

# Attention, Perception, & Psychophysics

## Skipping of Chinese characters does not rely on word-based processing

Journal:	<i>Attention, Perception, &amp; Psychophysics</i>
Manuscript ID	PP-ORIG-17-050.R1
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Lin, Nan; Institute of Psychology, Chinese Academy of Sciences, Angele, Bernhard; Bournemouth University Hua, Huimin; Institute of Psychology, Chinese Academy of Sciences, Shen, Wei; Institute of Psychology, Chinese Academy of Sciences Zhou, Junyi; Institute of Psychology, Chinese Academy of Sciences Li, Xingshan; Chinese Academy of Sciences, Psychology
Keywords:	eye movements and reading, reading, word perception

SCHOLARONE™  
Manuscripts

Only

1  
2  
3 Skipping of Chinese characters does not rely on word-based processing  
4  
5

6 Nan Lin  
7

8 CAS Key Laboratory of Behavioral Science, Institute of Psychology  
9  
10 Department of Psychology, University of Chinese Academy of Sciences  
11  
12

13 Bernhard Angele  
14

15  
16 Psychology Research Centre, Faculty of Science and Technology, Bournemouth  
17  
18 University  
19

20  
21 Huimin Hua, Wei Shen, Junyi Zhou, and Xingshan Li  
22

23 CAS Key Laboratory of Behavioral Science, Institute of Psychology  
24  
25 Department of Psychology, University of Chinese Academy of Sciences  
26  
27

28  
29  
30  
31 **Author Note**  
32

33  
34  
35 This research was supported by grants from the Natural Science Foundation of  
36  
37 China (31571125, 31300842), a grant from British Academy (International  
38  
39 Partnership and Mobility Scheme PM150023), and the Beijing Advanced Innovation  
40  
41 Center for Imaging Technology (BAICIT-2016018).  
42

43 Correspondence concerning this article should be addressed to Xingshan Li, CAS  
44  
45 Key Laboratory of Behavioral Science, Institute of Psychology, 16 Lincui Road,  
46  
47 Beijing, 100101, China, E-mail address: lixs@psych.ac.cn  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Abstract

Previous eye-movement studies have indicated that people tend to skip extremely high frequency words (such as “the” in English and “的/de” in Chinese) in sentence reading. Two alternative hypotheses have been proposed to explain how this frequent skipping happens in Chinese reading: one assumes that skipping happens when the preview has been fully identified at the word level (word-based skipping); the other assumes that skipping happens whenever the preview character is easy to identify regardless of whether lexical processing has been completed or not (character-based skipping). Using the gaze-contingent display change paradigm, we examined the two hypotheses by substituting the preview of the third character of a four-character Chinese word with the high-frequency Chinese character “的/de”, which should disrupt the ongoing word-level processing. The character-based skipping hypothesis predicts that this manipulation will enhance the skipping probability of the target character (i.e. the third character of the target word) because the character “的/de” has much higher character frequency than the original character; the word-based skipping hypothesis instead predicts a reduction of the skipping probability of the target character because the presence of the character “的/de” is lexically infelicitous at word level. The results supported the character-based skipping hypothesis, indicating that in Chinese reading the decision of skipping a character can be made before integrating it into a word.

1  
2  
3  
4 In sentence reading, people do not fixate every single word. Word skipping, that  
5  
6 is, moving one's gaze past a word during first-pass reading without fixating it,  
7  
8 happens frequently in both alphabetic reading (Carpenter & Just, 1983; Rayner, 1998)  
9  
10 and Chinese reading (Inhoff & Liu, 1998). How do readers decide whether to skip a  
11  
12 word or not? In alphabetic reading, word-skipping probability is modulated by the  
13  
14 length and frequency of the previewed word and the predictability of the upcoming  
15  
16 word (Abbott, Angele, Ahn, & Rayner, 2015; Angele, Laishley, Rayner, &  
17  
18 Liversedge, 2014; Angele & Rayner, 2013; Rayner, 1998). Even though the effect of a  
19  
20 word's predictability on the probability of skipping is well established, Angele and  
21  
22 Rayner (2013) found that, at least in English, the influence of contextual information  
23  
24 on skipping is rather weak and can be easily overridden by preview information:  
25  
26 Using the gaze-contingent display change paradigm (Rayner, 1975), Angele and  
27  
28 Rayner (2013) manipulated the preview that was available for a three-letter target  
29  
30 word. When the upcoming word appeared to be the definite article "the" (a function  
31  
32 word with the highest word-frequency in English vocabulary) readers skip it very  
33  
34 frequently even if the syntactic sentence context does not allow for an article in that  
35  
36 position in the sentence (e.g. "the" as the preview for "ban" in "The council voted to  
37  
38 immediately ban smoking near public buildings"). A follow-up study (Angele et al.,  
39  
40 2014) has found a similar effect for skipping previews of short high-frequency content  
41  
42 words in infelicitous positions (e.g. "sad" as a preview for "sew" in "She would  
43  
44 always sew her own dresses").  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

55  
56 Furthermore, it seems that preview information is processed completely  
57  
58  
59  
60

1  
2  
3  
4 independently from contextual information: Abbott et al. (2015) found that  
5  
6 manipulating the semantic constraint of the sentence (i.e. the predictability of the  
7  
8 word in whose position the preview of “the” appeared, for example “die” in “If you  
9  
10 are shot in the head you will likely die” vs. “If you are shot in the foot you will likely  
11  
12 die”) does not reduce skipping of words that look like an infelicitous occurrence of  
13  
14 “the”. In fact, the effect of predictability on skipping that Abbott et al. (2015) found  
15  
16 was completely independent from the “the”-skipping effect. This led to the somewhat  
17  
18 counterintuitive result that readers in Abbott et al.’s (2015) experiment were more  
19  
20 likely to skip an infelicitous “the”-preview if the context up to the position of the  
21  
22 “the”-preview was highly predictive of the actual target word (which was never “the”)  
23  
24 than if the actual target word was of low predictability. These findings indicate that, in  
25  
26 quite a few cases, readers make their word-skipping decision before integrating a  
27  
28 previewed word into the sentence, while taking the contextual constraint into account  
29  
30 separately (and without checking if the word at the target position actually fit the  
31  
32 constraint).  
33  
34  
35  
36  
37  
38  
39  
40

41       Skipping also happens frequently in Chinese reading. Previous studies have  
42  
43 highlighted several characteristics of skipping of Chinese characters and words. First,  
44  
45 Chinese readers do not fixate every character. The probability of fixating one  
46  
47 particular character in a sentence is about 42.8%, and the average saccade length is  
48  
49 3.15 characters (Li, Bicknell, Liu, Wei, & Rayner, 2014). Second, words comprising  
50  
51 two or more characters are also sometimes skipped as a whole, and longer words are  
52  
53 skipped less often than short words. For example, Li, Liu & Rayner (2011) showed  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 that the skipping probability for two-character words was .34, while the skipping rate  
5  
6 for four-character words was .07. Third, all characters in a word have similar fixation  
7  
8 probabilities during first pass reading independent of their position within the word  
9  
10 (Li et al., 2011; Li et al., 2014). Fourth, the reason why Chinese readers do not fixate  
11  
12 every single character seems to be that Chinese readers can process more than one  
13  
14 character during each fixation. It has been shown that the perceptual span in Chinese  
15  
16 reading comprises one character to the left and two to three characters to the right of  
17  
18 fixation (Inhoff & Liu, 1997, 1998; Tsai & McConkie, 1995).  
19  
20  
21  
22  
23

24       Although we now know about many aspects of word/character skipping during  
25  
26 Chinese reading, the exact mechanism of how to determine when to skip a word or a  
27  
28 character is still not clear. Interestingly, a recent study of Chinese reading (Zang et al.,  
29  
30 in press) has obtained some findings highly similar to those of Angele and Rayner  
31  
32 (2013). In their study, Zang et al. investigated whether the preview of a  
33  
34 single-character word – “的/de”, which is a function word (more precisely, a  
35  
36 grammatical particle, indicating, among other things, ownership) and the word with  
37  
38 the highest word-frequency in the Chinese vocabulary, can enhance skipping  
39  
40 probability. They found that, just as occurs with “the” in English reading, a preview of  
41  
42 “的/de” increased the probability of skipping the character, even when it occurs in a  
43  
44 position where it violates syntactic constraints from the sentence context.  
45  
46  
47  
48  
49  
50

51       What triggers this “automatic” skipping response? One possibility is that readers  
52  
53 might process the preview word lexically and then make decision whether it will be  
54  
55 necessary to fixate that word in order to achieve full lexical access or not. This would  
56  
57  
58  
59  
60

1  
2  
3  
4 be predicted by the E-Z Reader model (Reichle, Pollatsek, & Rayner, 2006). The E-Z  
5  
6  
7 reader model assumes that words are processed serially, and the decision on when to  
8  
9 move the eyes and where to move the eyes are directly determined by different stages  
10  
11 of word processing. The completion of the “familiarity check” stage L1 on the  
12  
13 currently fixated word (word  $n$ ) will initiate a saccade program to move the eyes to  
14  
15 next word (word  $n + 1$ ). The currently fixated word will then finish lexical access  
16  
17 (stage L2) concurrently with the planning of the upcoming saccade. Attention will  
18  
19 shift to word  $n+1$  if word  $n$  can be fully identified during this stage. Therefore,  
20  
21 attention is likely to be shifted to word  $n+1$  when the eyes are still on word  $n$  and the  
22  
23 lexical processing of word  $n + 1$  is likely to begin (again with the familiarity check  
24  
25 stage L1) while the eyes are still on word  $n$ . The eye movement control module needs  
26  
27 some time to program and prepare the saccade after they receive the saccade program  
28  
29 signal (stages M1 and M2). In some situations, the familiarity check (L1) of word  $n$   
30  
31  $+1$  finishes quickly enough that saccade programming still is at a stage where the  
32  
33 saccade to word  $n+1$  (now no longer necessary) can still be cancelled (i.e. before the  
34  
35 end of stage M1). In this situation, a new saccade to word  $n+2$  (the word to the right  
36  
37 of word  $n+1$ ) is prepared and, as a consequence, word  $n + 1$  is skipped.  
38  
39  
40  
41  
42  
43  
44  
45

46  
47 Essentially, the oculomotor system “bets” that it will be able to finish identifying  
48  
49 the current word before the saccade preparation is complete. If both the currently  
50  
51 fixated word and the upcoming word are easy to identify, the model predicts that there  
52  
53 will be enough time to identify both words and move attention to the subsequent word.  
54  
55 If this happens early enough, the saccade to the upcoming word will be cancelled and  
56  
57  
58  
59  
60

1  
2  
3 a skipping saccade programmed instead. We will refer to this hypothesis as the  
4  
5  
6 *word-based skipping hypothesis*.  
7

8  
9 Another hypothesis has been proposed to explain how skipping happens,  
10  
11 especially in Chinese reading. Since there are no inter-word spaces to mark word  
12  
13 boundaries, Chinese readers usually do not target any specific position within a word  
14  
15 (Li, Liu & Rayner, 2011). Instead, some studies have suggested that Chinese readers  
16  
17 might use a processing-based strategy to plan their eye movements: During any given  
18  
19 fixation, they try to process as much information as possible foveally and parafoveally,  
20  
21 and then plan their next eye movements to obtain information beyond what they have  
22  
23 already processed (Li et al., 2011; Wei, Li, & Pollatsek, 2013). Thus, if the preview  
24  
25 character is easy to process (e.g. a high-frequency character such as “的/de”), the  
26  
27 resulting saccade is likely to be long, increasing the chance that the preview character  
28  
29 may be skipped. Instead of emphasizing the role of words, the processing-based  
30  
31 strategy assumes that the amount of information that has been processed for upcoming  
32  
33 characters is what determines where to move the eyes. We will refer to this possibility  
34  
35 as the *character-based skipping hypothesis*.  
36  
37  
38  
39  
40  
41  
42  
43

44 How can we distinguish between the word-based skipping and character-based  
45  
46 skipping hypotheses? If the character-based skipping hypothesis is true, characters  
47  
48 that are easy to recognize (e.g. characters with low visual complexity or high  
49  
50 frequency) should be skipped more often than those that are difficult to recognize, no  
51  
52 matter whether the character to be skipped is a word on its own, such as in all the  
53  
54 previously reported experiments, or if it is embedded in another word. However, if the  
55  
56  
57  
58  
59  
60



1  
2  
3  
4 decision to skip a character is based on successful word-level processing, embedding  
5  
6 the character in a longer word should lower the probability that it is skipped since its  
7  
8 presence should affect the processing of the embedded word.  
9

10  
11 In the present study, we aimed to examine how the preview of the extremely high  
12  
13 frequency character “的/de” embedded in a multi-character Chinese word affects the  
14  
15 skipping probability. We chose four-character words as target words and we used a  
16  
17 gaze contingent display change paradigm in which participants were asked to read  
18  
19 sentences on a computer screen. Before the eyes crossed an invisible boundary located  
20  
21 to the left of the target character (the third character of the target word), the preview  
22  
23 was substituted with the character “的/de”. After the eyes crossed the boundary, the  
24  
25 third character changed to the correct character as in the target word. The above  
26  
27 design allowed us to examine whether the presence of a high frequency preview  
28  
29 character would make the skipping probability increase when it resulted in lexical  
30  
31 violation at word level. We expected that if Chinese readers decide to skip a  
32  
33 previewed character based on character level processing and before integrating it into  
34  
35 a word, as assumed by the character-based skipping hypothesis, then the previewed  
36  
37 “的/de” may make the skipping probability increase; otherwise, if the skipping  
38  
39 decision is on the premise of successful word-level processes associated with the  
40  
41 preview character, as assumed by the word-based skipping hypothesis, the lexical  
42  
43 violations associated with the previewed “的/de” would result in a reduction of  
44  
45 skipping probability as readers should fixate, rather than skip, a part of the sentence  
46  
47 that is difficult to process.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 A key point of the above experimental design was to ensure that the presence of  
5  
6 the incorrect preview character “的/de” would result in lexical violation at word level.  
7  
8  
9 In this context, it should be noted that the majority of Chinese characters can be used  
10  
11 either as a component of a multi-character word or as a single-character word. For  
12  
13 example, “方针” (“fangzhen”) is a two character word that means “policy” while its  
14  
15 component characters can both be used as single-character words: “方” (“fang”)  
16  
17 means “square”, and “针” (“zhen”) means “needle”. Therefore, if the character “的/de”  
18  
19 is embedded into a short target word (e.g. “方的针”), readers may treat the remaining  
20  
21 character(s) of the target word and the embedded character “的/de” as independent  
22  
23 single-character words (e.g. “方的针” can be interpreted as “a square needle”), which  
24  
25 would possibly constitute a semantic or syntactic violation, but not necessarily a  
26  
27 lexical violation. To minimize this problem, we chose to embed the character “的/de”  
28  
29 in a long word where the remaining characters and the embedded character “的/de”  
30  
31 can hardly be perceived as single-character words.  
32  
33  
34  
35  
36  
37  
38

39 We chose four-character words but not longer words as our target words for two  
40  
41 reasons. First, according to the Chinese Linguistic Data Consortium (2003) corpus,  
42  
43 2.8% of words are one character words, 63.9% are two-character words, 17.5% are  
44  
45 three-character words, 14.2% are four-character words, and only 1.7% are longer than  
46  
47 four characters. Therefore, a typical Chinese word consists of no more than four  
48  
49 characters. Therefore, a typical Chinese word consists of no more than four  
50  
51 characters<sup>1</sup>. Second, previous studies have provided adequate evidence that  
52  
53  
54

---

55  
56 <sup>1</sup> A large portion of four-character Chinese words are four-character idioms. Their role and usage as words are  
57 well-accepted. For example, they are included as words in the Contemporary Chinese Dictionary (Chinese  
58 Academy of Sciences 2005), which is the most authoritative Chinese language dictionary. Therefore, we did not  
59 differentiate four-character idioms from the other four-character words in the present study.  
60

1  
2  
3  
4 four-character words are represented in Chinese readers' mental lexicon. For example,  
5  
6 Li et al. (2009) found that, for a four-character string, Chinese readers can recognize  
7  
8 its third and fourth characters much faster when they constitute a word than when they  
9  
10 constitute two unrelated words or a congruent two-word phrase. Gu et al. (2015)  
11  
12 found that it is more difficult to detect the transposition between the second and third  
13  
14 characters of a four-character string when it is a four-character word than when it is a  
15  
16 phrase consisting of two two-character words. Therefore, we used four-character  
17  
18 words as target words in the current study.  
19  
20  
21  
22  
23

24 We chose only four-character words whose first two characters cannot form a  
25  
26 word on their own as our target words and substituted the preview of their third  
27  
28 characters with the character “的/de”. This effectively made an interpretation of “的  
29  
30 /de” as a function word impossible. Within this design, the first, second, and last  
31  
32 characters of the target word provide rich clues to prime the representation of the  
33  
34 target word, minimizing the possibility that readers treat them as independent  
35  
36 single-character words; the character “的/de” is embedded in the middle of the target  
37  
38 word, preventing it from being perceived as a single-character word preceding or  
39  
40 following the target word; the first two characters cannot form a word either by  
41  
42 themselves or together with the following character “的/de”, ensuring that the  
43  
44 presence of the character “的/de” will immediately result in lexical violations at word  
45  
46 level even before the last character of the target word is recognized.  
47  
48  
49  
50  
51  
52

53  
54 Because there is evidence that the probability of skipping a character is also  
55  
56 affected by its visual complexity (Liversedge et al., 2014; Ma & Li, 2015) and its  
57  
58  
59  
60

1  
2  
3  
4 grammatical properties (e.g. a function word might be skipped more often than a  
5  
6 content word; Carpenter & Just, 1983; cf. Angele et al., 2014), we also added two  
7  
8 control conditions to examine whether the relatively low visual complexity and the  
9  
10 grammatical properties of “的/de” enhance the skipping probability. Given that the  
11  
12 number of function words in Chinese is very limited, there is no other character that  
13  
14 matches “的/de” on both the visual and grammatical dimensions. Therefore we used  
15  
16 two separate control conditions. The characters “钓/diao” serves as a control  
17  
18 condition that closely matches the sub-character visual properties of “的/de”: They  
19  
20 have the same stroke numbers and share the same right radical (i.e., “勺”). The  
21  
22 character “乎/hu” serves as another control condition in that its grammatical  
23  
24 properties are similar to those of “的/de” (though is much less frequent): Both “乎/hu”  
25  
26 and “的/de” can be used as single-character function words and can be used as  
27  
28 auxiliaries as well as modal particles (Sun, Huang, Sun, Li, & Xing, 1997). The  
29  
30 inclusion of the control conditions can help distinguish the two target hypotheses:  
31  
32 According to the word-based skipping hypothesis, the “钓/diao” and “乎/hu”  
33  
34 previews, like the “的/de” preview, should disrupt word-level processing and thus  
35  
36 result in a reduction of skipping probability. On the other hand, according to the  
37  
38 character-based skipping hypothesis, if the “的/de” preview is skipped more often due  
39  
40 to its extremely high frequency but not because of its visual complexity or its  
41  
42 grammatical properties, then the skipping probability of the “钓/diao” and “乎/hu”  
43  
44 previews should not be significantly different from that of the unsubstituted previews.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

## 56 **Method**

57  
58  
59  
60

## Participants

Twenty-eight students (17 females) from colleges around the Institute of Psychology, Chinese Academy of Sciences participated in the experiment. All participants were native Chinese speakers with normal or corrected to normal vision and were naive to the purpose of the experiment.

## Apparatus

Participants' eye movements were recorded using an SR Research EyeLink 1000 eye tracker with a sampling rate of 1000 Hz. Sentences were displayed on a 21-inch CRT monitor with a refresh rate of 150 Hz. Each character subtended a visual angle of approximately 1°.

## Materials

Participants read 80 experimental sentences with four experimental conditions. The sentences were no more than 32 characters long (ranging from 18 to 32 characters) and were all single-line sentences (See Table 1 for examples). Each experimental sentence contained a four-character target word (located at least five characters away from the beginning and the end of the sentence). Frequencies of target words in the Chinese Linguistic Data Consortium (2003) corpus ranged from 0.05 to 6.30 occurrences per million ( $M=0.87$ ;  $SD=0.88$ ).

There were four preview conditions for each target word. The first condition was the identity condition, in which the preview was identical to the target word (e.g., 道貌岸然 as the preview of 道貌岸然, which means “sanctimonious”). The second condition was the “的/de” condition, in which the preview of the third character of the

1  
2  
3  
4 target word (i.e. the target character) was substituted with the character “的/de” (e.g.,  
5  
6 道貌的然 as the preview of 道貌岸然). The third and fourth conditions were control  
7  
8 conditions, in which the preview of the target character was substituted with the  
9  
10 characters “钓/diao” (e.g., 道貌钓然 as the preview of 道貌岸然) and “乎/hu” (e.g.,  
11  
12 道貌乎然 as the preview of 道貌岸然), respectively. The characters in the control  
13  
14 conditions were chosen to investigate whether the characteristics of “的/de” other than  
15  
16 its extremely high frequency, specifically, its visual complexity (e.g. number of  
17  
18 strokes) and its grammatical properties (e.g. a character that is very frequently used as  
19  
20 a function word), may enhance skipping. The control conditions are very similar to  
21  
22 “的/de” in most aspects, but there are some small differences (e.g. “钓/diao” is a verb  
23  
24 and “乎/hu” is visually simpler than “的/de”). Because of this, the findings of the  
25  
26 control conditions should be interpreted with caution.  
27  
28  
29  
30  
31  
32

33  
34 We selected the four-character target words from the Chinese Linguistic Data  
35  
36 Consortium (2003) corpus considering the following factors: First, we ensured that  
37  
38 the first two characters of each target word cannot constitute a two-character word so  
39  
40 that substituting a different character for the target character would always result in a  
41  
42 lexical violation. Second, the mean stroke number of the target characters ( $M=7.95$ ;  
43  
44  $SD=2.41$ ) was matched to the number of strokes of the character “的/de” (which is 8).  
45  
46 Third, the mean character frequency (occurrences per million) of the target characters  
47  
48 ( $M=192.51$ ;  $SD=164.96$ ) was much lower than the character frequency of the  
49  
50 character “的/de” (which is 22797) and no lower than that of the character “钓/diao”  
51  
52 or “乎/hu” (which are 23 and 162, respectively).  
53  
54  
55  
56  
57  
58  
59  
60

## Procedure

Following a three-point calibration procedure, participants were asked to read each sentence silently for comprehension. Participants read seven practice sentences and then 80 experimental sentences intermixed with 80 filler sentences. The gaze-contingent boundary paradigm was used to present the stimuli. When the eyes of participants crossed an invisible boundary located left of the target character, the preview character was replaced by the target character on the screen. After the experiment, participants were asked whether or not they noticed anything unusual during this experiment. No participant reported that they noticed the boundary manipulation during the experiment. Among the experimental and filler sentences, 54 sentences were followed by a forced-choice question to ensure that participants read the sentences carefully.

## Results

The accuracy of the comprehension questions was high (96.2 %), indicating that participants read and understood the sentences well. We eliminated 2.8% of the trials for blinks or track loss, 22.2% of the trials for display changes that completed either late (after fixation onset) or extremely early (50 ms earlier than fixation onset, which is quite rare and may be triggered by very infrequent long saccades), and 3.3% of the trials that were triggered by saccades that terminated to the left of the boundary (“j-hooks”). In sum, 28.3% of the data were eliminated and data from a total of 1606 trials were included in the following analysis.

We analyzed skipping probability and fixation times on the target character and

1  
2  
3  
4 the pre-target region (the two characters before the boundary). In addition, the launch  
5  
6 site of the saccades that landed at the target character was also calculated. The  
7  
8 analyses were conducted using linear mixed models (LMMs), which were fitted using  
9  
10 the *lmer* function from the lme4 package (Version 1.1-7) in the R Environment for  
11  
12 Statistical Computing (Version 3.3.1). The preview condition was included as a fixed  
13  
14 factor and three treatment contrasts (“的/de” vs. identity; “的/diao” vs. identity; “乎  
15  
16 /hu” vs. identity) were used to test for the preview effect. Before reporting specific  
17  
18 LMM preview contrasts, we conducted an omnibus likelihood-ratio test for the effect  
19  
20 of preview to examine the significance of the main effect. We reported the three  
21  
22 contrasts between identity preview condition and other preview conditions only when  
23  
24 the main effect was significant. For skipping probability, we report regression  
25  
26 coefficients (*b*), standard errors (*SE*), and *z*-values from generalized LMMs using a  
27  
28 logit-link. For fixation times, which include *first fixation duration* (*FFD*; the duration  
29  
30 of the first first-pass fixation on the target character) and *gaze duration* (*GD*; the sum  
31  
32 of all fixations on the target character before leaving it, including refixations), we  
33  
34 calculated regression coefficients (*b*), standard errors (*SE*), and *t*-values from LMMs  
35  
36 on the basis of raw fixation time data. We ran LMMs with participant and item as  
37  
38 crossed random effects. Including random slopes for the preview condition over  
39  
40 participants and items did not improve the model fit. We do not report *p*-values since  
41  
42 it is not clear how to compute the degrees of freedom for *t*-values testing the effect of  
43  
44 coefficients in LMMs. However, since we have relatively high numbers of subjects  
45  
46 and items, the *t*-values should be approximately normally distributed and can be  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4 interpreted like  $z$ -values. Therefore, we use the same two-tailed criterion  $|z|$  or  $|t| \geq 1.96$   
5  
6 for  $z$  and  $t$ -values in order to test for significance test at the 0.05  $\alpha$ -level. The mean  
7  
8 skipping probability, fixation times, and launch site of the four preview conditions are  
9  
10 shown in Table 2 and the results of the LMM analyses are shown in Table 3.

### 14 Pre-target character

16 The result of the omnibus test on FFD, expressed as a chi-squared statistic,  
17  
18 showed that the model including the preview condition as a factor provided a  
19  
20 significantly better fit to the data than the model not including it,  $\chi^2(3)=11.51$ ,  
21  
22  $p=0.009$ . However, the contrasts between identity preview condition and other  
23  
24 preview conditions showed no significant difference (“的/de” vs. identity: 259ms vs.  
25  
26 253ms,  $b=7.65$ ,  $SE=6.71$ ,  $t=1.14$ ; “钓/diao” vs. identity: 243ms vs. 253ms,  $b=-9.17$ ,  
27  
28  $SE=6.73$ ,  $t=-1.36$ ; “乎/hu” vs. identity: 262ms vs. 253ms,  $b=12.37$ ,  $SE=6.64$ ,  $t=1.86$ ).  
29  
30  
31 The results of the omnibus test on skipping probability and GD were nonsignificant  
32  
33 (skipping probability:  $\chi^2(3)=1.62$ ,  $p=0.66$ ; GD:  $\chi^2(3)=7.42$ ,  $p=0.06$ ).  
34  
35  
36  
37  
38

### 39 Target character

41 The results of the omnibus test on skipping probability, FFD, and GD were all  
42  
43 significant (skipping probability:  $\chi^2(3)=10.90$ ,  $p=0.012$ ; FFD:  $\chi^2(3)=16.49$ ,  $p<0.001$ ;  
44  
45 GD:  $\chi^2(3)=9.80$ ,  $p=0.020$ ) while the result on launch site was nonsignificant,  $\chi^2(3)$   
46  
47  $=5.26$ ,  $p=0.154$ . Readers skipped the previews of “的/de” more often than the  
48  
49 identical previews (0.65 vs. 0.58;  $b=0.37$ ,  $SE=0.15$ ,  $z=2.44$ ). The other preview  
50  
51 conditions show no significant difference from the identity condition (“钓/diao” vs.  
52  
53 identity: 0.58 vs. 0.58;  $b=0.04$ ,  $SE=0.15$ ,  $z=0.24$ ; “乎/hu” vs. identity: 0.61 vs. 0.58;  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  $b=0.23$ ,  $SE=0.15$ ,  $z=1.53$ ). In order to account for any possible effects of target word  
5  
6 frequency, we further conducted a supplementary analysis with both target word  
7  
8 frequency and the preview manipulation as fixed effects. The result pattern was highly  
9  
10 consistent with that of the main analysis (“的/de” vs. identity: 0.65 vs. 0.58;  $b=0.41$ ,  
11  
12  $SE=0.15$ ,  $z=2.68$ ; “钧/diao” vs. identity: 0.58 vs. 0.58;  $b=-0.01$ ,  $SE=0.15$ ,  $z=-0.07$ ;  
13  
14 “乎/hu” vs. identity: 0.61 vs. 0.58;  $b=0.26$ ,  $SE=0.15$ ,  $z=1.70$ ).

15  
16  
17  
18  
19 In all display change experiments where the preview condition strongly  
20  
21 influences skipping probability, it is important to keep in mind that the fixation time  
22  
23 measures on the target word or character reflect only the (minority of) cases where the  
24  
25 target word or character was fixated and not skipped. When the target character was  
26  
27 fixated, readers' FFD and GD on the target character were shorter in the identity  
28  
29 condition than in the “钧/diao” and “乎/hu” conditions (FFD: “钧/diao” vs. identity:  
30  
31 245ms vs. 223 ms;  $b=29.94$ ,  $SE=10.30$ ,  $t=2.91$ ; “乎/hu” vs. identity: 243 ms vs.223  
32  
33 ms;  $b=23.66$ ,  $SE=11.48$ ,  $t=2.06$ ; GD: “钧/diao” vs. identity: 253 ms vs. 224 ms;  
34  
35  $b=31.99$ ,  $SE=8.53$ ,  $t=3.75$ ; “乎/hu” vs. identity: 247ms vs.224ms;  $b=29.53$ ,  $SE=8.82$ ,  
36  
37  $t=3.35$ ). There was no significant difference between the “的/de” and identity preview  
38  
39 conditions in FFD and GD (FFD: 232ms vs. 223ms,  $b=9.57$ ,  $SE=9.15$ ,  $t=1.05$ ; GD:  
40  
41 232ms vs. 224ms,  $b=9.72$ ,  $SE=9.00$ ,  $t=1.08$ ).

## 42 43 44 45 46 47 48 49 **Discussion**

50  
51 We investigated the influence of substituting the preview of the third character of  
52  
53 a four-character word with the character “的/de” on skipping probability. The preview  
54  
55 of the character “的/de” significantly enhanced the skipping probability in comparison  
56  
57  
58  
59  
60

1  
2  
3  
4 with the unsubstituted preview, despite its appearance in the preview resulting in  
5  
6 lexical violations at word level. This suggests that readers may decide to skip a  
7  
8 previewed character mainly based on how easily it can be processed at character level.  
9  
10  
11 Word-level processing does not seem to play a large role in skipping decisions, if any.  
12

13  
14 Our results support the character-based skipping hypothesis, and did not support  
15  
16 the word-based skipping hypothesis. The word-based skipping hypothesis assumes  
17  
18 that the skipping decision is based on successful word-level processing associated  
19  
20 with the preview character. In the experimental preview conditions of the current  
21  
22 study, substituting the target character with the unrelated characters “的/de”, “  
23  
24 /diao”, or “乎/hu” should disrupt the processing of the target word, which, according  
25  
26 to the word-based skipping hypothesis, should result in a reduction of skipping  
27  
28 probability. However, this was not the case. Readers showed higher skipping  
29  
30 probability in the “的/de” condition than in the identity condition and showed no  
31  
32 significant difference in skipping probability between the identity, “  
33  
34 /diao”, and “乎  
35  
36 /hu” conditions.  
37  
38  
39

40  
41 On the other hand, the results are consistent with the predictions of the “character  
42  
43 based” skipping hypothesis. The character-based skipping hypothesis assumes that  
44  
45 Chinese readers decide to skip a previewed character based on character level  
46  
47 processing and before integrating it into a word. Thus, they predicts that a character  
48  
49 will be more likely to be skipped the easier it can be processed (just as the character  
50  
51 “的/de” in the current study). This is exactly what we observed in the current study.  
52  
53  
54

55  
56 The results of the current study is consistent with the processing-based  
57  
58  
59  
60

1  
2  
3  
4 hypothesis of eye movement control (Li et al., 2011; Wei, Li, & Pollatsek, 2013),  
5  
6 which assumes that Chinese readers plan their saccades to process as much  
7  
8 information as possible at a given fixation and move their eyes beyond the recognized  
9  
10 characters. Thus, if the characters in the parafovea are easy to process, readers should  
11  
12 be able to process more characters and, as a consequence, make longer saccades. In  
13  
14 the current study, the character “的/de” is easy to process, thus it is more likely to be  
15  
16 processed with parafoveal vision, and thus is more likely to be skipped.  
17  
18  
19

20  
21 Our results also provided evidence on the relationship between parafoveal  
22  
23 processing and expectations of what the next character should be based on lexical  
24  
25 representations (the character-level expectation) in terms of skipping decisions: We  
26  
27 found that automatic skipping of “的/de” previews occurred even though “的/de” was  
28  
29 incompatible with the character-level expectation based on the preceding characters of  
30  
31 a same word. In our experiment, the preceding characters were highly predictive of  
32  
33 the target character – in fact, for each target word, the third character was the only  
34  
35 legal continuation after the first two characters. Therefore, the preview character “的  
36  
37 /de” violated not only the word knowledge of readers but also went against any  
38  
39 expectations that they might have formed for the preview character based on the  
40  
41 preceding characters. This result is to some extent similar to Abbott et al.’s (2015)  
42  
43 results, who found that the-skipping was not reduced in cases where the target word  
44  
45 was highly predictable. Therefore, it seems that in both English and Chinese reading,  
46  
47 parafoveal processing overrides violations of expectation (at character level or at word  
48  
49 level) when it comes to skipping decisions. Note that this does not mean that there is  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 no effect of expectations and predictability at all, but when the expectations based on  
5  
6 previous processing and the results from current parafoveal processing disagree, it  
7  
8 seems that the skipping decision is determined by the parafoveal preview rather than  
9  
10 previously generated expectations.  
11  
12

13  
14 Finally, we used two additional control conditions to test what type of features of  
15  
16 “的/de” other than its extremely high frequency may explain the increase in skipping  
17  
18 probability. The preview of the character “鈞/diao”, whose visual complexity is  
19  
20 highly similar to that of the character “的/de”, did not enhance the skipping  
21  
22 probability in comparison with the unsubstituted preview, indicating the  
23  
24 skipping-probability enhancement for “的/de” is not due to its low visual complexity.  
25  
26 There was a non-significant increase in skipping probability when the preview was  
27  
28 the function-word character “乎/hu” compared to the identical preview condition. If  
29  
30 this effect was significant, it might be taken to indicate that the grammatical  
31  
32 properties of “的/de” may contribute to the tendency of readers to skip it. However, it  
33  
34 should be noted that the relative high skipping probability of “乎/hu” may also result  
35  
36 from its visual simplicity: it is a single-component character with fewer strokes than  
37  
38 “的/de” (6 vs. 8) and the actual target characters (6 vs. 7.95;  $t(79)=10.96$ ,  $p < 0.001$ ).  
39  
40  
41  
42  
43  
44  
45

46  
47 In addition to the skipping probability, we also analyzed the fixation times of the  
48  
49 pre-target region and target characters. For the pre-target region, the conditions with  
50  
51 incorrect previews did not show longer fixations than the identity preview condition  
52  
53 did. Therefore, the observed skipping probability differences between the “的/de”  
54  
55 preview condition and the identity preview condition on the target characters cannot  
56  
57  
58  
59  
60

1  
2  
3  
4 be explained according to the fixation time on the pre-target region. The FFDs and  
5  
6 GDs on the target character was significantly longer in the “钓/diao” and “乎/hu”  
7  
8  
9 conditions than in the identity condition. However, there was no significant difference  
10  
11 in fixation times between the identity and “的/de” conditions. If the difference is due  
12  
13 to a classic preview benefit effect, it should be equal across the three comparisons  
14  
15 between the identity condition and the conditions with display change. Because of this,  
16  
17 the fixation-time difference between conditions might mainly reflect disruptions from  
18  
19 incorrect preview information rather than preprocessing of the identical character.  
20  
21 Because the character “的/de” has extremely high frequency, its processing might be  
22  
23 more automatic and attract less attention than the processing of “钓/diao” and “乎/hu”  
24  
25 and thus resulted in less disruption. Risse and Kliegl (2014) observed a similar effect  
26  
27 where the frequency of a preview word influenced fixation times after crossing the  
28  
29 boundary.  
30  
31  
32  
33  
34  
35

36 In summary, our results are compatible with the hypothesis that skipping of  
37  
38 high-frequency characters such as “的/de” is triggered by character-level preview  
39  
40 processing rather than by word-level processing. The preview of the character “的/de”  
41  
42 led to an increased skipping probability despite the word-level lexical violations it  
43  
44 causes and despite its very low predictability. Our findings indicate that Chinese  
45  
46 readers frequently make the decision of skipping a high-frequency character before  
47  
48 integrating it into a word.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- Abbott, M. J., Angele, B., Ahn, Y. D., & Rayner, K. (2015). Skipping syntactically illegal the previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1703-1714.
- Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high- and low-frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1181-1203.
- Angele, B., & Rayner, K. (2013). Processing the in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 649-662.
- Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). Hillsdale, NJ: Academic Press.
- Chinese Academy of Social Sciences (2012). 现代汉语词典(第六版) [The Contemporary Chinese Dictionary], 6th Edn. Beijing: The Commercial Press.
- Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation.

- 1  
2  
3  
4 Inhoff, A. W., & Liu, W. (1997). The perceptual span during the reading of Chinese  
5  
6 text. In H.-C. Chen (Ed.), *The cognitive processing of Chinese and related*  
7  
8 *Asian languages*. Hong Kong: Chinese University Press.
- 9  
10  
11 Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during  
12  
13 the reading of Chinese sentences. *Journal of Experimental Psychology: Human*  
14  
15 *Perception and Performance*, 24(1), 20-34.
- 16  
17  
18 Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally  
19  
20 similar across disparate writing systems: a systematic characterization of how  
21  
22 words and characters influence eye movements in Chinese reading. *Journal of*  
23  
24 *Experimental Psychology: General*, 143(2), 895-913.
- 25  
26  
27  
28  
29 Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is  
30  
31 there a preferred viewing location? *Vision Research*, 51(10), 1146–1156.
- 32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The  
effect of visual complexity and word frequency on eye movements during  
Chinese reading. *Visual Cognition*, 22(3-4), 441–457.
- Ma, G., & Li, X. (2015). How character complexity modulates eye movement  
control in Chinese reading. *Reading and Writing*, 28(6), 747-761.
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive  
Psychology*, 7(1), 65-81.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years  
of research. *Psychological Bulletin*, 124(3), 372-422.



- 1  
2  
3  
4 Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control,  
5  
6 serial-attention model of eye-movement behavior during reading. *Cognitive*  
7  
8 *Systems Research*, 7(1), 4-22.
- 9  
10  
11 Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty  
12  
13 in parafoveal processing of word  $n + 1$  during reading. *Journal of Experimental*  
14  
15 *Psychology: Human Perception and Performance*, 40(2), 653–668.
- 16  
17  
18 Sun, H., Huang, J., Sun, D., Li, D., & Xing, H. (1997). Introduction to language  
19  
20 corpus system of modern Chinese study. In M. Hu (Ed.), *Paper collection for*  
21  
22 *the fifth world Chinese teaching symposium* (pp. 459–466). Beijing, China:  
23  
24 Peking University Press.
- 25  
26  
27  
28  
29 Tsai, C. H., & McConkie, G. W. (1995). *The perceptual span in reading Chinese text:*  
30  
31 *A moving window study*. Paper presented at the Seventh International  
32  
33 Conference on the Cognitive Processing of Chinese and Other Asian Languages,  
34  
35 Hong Kong.
- 36  
37  
38  
39 Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect  
40  
41 outgoing saccade length in Chinese reading. *Vision Research*, 80, 1-6.
- 42  
43  
44 Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S.P. (in press). The  
45  
46 skipping of the structural particle de (的) in Chinese reading. *The Quarterly*  
47  
48 *Journal of Experimental Psychology*.
- 49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1. An example of experimental stimuli of the four preview conditions.**

Preview conditions	Character frequency of the preview character	Stroke number of the preview character	Example sentences
identity	192.5 ± 165.0	8.0 ± 2.4	我欣赏他不人云 <b>亦</b> 云的勇气，有与众不同的胆识。
的(de)	22797 ± 0	8 ± 0	我欣赏他不人云 <b>的</b> 云的勇气，有与众不同的胆识。
钧(diao)	23 ± 0	8 ± 0	我欣赏他不人云 <b>钧</b> 云的勇气，有与众不同的胆识。
乎(hu)	162 ± 0	6 ± 0	我欣赏他不人云 <b>乎</b> 云的勇气，有与众不同的胆识。

*Note:* The character frequency and stroke number were presented in form of Mean ± Standard Deviation. Character in bold are characters in the target region. The English translation of the example sentence is: I admire his distinctive courage and that he does not follow the herd. The vertical black line represents an invisible boundary. When the eyes crossed the boundary, the preview character was replaced by the target character.

**Table 2. Condition means for the pre-target region and the target character**

	<u>Type of preview</u>			
	identity	的(de)	钓(diao)	乎(hu)
Pre-target region				
Skip probability	21% (3.2%)	24% (3.2%)	24% (4.1%)	21% (3.2%)
First fixation duration	253 (8.13)	259 (8.74)	243 (8.65)	262 (8.01)
Gaze duration	282 (11.96)	285 (10.26)	263 (10.46)	280 (10.48)
Target character				
Skip probability	58% (4.2%)	65% (2.8%)	58% (3.7%)	61% (3.4%)
First fixation duration	223 (7.00)	232 (8.46)	245 (8.11)	243 (9.64)
Gaze duration	224 (7.12)	232 (8.46)	253 (10.80)	247 (9.54)
Launch site	2.53 (0.16)	2.71 (0.16)	2.50 (0.17)	2.45 (0.13)

*Note.* Standard errors are in parentheses.

Table 3. Linear mixed effects model analyses for the pre-target region and the target character

Effect	Skip probability			First fixation duration			Gaze duration			Launch site		
	<i>b</i>	<i>SE</i>	<i>z</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>
Pre-target region												
de vs. identity	0.11	0.18	0.64	7.65	6.71	1.14	5.17	9.43	0.55			
diao vs. identity	0.15	0.18	0.82	-9.17	6.73	-1.36	-18.63	9.45	<b>-1.97</b>			
hu vs. identity	-0.05	0.18	-0.30	12.37	6.64	1.86	1.85	9.36	0.20			
Target character												
de vs. identity	0.37	0.15	<b>2.44</b>	9.57	9.15	1.05	9.72	9.00	1.08	0.16	0.10	1.56
diao vs. identity	0.04	0.15	0.24	29.94	10.30	<b>2.91</b>	31.99	8.53	<b>3.75</b>	-0.04	0.09	-0.43
hu vs. identity	0.23	0.15	1.53	23.66	11.48	<b>2.06</b>	29.53	8.82	<b>3.35</b>	-0.06	0.10	-0.63

Note. Each column represents a model fit to one of the dependent variables. SP=skipping probability; FFD=first fixation duration; GD=gaze duration; *b*=regression coefficient; *SE*=standard error; *t* or *z*=test statistic (*b/SE*). Cells marked in bold represent  $|t| \geq 1.96$ .