundary. Therefore, any disruption of normal reading was minimized.

Any fixations shorter than 80 ms or longer than 1200 ms were removed from the analyses. Trials were removed if a) a track loss occurred or there were fewer than three fixations in total (0.9 %); b) participants blinked while the display changed or the boundary changed (7.6 %); c) the display changes triggered late (13.8 %); and d) a saccade crossed the boundary (triggering the display change) but hooked back to

come to rest to the left of the boundary (9.1 %). Finally, for each measure any observations were removed that were three standard deviations from the mean of each participant (for the target word analyses: 1.0 %; for the pretarget region analyses: 1.9 %). The total percentage of trials excluded from the analyses was in line with other display change experiments (e.g., Kliegl, Risse, & Laubrock, 2007). This relatively large exclusion rates were probably due to the lack of spaces between words in Chinese and the fact that we used two boundaries in the study which might increase the rate of hooking (premature triggering of the boundary).

Fixation times and skipping probability on the target and the pretarget region were examined. Specifically, the following measures were computed: *first fixation dur*ation (FFD, the duration of the first fixation on a region), *single fixation duration (*SFD, the fixation duration on a region when it received only one fixation during first-pass reading), *gaze duration* (GD, the sum of all first pass fixations on a region), and *total fixation duration* (TFD, the sum of all fixations on a region).

Data were analyzed using linear mixed-effects models (LMM) with the lme4 package (version 1.1.21.; Bates, Maechler, Bolker, & Walker, 2015) in R (version 3.4.4; R Core Team, 2018). We treated Target Preview, Color Preview and their interactions as fixed factors, and for the color preview, we conducted the successive contrasts with comparisions of black vs congruent, and congruent vs incongruent color previews. In addition, we entered participants and items as crossed random factors. Models were run for each measure with the maximum random effects structure (Barr, Levy, Scheepers, & Tily, 2013), allowing both random intercepts and slopes for the previews over both participants and items. If the model failed to converge, it was further trimmed starting with removing correlations between factors, then interactions, then random factors for items then for participants until it converged. For fixation time analyses, data were log-transformed to increase the normality. Note that analyses of the untransformed fixation time measures produced similar patterns of effects to the transformed data, with some measures producing even more pronounced effects (see below for details). The regression coefficient (*b*, effect relative to the intercept, indicating effect size), standard error (*SE*), and *t* or *z* values are reported in these analyses (Baayen, 2008, Baayen, Davidson, & Bates, 2008). Means and standard deviations for the eye movement measures on the target word and the pretarget region are shown in Table 1, and the corresponding fixed effect estimations for these measures are shown in Table 2.

The target word

For all the eye movement measures, there was a reliable effect of target preview such that readers made longer fixations when they had a pseudocharacter preview rather than an identity preview (all *t* > 6.12), replicating the standard preview effect (Liversedge & Findlay, 2000; Rayner, 2009). Furthermore, relative to the no color (black) preview, readers made longer fixations when they received a congruent color preview (all *t* > 6.48). This reflects the change from one color to another (incongruent or congruent color to black) relative no-color change (black to black). Importantly, the difference between black and congruent color previews interacted with the target preview across all fixation time measures (all |*t*| > 2.43, see Fig. 2). The planned contrasts showed that this target word color preview effect was more robust when previews were identical to the target (FFD: *b* = 0.14, *SE* = 0.02, *t* = 6.11; SFD: *b* = 0.16, *SE* = 0.02, *t* = 6.49; GD: *b* = 0.17, *SE* = 0.03, *t* = 6.64; TFD: *b* = 0.15, *SE* = 0.03, *t* = 5.34) than when they were pseudocharacters (FFD: *b* = 0.07, *SE* = 0.02, *t* = 3.15; SFD: *b* = 0.10, *SE* = 0.03, *t* = 3.66; GD: *b* = 0.08, *SE* = 0.02, *t* = 3.30; TFD: *b* = 0.08, *SE* = 0.03, *t* = 3.13). This shows that readers gain access to the meaning of the target word more rapidly after an identity preview than a pseudocharacter preview, and this in turn influences the immediacy and magnitude of the colour preview effect that is seen at the target. The more rapidly available the target word meaning, then the more immediate and greater the magnitude of the colour preview effect. This effect is entirely consistent

with a basic preview benefit effect that then leads to a foveal word stroop effect which is more immediate and larger than when the preview benefit is limited due to the presence of the pseudocharacter.

The congruent color previews produced slightly shorter total fixation durations compared to the incongruent color previews (*t* = -1.89, *p* = 0.06),[4] indicating a numerical, but non-significant target word Stroop effect. It is important to note that this is a main effect of preview color that occurs at the target word. This means that regardless of whether the target preview did, or did not, deliver meaningful linguistic information, the color congruency effect occurred at the target. Given this, it is very likely that the effect arises because the reader perceives and processes the colour of the string in the target position in the parafovea, and then, upon fixating the target, which immediately changes to be presented in black and in its identity form, the congruency of the color that was perceived and processed earlier and the word meaning have a facilitatory influence. Critically, if the effect was linguistically mediated based on the preview, then an interactive effect between preview type and preview color should have occurred at the target. This did not occur (all |*t*| < 1.27).

Finally, there were no reliable differences between the two conditions in any other measures, nor any other interactions.[5]

The pretarget region

The two characters between the two boundaries were defined as the pretarget region. The effect of target preview did not achieve significance in FFD (*t* = -1.78, *p* = 0.08) and GD (*t* = -0.86), but was reliable in SFD (*t* = -2.10) and TFD (*t* = 4.86) with longer fixations on the pretarget region for identical compared to pseudocharacter target word previews in SFD but shorter pretarget TFD for identical than for pseudocharacter target previews. This SFD result seems to be an instance of a parafoveal attraction effect, whereby a parafoveal preview of a pseudocharacter is perceived as being visually unfamiliar, and this oddity attracts the point of fixation more rapidly than a normal Chinese character preview that is more visually familiar, causing the current fixation on the pretarget region to terminate earlier in the pseudocharacter than in the identity preview condition (see Hyönä, 1995; Hyönä & Bertram, 2004; Kennedy, 1998). Note also, though, that the opposite pattern occurred for the TFD in that readers retured to the pretarget region more often to spend more time processing that word overall. This is likely due to the fact that the eyes left the word prematurely during first pass (since the eyes were rapidly attracted to the upcoming pseudocharacter preview) and therefore needed to return to the pretarget region to ensure that it was processed fully. The main point to take from this particular set of results is that they demonstrate orthographic parafoveal-on-foveal influences during first pass reading, that is, a sensitivity to orthographic oddity in the parafovea. However, they offer no support for parallel lexical processing accounts that specify lexical PoF effects.

In addition, the target preview interacted with the difference between the congruent color and black previews in GD (*t* = -2.24, see Fig. 3). The planned contrast showed that for the identity preview, there was no difference between the congruent and black previews (*b* = 0.02, *SE* = 0.02, *t* = 1.09), but for the pseudocharacter preview, GD on the pretarget region was shorter for the congruent color preview than for the black preview (*b* = -0.04, *SE* = 0.02, *t* = -2.08). First, in relation to this

---

[4] Analyses from the raw data of total fixation duration showed a reliable difference between the congruent and incongruent color previews (*b* = 16.26, *SE* = 7.62, *t* = 2.13). Also the difference between the two color previews showed an interactive pattern with the target preview in first fixation duration (*b* = −15.54, *SE* = 9.16, *t* = −1.70, *p* = 0.09), however the planned contrasts showed no reliable effects of color preview for both word and pseudocharacter previews (all *t* < 1.33).

[5] We also conducted analyses for the target word exclusively for those trials in which the pretarget region was fixated (i.e., not skipped) and the pattern of effects was exactly the same as that reported here.

Table 1

Means and standard deviations for the eye movement measures on the target word and the pretarget region in Experiment 1.

Analysis region		Identity			Pseudocharacter		
		Black	Congruent	Incongruent	Black	Congruent	Incongruent
The target word	FFD (ms)	253 (53)	295 (61)	307 (78)	298 (68)	321 (65)	312 (60)
	SFD (ms)	254 (54)	300 (65)	313 (85)	299 (69)	331 (66)	320 (55)
	GD (ms)	262 (63)	314 (69)	328 (92)	326 (86)	351 (72)	350 (66)
	TFD (ms)	301 (86)	351 (99)	365 (105)	377 (126)	399 (114)	418 (136)
The pretarget region	FFD (ms)	220 (36)	222 (39)	225 (39)	220 (39)	215 (37)	222 (36)
	SFD (ms)	219 (38)	221 (39)	224 (39)	218 (40)	215 (40)	219 (36)
	GD (ms)	241 (59)	247 (75)	254 (79)	250 (70)	241 (720)	246 (66)
	TFD (ms)	332 (103)	330 (124)	340 (131)	360 (118)	356 (139)	360 (134)

Note: Standard deviations are provided in parentheses. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration.

Table 2

Fixed effects estimates for the eye movement measures on the target word and the pretarget region in Experiment 1.

Analysis region		FFD			SFD			GD			TFD		
		b	SE	t	b	SE	t	b	SE	t	b	SE	t
The target word	**Target Preview**												
	Pseudocharacter vs Identity	0.08	0.01	**6.13**	0.09	0.01	**7.17**	0.13	0.01	**9.20**	0.14	0.02	**7.80**
	Color Preview												
	Congruent vs Black	0.11	0.02	**6.94**	0.13	0.02	**7.91**	0.13	0.02	**7.58**	0.12	0.02	**6.49**
	Incongruent vs Congruent	0.00	0.02	0.17	0.00	0.02	0.27	0.02	0.02	1.00	0.03	0.02	<u>1.89</u>
	Interaction												
	Target Preview × Congruent vs Black	−0.09	0.03	**−2.96**	−0.09	0.03	**−2.61**	−0.12	0.03	**−3.53**	−0.09	0.04	**−2.44**
	Target Preview × Incongruent vs Congruent	−0.04	0.03	−1.26	−0.04	0.03	−1.21	−0.03	0.03	−0.83	0.01	0.04	0.32
The pretarget region	**Target Preview**												
	Pseudocharacter vs Identity	−0.01	0.01	<u>−1.78</u>	−0.02	0.01	**−2.10**	−0.01	0.01	−0.86	0.06	0.01	**4.86**
	Color Preview												
	Congruent vs Black	−0.00	0.01	−0.48	−0.00	0.01	−0.53	−0.01	0.01	−0.71	−0.02	0.02	−1.08
	Incongruent vs Congruent	0.02	0.01	**2.20**	0.02	0.01	<u>1.82</u>	0.02	0.01	1.63	0.02	0.01	1.51
	Interaction												
	Target Preview × Congruent vs Black	−0.01	0.02	−0.91	−0.04	0.02	−0.73	−0.05	0.02	**−2.24**	−0.01	0.03	−0.35
	Target Preview × Incongruent vs Congruent	0.01	0.02	0.32	0.00	0.02	0.21	0.00	0.02	0.03	−0.01	0.03	−0.17

Note. Significant terms are presented in bold, and terms approaching significance are underlined.
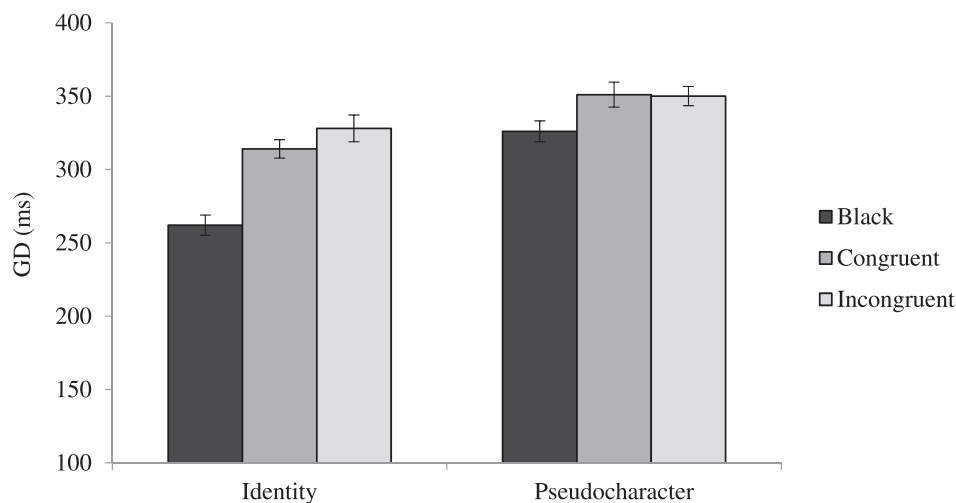

Fig. 2. Gaze duration as a function of target preview and color preview condtions on the target word.

effect, note this does not reflect any type of Stroop influence because the difference occurred for the pseudocharacter previews that carry no semantic information. It is likely that readers moved their eyes from the pretarget region onto the target word more rapidly when the target preview was colored compared with when it was black, especially when the preview was a pseudocharacter, because the briefly presented colored pseudocharacter preview captured attention. This would have resulted in shorter pretarget fixations in the congruent color compared to the black preview for the pseudocharacter previews.

Finally, the differences between the incongruent and congruent color previews were reliable in FFD ($t = 2.20$), and approached significance in SFD ($t = 1.82$, $p = 0.07$), such that readers spent more time when they had an incongruent color preview rather than a congruent color preview. It is not clear why this result has occurred given that for there to be
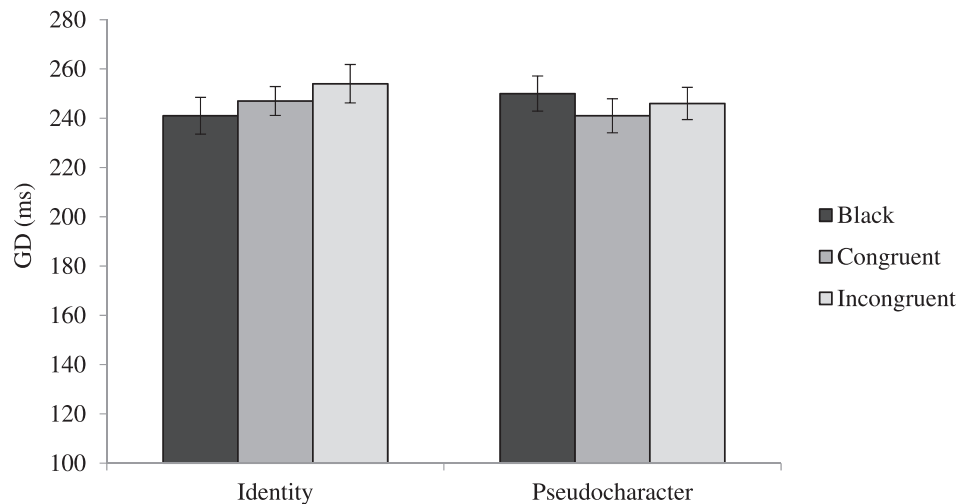
Fig. 3. Gaze duration as a function of target preview and color preview condtions on the pretarget region.

any meaningful influence of congruency at the pretarget region, then this must be interactive with preview type – that is, in order for congruency to have an effect, the identity preview must be available. None of the other interactions were reliable.

In summary, the results show a clear target preview effect, a color preview effect and a non-significant Stroop effect that was, numerically, of the type we predicted at the target word. Additionally, there was no suggestion of interactive effects at the pretarget region reflecting semantic PoF effects deriving from incongruency between target word color and meaning. The results demonstrate that our novel Stroop boundary paradigm was effective in inducing Stroop type effects. Further, the results provide no evidence for parallel accounts that clearly specify semantic PoF effects during reading, but are in line with serial processing accounts. All of these things said, it is also the case that the results from Experiment 1, overall, were not as robust in all the measures as we might have liked, and this may be slightly surprising given the large number of participants and stimuli that we used in the study. Furthermore, it is the case that within the results there were a small number of effects that were difficult to interpret (and to some extent, are difficult to consider meaningful). Given this, we carried out a further set of analyses in which we sought to examine the data sets with greater scrutiny to better understand how we might undertake further experimental work to deliver a more compelling and clearer set of results.

Additional analyses

As mentioned earlier, four colors were used in Experiment 1 as per the classical Stroop study. However, upon reflection, it was apparent that the colors of blue (蓝) and yellow (黄) may not have been as effective as parafoveal Stroop stimuli as we might have liked. On our CRT monitor the blue text appeared quite similar to the black text (thereby not delivering a strong color cue), and the yellow text was difficult to distinguish from the white background of the display meaning that its orthographic identity may not have been apparent. Both these colours, therefore, may have provided a relatively weak colour (in)congruency cue. If this was the case, then it might at least partially explain why we obtained some slightly odd and statistically weak effects, and more importantly, it may also explain why we observed no evidence of semantic PoF effects.

In order to investigate whether our concerns regarding the efficacy of our color stimuli were reasonable, and to focus on the portion of the full data set for which the chances of observing lexical PoF effects were maximal, we analysed a sub-set of our data. In these analyses, we considered only those trials where red and green previews were included, and only those trials in which the target word was fixated immediately after a fixation on the pretarget region. Overall, the results

from these analyses were very similar to those reported above, and for this reason, here we only report results that differ from those based on the full dataset (for the full set of results, please see Table S1-2 in the Supplemental Section).

For the target word analyses, the interaction between the congruent color vs black and the target preview (identity or pseudocharacter) was not reliable (all ($|t| < 1.63$). However, interestingly, there was a numerical pattern that was suggestive of an interaction between the color congruency of the preview (incongruent vs congruent) and the type of preview, though these effects did not achieve significance (FFD: $|t| = 1.67$, $p = 0.10$; SFD: $|t| = 1.72$, $p = 0.09$). The planned contrasts for the identity previews showed Stroop effects that approached, but did not achieve, significance. Incongruent color previews produced numerically longer first ($b = 0.07$, $SE = 0.04$, $t = 1.80$, $p = 0.07$) and single fixation durations ($b = 0.07$, $SE = 0.04$, $t = 1.70$, $p = 0.09$) compared to the congruent color previews. However, the counterpart congruency effects for the pseudocharacter previews, were absolutely minimal and this was reflected in the planned comparisons (all $|t| < 1$).[6]

For the pretarget region analyses, again, the basic patterns were very similar to those obtained for the full data set, though the effects were more stable, supporting our suggestion that the blue and yellow stimuli may have been less effective than the red and green (and note that this is the case despite significantly reduced power). Relative to the black preview, the congruent color target preview produced shorter fixations on the pretarget region, and these effects were statistically robust (all $|t| > 2.04$). This is proably another instance of attentional capture, due this time to the colored target preview. Clearly, there was no suggestion of any semantic PoF Stroop effect.

To reiterate, the additional analyses that we undertook were necessarily based on a subset of our data with the consequence of reduced statistical power. It might reasonably be suggested, therefore, that this is (again) the reason for the absence of semantic PoF effects at the

---

[6] Anlsyses from the raw data showed that the difference between the incongruent and congruent color previews interacted significantly with the target preview in FFD ($b = -35.90$, $SE = 17.00$, $t = -2.11$), SFD ($b = -37.42$, $SE = 17.61$, $t = -2.13$) and GD ($b = -42.24$, $SE = 21.43$, $t = -1.97$). The planned contrasts showed significantly longer fixations for incongruent than congruent color previews, but this color incongruency effect only occurred for identity previews (FFD: $b = 27.00$, $SE = 12.23$, $t = 2.21$; SFD: $b = 28.41$, $SE = 12.53$, $t = 2.29$; GD: $b = 40.17$, $SE = 15.62$, $t = 2.57$), not for pseudocharacter previews (FFD: $b = -11.47$, $SE = 11.60$, $t = -0.32$; SFD: $b = -12.32$, $SE = 12.60$, $t = -0.98$; GD: $b = -4.34$, $SE = 14.86$, $t = -0.29$). These patterns were reliably replicated in Experiment 2 where we exclusively used red and green previews to ensure a more effective Stroop manipulation.

pretarget region. Given this, we carried out Experiment 2 in which we used the same number of stimuli as we employed in Experiment 1, but for these stimuli we only used the more effective red and green preview color congruency manipulation.

Experiment 2

Method

Participants

Ninety-six students (Mean Age = 21 years, SD = 3; Male = 15, Female = 81) from Tianjin Normal University who did not take part in Experiment 1, participated in the experiment for monetary compensation. They were all native Chinese speakers, had normal color vision and normal or corrected to normal vision, without any history of reading impairments. They were all naïve regarding the purpose of the experiment.

Materials and design

As in Experiment 1, the same 72 experimental sentence frames were used, but the color words were only red and green. The sentences were all natural with the mean naturalness of 4.13 (SD = 0.28) and the color-words were unpredictable given the preceding context with the mean predictability of 0.004 (SD = 0.02), rated by 32 participants (16 in each norming task) who did not take part in any of the two eye-tracking experiments. The design of Experiment 2 was the same as Experiment 1. In addition, there were 144 filler sentences, 48 of these contained other color words such as blue (蓝), yellow (黄), orange (橙) and cyan (青) but these were presented normally in the sentence. The power of Experiment 2 (96 participants and 72 sets of stimuli) is estimated to be 0.88, that is greater than the recommended level of 0.80.

Apparatus and procedure

The same apparatus and procedure were used as in Experiment 1, though the monitor was Viewsonic P225f with the same display parameters as in Experiment 1.

Results and discussion

Participants' mean comprehension accuracy was 88 %, again suggesting that they read and fully understood the sentences. Similar to Experiment 1, 98 % participants reported that they were aware of display changes. Among these, they estimated that on average there were 25 % trials (SD = 21 %) including character changes, 30 % (SD = 22 %) including color changes, and 18 % (SD = 19 %) including both changes. These estimates are slightly higher than our real changes (16 % character changes, 21 % color changes, and 11 % both changes, among all the sentences including filler sentences that participants read). Again, though participants were aware of display changes, the disruption to reading was minimised.

The same data exclusion criteria were used as in Experiment 1. Fixations shorter than 80 ms or longer than 1200 ms were removed. Trials were removed when a) participants blinked while the display changed or the boundary changed (4.3 %); b) the display changes triggered late (14.4 %); and c) a saccade crossed the boundary (triggering the display change) but hooked back to the left (9.4 %). Finally, for each measure observations were removed that were three standard deviations from the mean of each participant (for the target word analyses: 1.0 %; for the pretarget region analyses: 2.0 %). Again, as in Experiment 1, the total percentage of trials excluded from the analyses quite high but in line with other display change experiments (e.g., Kliegl et al., 2007). Means and standard deviations for the eye movement measures on the target word and the pretarget region are shown in Table 3, and the corresponding fixed effect estimations for these measures are shown in Table 4.

The target word

As in Experiment 1, for all the eye movement measures, there was a reliable effect of target preview with shorter fixations for the identity preview compared with the pseudocharacter preview (all $t > 7.35$), replicating the standard preview effect (Liversedge & Findlay, 2000; Rayner, 2009). Again, this suggests that our preview manipulation was effective. Furthermore, readers made longer fixations when they had a congruent color than a black preview (all $t > 4.00$). The interaction between black and congruent color previews and target preview type approached but did not achieve significance in TFD ($t = -1.87$, $p = 0.06$). The planned contrasts showed that the Stroop effect at the target word was stronger for identical previews ($b = 0.10$, $SE = 0.03$, $t = 3.20$) than pseudocharacter previews ($b = 0.05$, $SE = 0.03$, $t = 1.64$, see Fig. 4). Most importantly, there were reliable differences at the target word between the incongruent and congruent color preview conditions in all fixation time measures (all $t > 2.44$), with shorter fixations for congruent than incongruent color previews. Note, that the counterpart parafoveal target word Stroop effect in Experiment 1 approached significance only for the TFD, whereas here it patterned similarly and was consistently reliable for all fixation time measures.

When the preview color was incongruent with the meaning of the target word, it produced disruption to identification of the target. When readers process the target word in the parafovea, they clearly identify its color and this is maintained in memory. When they then move their eyes to fixate the target, they cross the second boundary causing the target word to be presented in black. However, upon its fixation, the target can be fully lexically identified and when this happens, and the semantic meaning of the word is incongruent with the color representation that was encoded from the preview in the parafovea, this causes disruption to processing. Note that these results do not suggest that the semantics of the word are processed when it is in the parafovea. It is the color of the parafoveal word that is processed and the memory representation for this produces the stroop effect when the target word is fixated and identified.[7]

The pretarget region

The effect of target preview was reliable in TFD with shorter total fixations for identity than pseudocharacter previews ($t = 3.78$), replicating the orthographic parafoveal-on-foveal effect observed in Experiment 1. In addition, the congruent color preview produced shorter fixations on the pretarget region compared to the black preview for the fixation time measures (all $|t| > 2.53$, though it did not achieve significance in FFD, $t = -1.73$, $p = 0.08$). As in Experiment 1, this was probably due to the parafoveal visual colour oddity capturing visual attention and drawing the point of fixation from the pretarget to the target region rapidly. None of the other effects or interactions were reliable. Together the pretarget region results are consistent with a serial lexical processing account of reading, and like the results of Experiment 1, they offer no evidence to support parallel accounts that specify semantic PoF effects during reading.

Additional analyses

Given that we observed a very similar pattern of effects to that obtained in Experiment 1, namely, target preview effects and parafoveal Stroop effects but no hint of PoF Stroop effects, we combined the data from Experiment 2 with the data from the trials in which the red and green stimuli were used in Experiment 1. This gave a total number of participants for these analyses of 198 (102 from Experiment 1 and 96 from Experiment 2, amounting to 1764 observations per condition. These values provide a 0.96 power value demonstrating that we have

---

[7] Note that as in Experiment 1, we also conducted analyses for the target word for only those trials in which the pretarget region was fixated (i.e., we removed trials where the pretarget region was skipped). The pattern of effects was exactly the same as that reported here.

Table 3

Means and standard deviations for the eye movement measures on the target word and the pretarget region in Experiment 2.

Analysis region		Identity			Pseudocharacter		
		Black	Congruent	Incongruent	Black	Congruent	Incongruent
The target word	FFD (ms)	252 (57)	277 (58)	292 (77)	280 (61)	309 (70)	329 (86)
	SFD (ms)	254 (55)	278 (58)	290 (75)	289 (67)	317 (72)	334 (92)
	GD (ms)	267 (69)	289 (68)	302 (78)	310 (83)	334 (80)	353 (95)
	TFD (ms)	299 (88)	339 (95)	363 (118)	381 (134)	384 (112)	416 (140)
The pretarget region	FFD (ms)	213 (39)	209 (30)	214 (64)	214 (36)	207 (36)	205 (32)
	SFD (ms)	212 (38)	208 (30)	215 (71)	214 (37)	206 (37)	205 (31)
	GD (ms)	235 (57)	225 (47)	230 (86)	235 (55)	226 (64)	226 (50)
	TFD (ms)	340 (111)	330 (105)	333 (119)	362 (116)	349 (132)	362 (129)

Note: Standard deviations are provided in parentheses. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration.

Table 4

Fixed effects estimates for the eye movement measures on the target word and the pretarget region in Experiment 2.

Analysis region		FFD			SFD			GD			TFD		
		b	SE	t	b	SE	t	b	SE	t	b	SE	t
The target word	**Target Preview**												
	Pseudocharacter vs Identity	0.10	0.01	**7.36**	0.14	0.01	**9.77**	0.16	0.01	**11.26**	0.16	0.02	**9.93**
	Color Preview												
	Congruent vs Black	0.10	0.02	**5.46**	0.09	0.02	**5.40**	0.09	0.02	**4.89**	0.08	0.02	**4.01**
	Incongruent vs Congruent	0.05	0.02	**3.07**	0.04	0.02	**2.45**	0.05	0.02	**2.96**	0.07	0.02	**3.39**
	Interaction												
	Target Preview × Congruent vs Black	−0.02	0.03	−0.52	−0.02	0.03	−0.48	−0.03	0.04	−0.74	−0.08	0.04	<u>−1.87</u>
	Target Preview × Incongruent vs Congruent	0.02	0.03	0.60	0.00	0.03	0.14	0.00	0.03	0.14	−0.01	0.04	−0.14
The pretarget region	**Target Preview**												
	Pseudocharacter vs Identity	−0.00	0.01	−0.37	−0.01	0.01	−0.68	0.00	0.01	0.06	0.05	0.01	**3.78**
	Color Preview												
	Congruent vs Black	−0.02	0.01	<u>−1.73</u>	−0.03	0.01	**−2.54**	−0.04	0.01	**−3.68**	−0.05	0.02	**−2.74**
	Incongruent vs Congruent	−0.00	0.01	−0.19	0.00	0.01	0.01	0.01	0.01	0.45	0.02	0.02	1.34
	Interaction												
	Target Preview × Congruent vs Black	−0.03	0.02	−1.44	−0.03	0.02	−1.41	−0.01	0.02	−0.47	−0.03	0.03	−0.74
	Target Preview × Incongruent vs Congruent	0.00	0.02	0.11	0.01	0.02	0.42	0.02	0.02	1.01	0.04	0.03	1.26

Note. Significant terms are presented in bold, and terms approaching significance are underlined.
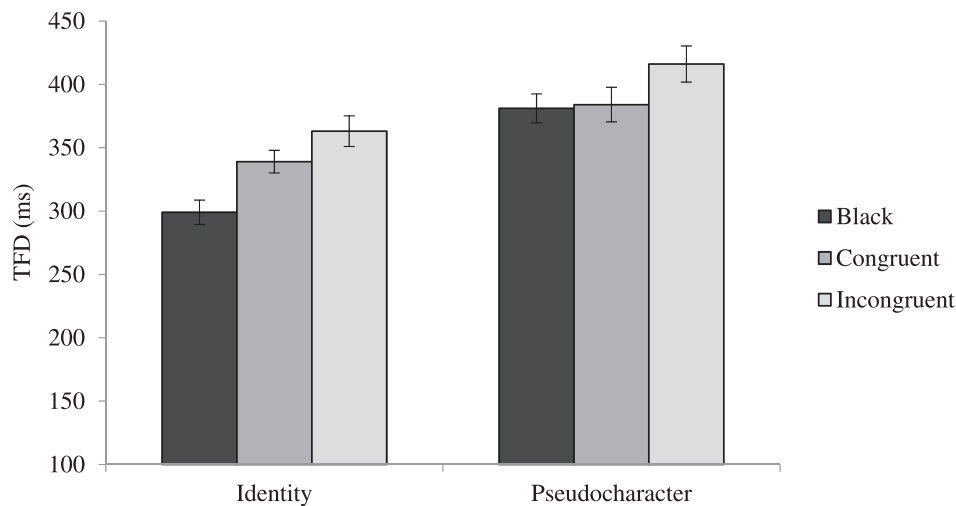

Fig. 4. Total fixation duration as a function of target preview and color preview condtions on the target word.

substantial power). We then carried out analyses for the pretarget region to further test whether we might obtain any evidence to support a PoF Stroop effect. The means and standard deviations for the eye movement measures for the combined analyses of data from both experiments on the pretarget region are shown in Table 5, and the corresponding fixed effect estimations for these measures are shown in Table 6.

In this final set of analyses, at the pretarget region, the effect of target

preview did not achieve significance in FFD ($t = -1.68$, $p = .09$), but was robust in SFD ($t = -2.08$) and TFD ($t = 5.68$), with longer fixations for identical previews in SFD but shorter fixations in TFD than for pseudocharacter previews, replicating the patterns observed from Experiments 1 and 2. The difference between the congruent color preview and the black preview was reliable in all eye movement measures with shorter fixations for the former than the latter (all $|t| > 2.89$). The

Table 5

Means and standard deviations for the eye movement measures on the pretarget region from Experiments 1 and 2.

	Identity			Pseudocharacter		
	Black	Congruent	Incongruent	Black	Congruent	Incongruent
FFD (ms)	216 (38)	212 (31)	217 (59)	217 (37)	208 (37)	206 (32)
SFD (ms)	215 (38)	210 (30)	217 (64)	217 (38)	207 (39)	205 (33)
GD (ms)	237 (58)	230 (55)	242 (92)	242 (62)	228 (69)	229 (60)
TFD (ms)	335 (106)	328 (112)	342 (139)	361 (113)	358 (139)	365 (142)

Note: Standard deviations are provided in parentheses. FFD = first fixation duration; SFD = single fixation duration; GD = gaze duration; TFD = total fixation duration.

Table 6

Fixed effects estimates for the eye movement measures on the pretarget region from Experiments 1 and 2.

	FFD			SFD			GD			TFD		
	b	SE	t	b	SE	t	b	SE	t	b	SE	t
Target Preview												
Pseudocharacter vs Identity	−0.01	0.01	<u>−1.68</u>	−0.01	0.01	**−2.08**	−0.00	0.01	−0.40	0.06	0.01	**5.68**
Color Preview												
Congruent vs Black	−0.02	0.01	**−3.06**	−0.03	0.01	**−3.78**	−0.05	0.01	**−4.75**	−0.04	0.01	**−2.90**
Incongruent vs Congruent	0.00	0.01	0.25	0.01	0.01	0.64	0.01	0.01	1.37	0.02	0.01	<u>1.72</u>
Interaction												
Target Preview × Congruent vs Black	−0.03	0.02	<u>−1.77</u>	−0.03	0.02	<u>−1.90</u>	−0.03	0.02	<u>−1.79</u>	−0.02	0.03	−0.58
Target Preview × Incongruent vs Congruent	0.00	0.02	0.25	0.01	0.02	0.33	0.01	0.02	0.58	0.02	0.03	0.63

Note. Significant terms are presented in bold, and terms that approached significance are underlined.

interaction between target preview and the difference between the congruent and black previews did not achieve significance in FFD ($t = -1.77$, $p = 0.07$), SFD ($t = -1.90$, $p = 0.06$) and GD ($t = -1.79$, $p = 0.07$, see Fig. 5). This effect reflects the point of fixation being rapidly attracted to a colored parafoveal preview.

Importantly, the interaction between target preview and the difference between the incongruent and congruent previews was not reliable in any of the eye movement measures (all $t < 0.64$). Bayes factor analyses for LMMs (Morey, Rouder, Jamil, Urbanek, Forner, & Ly, 2018) were calculated to quantify the level of uncertainty regarding the interaction relative to a null hypothsis. The default scale prior (0.5) and 100,000 Monte Carlo iterations of the BayesFactor package were used. The *BF* values equalled 0.06, 0.06, 0.05 and 0.06 in FFD, SFD, GD and TFD respectively, which favoured a null hypothesis (all *BF* < 1). Using different priors (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8), the sensitivity analysis showed consistent results, all *BF*s < 0.15. Recall that we predicted that if the semantic meaning of the target word is extracted parafoveally and the extraction is early enough to cause semantic PoF effects, then this would produce an incongruency effect (i.e., a stroop effect) at the pretarget region, that is, as an interaction between preview

congruency and preview type. Such effects did not occur even when we combined the data sets to provide substantial power, and Bayes Factor analyses for all of the measures supported the null hypothesis. Therefore, we obtained no evidence for semantic PoF Stroop effects at the pretarget region.

In sum, in these experiments we examined a very robust effect, namely the Stroop effect, to measure lexical influences of parafoveal words on eye movements. Note that we also undertook our experiments using a visually dense orthography, written Chinese, and we used a large number of experimental stimuli and a substantial number of participants. Finally, we combined the data from both of our experiments to optimise statistical power in our analyses. We opted for all of these aspects of our experimental approach in order to maximise the possibility of obtaining lexical (semantic) PoF effects and to provide the strongest test of serial vs parallel accounts specifying semantic PoF effects during reading. It should be very clear that under these circumstances the results are almost entirely consistent with a serial processing account, and at best, the evidence to support parallel accounts that specify PoF effects can only be considered to be extremely modest.
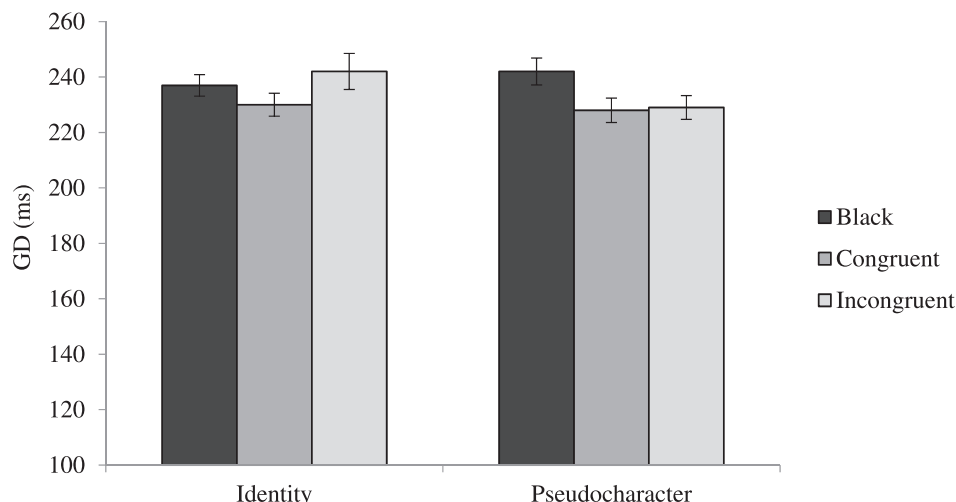

Fig. 5. Gaze duration as a function of target preview and color preview condtions on the pretarget region.

General discussion

Parafoveal vision is important for effective reading and information is systematically extracted from parafoveal words in order to ensure fast and efficient reading. There is, however, much equivocality regarding whether lexical or semantic processing of parafoveal words occurs sufficiently rapidly that it affects fixation times on foveal words. To be clear, the issue here concerns whether fixation durations on a currently fixated word may vary as a function of the lexical or semantic information that is available and is processed from parafoveal vision. Recall that evidence around the existence of lexical semantic PoF effects has been controversial, and arguments over this relate to whether lexical processing occurs serially or in parallel (as per different models of eye movement control during reading). To reiterate, generally, serial lexical processing accounts such as E-Z Reader do not predict these effects, whereas parallel lexical processing accounts such as SWIFT do. Recall, as discussed in the Introduction, OB1 assumes that semantic information can be extracted in parallel from, but not integrated across, parafoveal and foveal words. Thus, it might be argued that OB1 may not predict semantic PoF effects. However, to us, it is not immediately obvious why Stroop PoF effects should not occur. OB1 clearly specifies that words are lexically identified in parallel. Presumably, lexical identification will deliver the meaning of the word. And upon accessing the meaning of the word, in the incongruent condition in the current paradigm, the color in which the word is written will cause semantic interference. To be very clear, no integration across words is required in order for the effects to occur. Instead, they are driven entirely by processing associated with the word itself that occurs when the word is identified and its meaning is accessed. As we have noted, OB1 clearly specifies that words are identified in parallel. Thus, to us, it seems reasonable to suggest that at some point during a particular fixation the parafoveal word will be lexically identified and disruption due to the colour incongruity manipulation should occur. If the point of fixation is on the word prior to the coloured word when this happens, then this fixation will be increased due to that disruption.

Even if this logic, for some reason, is incorrect, it remains the case that OB1 specifies that access to words' semantic characteristics in parallel occurs via a spatiotopic sentence-level representation that is guided by expectations about word length and syntactic structure. Thus, at the very least, the current experimental paradigm and the null results with respect to PoF Stroop effects question the efficacy of any such system. Of course, there may be good reasons for this. Recall that Chinese is an unspaced language in which word boundary ambiguity is prevalent and for which there is a significant degree of syntactic flexibility (relatively free word order). It is, therefore, unclear how the spatiotopic mappings that OB1 relies upon might be achieved during Chinese reading given the lack of word length cues, word boundary ambiguity and less rigid syntactic constraints on word order. These issues require consideration into the future.

To return to our main line of discussion, empirically, most of the evidence for lexical PoF effects has been reported in large-scale corpus based studies (e.g., Angele et al., 2015; Kennedy & Pynte, 2005; Kliegl et al., 2006; Li et al., 2014; see Drieghe, 2011 for a review) but not in tightly controlled experimental studies (e.g., Angele et al., 2015; Brothers et al., 2017). In the present study, we employed a novel Stroop boundary paradigm and manipulated the preview and print color of a target word such that it was either a pseudocharacter or the identity, and its colour was congruent or incongruent with the target word's meaning, or black (no colour). Our objective was to investigate whether readers extract semantic information parafoveally, and whether any such extraction of semantic information might rapidly influence processing of the foveal word. More specifically, whether we could obtain evidence for lexical or semantic PoF effects (i.e., PoF Stroop effects) during Chinese reading. The Stroop effect is, arguably, one of the most compelling pieces of empirical evidence for automatic processing of semantic information (Stroop, 1935). In Experiment 1, the classic four colors and

color words – red, green, yellow and blue, were used as previews, whereas in Experiment 2, only the more visually salient red and green colors were used. We also elected to conduct our study in Chinese as the characteristics of the orthography maximise the possibility of obtaining parafoveal processing effects.

Our main findings are straightforward: For the target word, both experiments showed a reliable target preview effect with longer fixations for pseudocharacter than identity previews, replicating the standard preview effect and demonstrating that our preview manipulation was effective (Liversedge & Findlay, 2000; Rayner, 2009); Second, both experiments showed that reading times were increased at the target word for colored previews compared to the black previews, indicating a visual interference effect due to the color of the target when in the parafovea. Furthermore, consistent with our predictions, a reliable parafoveal Stroop effect was observed at the target word for both identity and pseudocharacter previews, with shorter fixation times for congruent than incongruent color previews in all fixation time measures in Experiment 2, though this effect did not achieve significance in total fixation durations in Experiment 1. Apparently, the incongruent color preview produced disruption to identification of the target word when it was fixated, indicating that the color preview information is extracted parafoveally and integrated with the meaning of the target word upon its fixation, and the classic Stroop effect (MacLeod, 1991; Stroop, 1935) mediated by a parafoveal color cue and access to word meaning at fixation occurs across saccades during normal sentence reading.

Critically, for the pretarget region, if semantic information about the upcoming parafoveal color word was available, then we would have expected to obtain PoF Stroop effects for identity rather than pseudocharacter previews. This pattern of effects did not occur. To reiterate, we combined data from both experiments to provide strong power, and we conducted the experiment in Chinese to maximise the possibility of obtaining even subtle PoF effects. Our combined analyses only showed shorter reading times on pretarget regions when the preview was colored compared with black, demonstrating that parafoveal visual oddity of an upcoming word pulls the point of fixation rapidly to it. However, there was no statistically robust evidence for a PoF Stroop effect and the Bayes factor analyses strongly favored the null hypothesis regarding the interactions.

To us, the results of these experiments are very clear and lead us to conclude that readers do not access the semantic meanings of upcoming parafoveal words before they are fixated. However, they do access such information associated with those words very rapidly when they are fixated. It unambiguously appears to be the case that the results strongly support serial models of eye movement control, and offer little, if any, support for parallel eye movement models of eye movement control that specify semantic PoF effects. We acknowledge, though, that some may argue that null effects do not provide evidence against parallel processing (evidence against a position is not the same as a lack of evidence for a position). From this perspective, regardless of how we demonstrate any absence of an effect, such results will not be taken as satisfactory evidence against the parallel accounts specifying semantic PoF effects. To us, this represents the adoption of a scientifically untestable perspective, that is, the theory stipulates that an influence should occur, results are (repeatedly) presented that fail to demonstrate such effects, and these empirical data are dismissed as holding no value (since they do not represent evidence against the position).

An alternative, in our view, more constructive approach is to assess the extent to which experimental situations offer the possibility of observing particular effects and to objectively quantify the evidence for one, or other, position on that basis. That is to say, we consider that both serial and parallel accounts must be assessed against each other in relation to the degree of evidence in favour of each, using strong experimental tests, formal assessments of experimental power and advanced statistical approaches (e.g., Bayesian methods) to quantify the likelihood of an experimental analysis having the capacity to reasonably demonstrate any effect that should occur (and, of course, this holds in

respect of either theoretical position). This approach avoids a situation where one side of the evidence is dismissed as inconsequential. From this more forward-looking perspective it may be possible to more readily assess the balance of evidence in favour of alternative positions. We also note that this approach allows us to move forward and gain traction with respect to theoretical development. To return to the current theoretical debate, it is the case that a strict serial position stipulates that under normal processing circumstances, lexical PoF effects (such as the Stroop effects explored in the current study) should not occur. However, it is also just as much the case that the standard parallel position (at least as per the specifications of SWIFT) stipulates such effects, absolutely, should occur. And from such a parallel perspective, why would we not expect evidence of Stroop based PoF effects at the pre-target word (given that they are clearly evident from the very first fixation on the target, thereby demonstrating that the experimental circumstances afford excellent conditions for lexically based Stroop effects to occur)? Thus, such standard parallel assumptions specify that lexical semantic effects should be detectable in fixations on the word prior to the target. However, this was clearly not the case even under experimental conditions that were optimized for the detection of parallel processing effects.

The present results are entirely in line with the findings of Brothers et al. (2017) who carefully manipulated the frequency of parafoveal words in four experiments with 244 participants and showed no reliable effects of parafoveal word frequency (for similar results, see Angele et al., 2015 in their analyses of the pretarget word when the frequency of target word was experimentally manipulated). Brothers et al. further carried out a Bayesian meta-analysis including previous similar studies and their experiments with 988 participants, but failed to find evidence for the PoF effects. Cutter, Martin and Sturt (2020) asked participants to read sentences (e.g., *The tall lanky guard who alerted Charlie/Charlie alerted to the danger was young*) including a subject relative clause (SRC, *alerted Charlie*) or an object relative clause (ORC, *Charlie alerted*), and found increased gaze durations on the word *who* immediately preceding the ORC relative to the SRC, that is, they obtained an early effect of the type of relative clause on processing at the pretarget. Importantly, the capitalization of the initial letter of the proper noun acts as a parafoveal visual cue to its syntactic category under normal sentence presentation and reading conditions. However, when sentences were presented in upper case, or in a boundary paradigm version of the experiment, wherein the visual cue of the capitalized initial letter was diminished, or unavailable prior to its direct fixation, this effect did not occur. Cutter et al., thus, argued that the apparent syntactic PoF effect that they obtained in their first experiment was actually very likey caused by a visual cue rather than the parafoveal word being pre-processed to a lexical level. There is a consistency to these studies on alphabetic language reading, along with the present experiments investigating Chinese reading, in that they do not provide evidence for lexical PoF effects.

Snell and Grainger (2019) offered a perspective advocating parallelism in respect of lexical processing and oculomotor control. Perhaps more importantly in this piece, they suggest the need for a paradigm shift away from eye movement investigations of the serial parallel debate and towards alternative experimental approaches, particularly, non-natural reading tasks. As they state "…here transpires a challenge: if words are truly processed in parallel (i.e., without integrating high-level information across words), 'direct' measures of word recognition speed (e.g., word viewing times) as a function of semantic relationships cannot be used to test parallelism (pp. 539–540)". Beyond the question of what actually constitutes a "direct" measure in experimental psychology, in our view, the current results demonstrate unequivocally that (direct) eye movement measures of word recognition (and other aspects of linguistic processing) do reflect semantic relationships between words. Furthermore, it seems reasonable to suggest that such measures can be used to discriminate between serial and parallel processing accounts, particularly when supported with Bayesian analytical techniques to quantify the null. From our perspective, somewhat self-evidently, models of eye movement control in reading should explain the variance in eye movement behaviour that occurs during natural reading. To this extent, such models need to provide a mechanistic account of how visual and linguistic processes occur over time during natural reading, and how those processes influence oculomotor commitments concerning where and when readers move their eyes. If our perspective is correct, then direct measures from eye movement data remain the most appropriate measure to discriminate between serial and parallel processing accounts (see Schotter & Payne, 2019).

Zang, Liversedge and colleagues have proposed a possible (at least partial) solution to the impasse in the serialism/parallelism debate, via the Multi-Constituent Unit hypothesis (MCU, Zang, 2019; Zang et al., 2021; see also Cutter, Drieghe & Liversedge, 2017). According to the MCU account, if the current word *n* and the following word $n + 1$ are strongly associated and frequently used together (e.g., as in the spaced compound, *teddy bear*), they may be represented and stored lexically as a single multi-constituent unit (e.g., Conklin & Schmitt, 2008, 2012; Shaoul & Westbury, 2011; Siyanova-Chanturia et al., 2011; Titone, & Connine, 1999; Wray, 2002; Wulff & Titone, 2014). For lexicalised MCUs of this type, we consider it plausible that lexical identification processes may be operationalized over the fixated constituent and subsequent elements simultaneously. In this way, for these linguistic forms, word $n + 1$, and in principle other constituents beyond $n + 1$ that comprise the MCU, could be lexically processed to a significant degree prior to their fixation during reading. In these cases, the lexical characteristics of word $n + 1$ could influence reading times on word *n*, and both *n* and $n + 1$ would be processed simultaneously (since they are part of the same lexical representation). To be clear, we see this situation as being rather analogous to how readers operationalize lexical processing across longer single words such as *basketball* (and long agglutinate words in languages like Finnish or German). This suggestion might offer an explanation for why, on some occasions, lexical processing might appear to occur in parallel, and modest lexical PoF effects might occur. Recall, also, the CRM described earlier (Li & Pollatsek, 2020). As we noted, the CRM does not take a firm position with respect to the issue of whether lexical identification occurs serially and sequentially, or in parallel during reading. Instead, the CRM specifies that lexical identification and word segmentation in Chinese are part of the same process (essentially, the identification of a word leads to a simultaneous segmentation commitment). In our view, this property of the CRM lends itself very readily to the MCU hypothesis that we have advocated (see also Gu, Zhou, Bao, Liu, Perea & Li, 2022). Furthermore, a model operating according to this approach would, on some occasions, appear to exhibit serial processing characteristics, and on other occasions parallel. Again, at some level, to us, there appears to be some theoretical commonality here. Nonetheless, as the model's name suggests, the CRM was developed to explain processing underlying Chinese reading, and therefore, it remains to be examined whether the processing assumptions lend themselves to reading in alphabetic (as well as other) languages. Before leaving our discussion of lexical PoF effects, we must also note that there are other existing explanations for their basis including saccade targeting errors, binocular disparity effects, and calibration errors (Drieghe, 2011). These aspects of processing might lead to modest PoF effects meaning that they would have a relatively reduced influence on lexical processing (see Brothers et al., 2017).

In summary, for experiments that adopt standard experimental designs it remains very difficult to demonstrate lexical PoF effects. As Brothers et al. (2017) have suggested, it is possible that frequency based lexical PoF effects simply do not exist. Beyond this, the current experiments demonstrate that Stroop based PoF effects do not occur under the most stringent of experimental circumstances.

CRediT authorship contribution statement

Chuanli Zang: Conceptualization, Methodology, Writing – original draft, Funding acquisition. **Zhichao Zhang:** Investigation, Formal analysis. **Manman Zhang:** Methodology. **Federica Degno:**

Methodology. **Simon P. Liversedge:** Funding acquisition, Supervision, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data sets and analysis scripts are publicly available at: https://osf.io/7b85t/.

Acknowledgments

References

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

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

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

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language, 96*, 9–22.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72–89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2017). Is orthographic information from multiple parafoveal words processed in parallel: An eye-tracking study. *Journal of Experimental Psychology: Human Perception and Peformance, 43*, 1550–1567.

Cutter, M. G., Martin, A. E., & Sturt, P. (2020). Capitalization interacts with syntactic complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(6), 1146–1164.

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 839–856). New York, NY: Oxford University Press.

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 787–800). New York: Oxford University Press.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research., 42*, 621–636.

Gu, J., Zhou, J., Bao, Y., Liu, J., Perea, M., & Li, X. (2022). The Effect of Transposed-Character Distance in Chinese Reading. *Journal of Experimental Psychology: Learning Memory & Cognition*. In press.

He, L., Song, Z., Chang, M., Zang, C., Yan, G., & Liversedge, S. P. (2021). Contrasting off-line segmentation decisions with on-line word segmentation during reading. *British Journal of Psychology, 112*, 662–689. https://doi.org/10.1111/bjop.12482

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 68–81.

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.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 149–179). Oxford: Elsevier Science Ltd.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45*, 153–168.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*, 12–35.

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+2. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1250–1255.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General, 143*, 895–913.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review, 127*(6), 1139–1162.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4*, 6–14.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203.

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S. Forner, K., & Ly, A. (2018). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from http://CRAN.R-project.org/packageBayesFactor.

Pan, J., Laubrock, J., & Yan, M. (2016). Parafoveal processing in silent and oral reading: Reading mode influences the relative weighting of phonological and semantic information in Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1257–1273.

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

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*, 65–81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K. (2009). The thirty-fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Reichle, E. D. (2011). Serial-Attention Models of Reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 767–786). New York: Oxford University Press.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13*, 115–119.

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

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*(4–5), 391–421.

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

Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading. *Trends in Cognitive Sciences, 23*, 811–812.

Schotter, E. R., Reichle, E. D., & Rayner, K. (2014). Re-thinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N+2 preview effects. *Visual Cognition, 22*, 309–333.

Shaoul, C., & Westbury, C. F. (2011). Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon, 6*(1), 171–196.

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences., 23*, 537–546.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018a). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review, 125*(6), 969–984.

Snell, J., Declerck, M., & Grainger, J. (2018b). Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition and Neuroscience, 33*(5), 563–574.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 776–784.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*(12), 1655–1674.

Westfall, J. (2015). PANGEA: Power ANalysis for GEneral Anova designs. Unpublished manuscript. *Available at http://jakewestfall. org/publications/pangea. pdf*.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wulff, S., & Titone, D. (2014). Introduction to the special issue, bridging the methodological divide: Linguistic and psycholinguistic approaches to formulaic language. *The Mental Lexicon, 9*, 371–376.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Chinese readers extract semantic information from parafoveal words during reading. *Psychonomic Bulletin & Review, 16*, 561–566.

Yan, M., & Sommer, W. (2015). Parafoveal-on-foveal effects of emotional word semantics in reading Chinese sentences: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1237–1243.

Yang, J., Wang, S., Tong, X., & Rayner, K. (2012). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing, 25*, 1079–1091.

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision, 3*(50), 1–13.

Zang, C., Du, H., Bai, X., Yan, G., & Liversedge, S. P. (2020). Word skipping in Chinese reading: The role of high-frequency preview and syntactic felicity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 603–620.

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2021). Foveal and Parafoveal Processing of Chinese Three-character Idioms in Reading. *Journal of Memory and Language*. in press.

Zang, C., Liversedge, S. P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 961–978). New York: Oxford University Press.