

Object-based attention in Chinese readers of Chinese words: Beyond Gestalt principles

XINGSHAN LI AND GORDON D. LOGAN
Vanderbilt University, Nashville, Tennessee

Most object-based attention studies use objects defined bottom-up by Gestalt principles. In the present study, we defined objects top-down, using Chinese words that were seen as objects by skilled readers of Chinese. Using a spatial cuing paradigm, we found that a target character was detected faster if it was in the same word as the cued character than if it was in a different word. Because there were no bottom-up factors that distinguished the words, these results showed that objects defined by subjects' knowledge—in this case, lexical information—can also constrain the deployment of attention.

Historically, covert visual attention was thought to be space based, so that attention was focused on regions of space, rather than on the objects in the display. Whether an object was selected was determined by its distance from the focus of attention. Space-based attention was modeled as spotlights (B. A. Eriksen & C. W. Eriksen, 1974; Posner, 1980), zoom lenses (C. W. Eriksen & St. James, 1986), and gradients (Downing & Pinker, 1985). Subsequently, two lines of evidence have suggested that attention can also be object based. One line of evidence comes from the use of divided-attention paradigms, showing that two attributes of one object can be recognized more rapidly and accurately than two attributes of different objects (Awh, Dhaliwal, Christensen, & Matsukura, 2001; Duncan, 1984; Kramer, Weber, & Watson, 1997; Lee & Chun, 2001; Vecera, 1997; Vecera & Farah, 1994). The other line of evidence comes from spatial-cuing paradigms, showing that observers are faster when targets and cues are in the same object than when they are in different objects (Egly, Driver, & Rafal, 1994; Moore, Yantis, & Vaughan, 1998; Pratt & Sekuler, 2001; Vecera, 1994). Although there are different interpretations of these two lines of evidence, they support the view that the organization of displays into objects can influence attention deployment.

Although much evidence supports object-based attention, what an object is or how an object is defined is still uncertain (Logan, 1996). Most work on object-based attention uses Gestalt principles, such as continuation, collinearity, similarity, or common fate, to define objects, which are usually considered to be bottom-up factors. In a classical demonstration of object-based attention, Egly et al. (1994) used two parallel rectangles, arrayed horizontally or vertically, to contrast space-based and object-based attention. As is illustrated in Figure 1A, a cue was

presented on one end of a rectangle, and then a target was presented at the cued location, at the other end of the cued rectangle, or on one end of the other rectangle. Egly et al. found that reaction times (RTs) were faster when the target was at the cued location than when it was at uncued locations, which supported space-based attention. In a critical comparison of the invalid-cue conditions, they found that RTs were faster when the target was in the cued rectangle than when the target was in the other rectangle, although the distance from the cue and the target was the same in these two conditions. This result provided strong support for object-based attention when objects were defined bottom-up. Although most object-based attention studies used spatially connected objects, some studies (e.g., Dodd & Pratt, 2005) showed that spatially separate items could also produce an object-based effect when perceptually grouped together.

Objects can not only be defined bottom-up by Gestalt principles, but also be defined top-down. Robertson and Treisman (2006) found that a patient with Balint's syndrome, who could only perceive single objects, could identify familiar words (ON and NO) but not the relative location of the two letters ("o" and "n") in the display. They took this to suggest that familiar words could be perceived as objects. The present study was designed to explore whether objects defined top-down by subjects' lexical knowledge could constrain the deployment of attention.

Some studies have shown that top-down factors such as instruction and experience could affect the perception of objects and hence influence the deployment of attention. Chen (1998) found that the object effect depended on instruction. In her study, subjects were shown two different colored Vs that formed an X-shaped stimulus. One

X. Li, xingshan.li@vanderbilt.edu

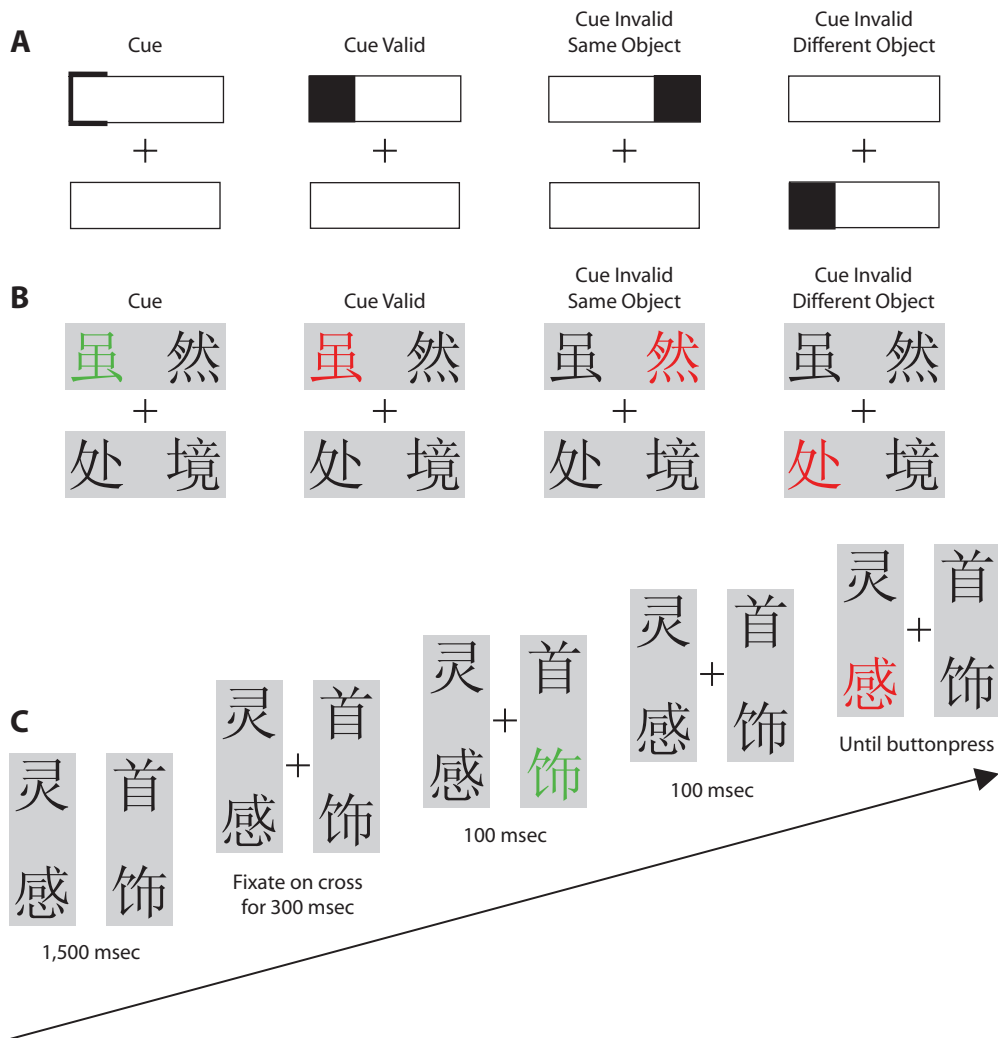


Figure 1. (A) The paradigm of Egly, Driver, and Rafal (1994). (B) The paradigm of the present study. (C) The procedure of the experiment. The shades in panels B and C represent the location of the words, which were not shown in the real stimuli.

group of subjects was instructed to interpret the stimuli as two Vs; the other group was instructed to interpret the stimuli as one X. If the stimulus was interpreted as two Vs, switching attention from one arm of a V to one arm of the other V was slower than switching attention between two arms within a V. There was no difference if the stimulus was interpreted as an X. Zemel, Behrmann, Mozer, and Bavelier (2002) found an object benefit if subjects learned the shape of the objects in the previous blocks; they found no object benefit if they had not learned the object shape. The stimuli in these studies were objects whose parts were spatially connected. The stimuli used in the present study differed from those in these studies in that they were not interconnected and would not be grouped as objects only on the basis of bottom-up factors.

We used the same paradigm as did Egly et al. (1994), but we used Chinese words instead of rectangles to define objects (Figure 1B), and we tested skilled Chinese readers. We presented four Chinese characters that con-

stituted two words, arrayed horizontally or vertically. If skilled Chinese readers perceived these words as objects, RTs in cue-invalid trials should be shorter when the target was in the cued word (invalid same-word condition) than when it was in the other word (invalid different-word condition). Each Chinese character is a single perceptual object defined by Gestalt principles. The strokes are usually connected or are very close to each other, and the characters are separated by small spaces. Thus, bottom-up processes would produce four small objects or one large object in each condition of the experiment. Previous Egly-style studies that presented objects in the four corners of a square did not find any object-based effect (Pratt & Sekuler, 2001). Hence, the word advantage was not expected in the present study if there was only bottom-up grouping. Note that the contrast between “bottom-up” and “top-down” is ambiguous in the literature. On the one hand, *top-down* can mean anything that is not purely perceptual, so knowledge can count as a

top-down influence no matter how it is deployed. On the other hand, top-down can also mean voluntary, in which case automatically applied knowledge is considered bottom-up. We use the first definition of top-down factors in this article.

Chinese words are ideal for our purpose. Words are the basic lexical units in Chinese, and there are no spaces between words in regular reading, so skilled Chinese readers may be accustomed to imposing top-down perceptual organization on texts that do not allow bottom-up segmentation. Moreover, two-character words are most common in Chinese. We assumed that skilled Chinese readers would combine pairs of characters to form words in a top-down fashion, overriding bottom-up factors. We ensured that subjects would group the displays as words by requiring them to identify the words in multiple-choice tests that followed one fourth of randomly selected trials.

METHOD

Subjects

Twenty native Chinese speakers, who were graduate students or spouses of students at Vanderbilt University, were paid to participate in the experiment. They all had normal or corrected-to-normal vision. All subjects had finished at least their college education in China, so all of them were fluent Chinese readers.

Apparatus

Stimuli were presented on a 21-in. CRT monitor controlled by a Dell PC. A chinrest located 57 cm from the monitor was used to minimize head movements. Eye movements were monitored via an EyeLink 2 tracker with eye position sampled at 250 Hz. The primary goal of tracking the eyes was to make certain that the subjects maintained fixation in the center of the display throughout each trial. Drift correction was conducted at the beginning of each trial.

Material

Four characters were shown in each trial. They appeared in two rows and two columns. In half of the trials, the two characters in each row constituted a word; in the other half, the two characters in each column constituted a word. All of the words were two characters long, with an average frequency of 88 (range, 20–1,000) per 1,116,417 words.¹ None of the words was used more than once in the experiment. There was no significant difference in word frequency between these two conditions. Special care was taken to make sure that the two words in a trial did not make sense when combined together. None of the characters other than the two words constituted a two-character word. Chinese characters can be combined with other characters to form many words of different lengths. Many characters are one-character words by themselves. The segmentation of Chinese words usually depends on context. In the present study, we encouraged subjects to segment the characters into two-character words by explicitly telling them that the four characters always constituted 2 two-character words and by asking them to do a memory task on these two-character words. All of the characters were shown in black on a white background. Each character occupied an 8 × 8 mm square (which subtended about 0.8° visual angle), with 8 mm between rows and columns. The stimuli were shown at the center of the screen.

Procedure

There were 768 experimental trials, equally divided into two sessions. Sixteen of the 20 subjects chose to finish the sessions in 1 day, which took about 2 h. For these subjects, there were 48 practice trials at the beginning of the first session. Four subjects chose to finish the two sessions on different days. There were 48 practice trials at the

beginning of each session for these subjects. The order of trials was randomized within a session.

In each trial, subjects were shown the four characters for 1,500 msec after drift correction in order to make sure that subjects had enough time to segment the words (Figure 1C). Then, a black cross was shown at the center of the four characters. Subjects were asked to fixate this cross. Subjects fixated the cross for 300 msec; then the cue character turned green for 100 msec. This cue was presented at the four character locations equally often. The cued character turned back to black for 100 msec, after which the target character turned red in 640 of 768 trials. The stimuli were presented until subjects responded by pressing a button on a Microsoft SideWinder Game Pad when a character turned red. They were told to do nothing on the 128 catch trials during which none of the characters turned red. They were told that RTs were recorded but that accuracy was more important. A trial was aborted and subjects were warned by a line of text (“you moved your eyes”) on the screen if they moved their eyes between the cue presentation and the target presentation. An error response was signaled by a sound (Utopia Program Error clip, Microsoft Windows 2000) and a line of text (“wrong response!”) on the computer monitor. After the practice trials and after each block of 96 experiment trials, there was a break. The target character appeared at the cued location in 75% of target-present trials. When the cue was invalid, the target was presented equally often in the same word as the cued character (invalid same-word condition, in 12.5% of target-present trials) and in the other word (invalid different-word condition, in 12.5% of target-present trials) in a position equidistant from the cued character (Egley et al., 1994). The target was never the character diagonal to the cued character.

To make sure that subjects read the words, they were given a multiple-choice test after one fourth of the trials. During the multiple-choice tests, the subjects were shown 4 two-character words, from which they were asked to choose one that had appeared on the previous trial. None of the alternative words, except the key word, was used during the experimental trials or the practice trials. The two words in the experimental trials were equally likely to be tested. Error responses were indicated by a warning sound (tada.wav, Microsoft Windows 2000) and a sentence on the display (“you made the wrong choice”).

RESULTS

Word Comprehension

Subjects chose the right words in the multiple-choice questions in 99% of trials, which indicated that they grouped the characters into words.

Eye Movements

Subjects made eye movements between the presentation of the cue and the presentation of the target in 4.3% of the trials. These trials were aborted. There was no difference in the number of aborted trials between the horizontal- and the vertical-word conditions ($p > .1$).

Response Accuracy

The hit rate (a response when the target was present) was 99.8%, and the false alarm rate (a response when the target was not present) was 7.9%. The false alarm rates for the horizontal- and the vertical-word conditions did not differ significantly ($p > .1$).

RTs

Only target-present trials with a correct response and without eye movements were included in the analysis. Tri-

als with RTs shorter than 150 msec were considered to be anticipations and were not included in the analysis. Trials with RTs more than three standard deviations from the cell mean were excluded from analysis (1.8% of trials). Altogether, 6.3% of the target-present trials were excluded from analysis.

Mean RTs were submitted to a 2 (row: the row in which the target was presented) × 2 (column: the column in which the target was presented) × 3 (condition: cue valid, invalid same word, invalid different word) × 2 (orientation of words) within-subjects ANOVA. As can be seen in Figure 2, RTs on cue-valid trials ($M = 446$ msec; $SE = 22$ msec) were much shorter than RTs on invalid same-word trials ($M = 521$ msec, $SE = 31$ msec) and on invalid different-word trials ($M = 529$ msec; $SE = 31$ msec) [$F(2,38) = 45.26, p < .0001, \eta_p^2 = .70$]. A planned contrast showed that the 8-msec difference between RTs in the invalid same-word condition and in the invalid different word condition was significant [$F(1,19) = 6.57, p < .05, \eta_p^2 = .26$], indicating a top-down organization of the displays into objects defined by words. This difference is similar to the 13-msec difference observed by Egly et al. (1994) for bottom-up objects. There was a three-way interaction of row, column, and word orientation [$F(1,19) = 9.00, p < .01, \eta_p^2 = 0.32$]. As is shown in Figure 3, RTs were shorter when the probe was at the bottom right character than when it was at other locations in the horizontal-word condition. In contrast, RTs were longer when the probe was at the top left character than when it was at other locations in the horizontal-word condition. This interaction did not compromise the comparison between invalid same-word and invalid different-word conditions.

To test whether the object-based effect was influenced by word orientation, mean RTs (Figure 2) were submitted to a 2 (condition: invalid same word, invalid different word) × 2 (orientation of words: horizontal, vertical) within-subjects ANOVA. There was a main effect of condition [$F(1,19) = 7.01, p < .05, \eta_p^2 = .27$], but no interaction between condition and orientation of words ($p > .1$).

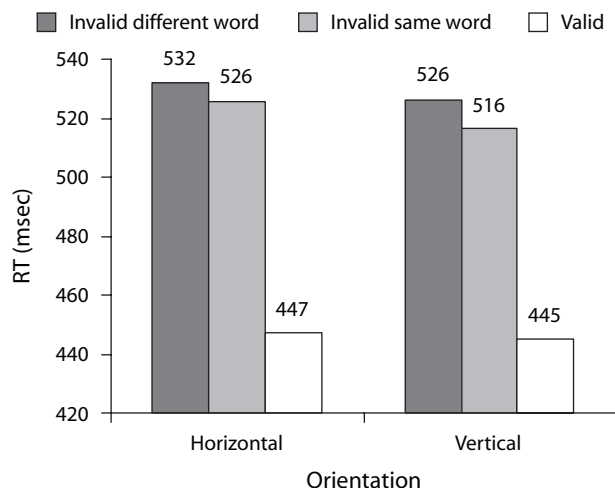


Figure 2. Reaction times (RTs, in milliseconds) of the valid, invalid same-word, and invalid different-word conditions.

Horizontal Words

503 (27)	505 (27)
+	
504 (31)	495 (30)

Vertical Words

505 (28)	492 (28)
+	
492 (26)	493 (27)

Figure 3. Reaction times (RTs, in milliseconds) when the target was at different locations. Left panel, the word was horizontal; right panel, the word was vertical. The numbers in parentheses were the standard errors of the RTs in that location (in milliseconds).

DISCUSSION

Using the same paradigm as that used in Egly et al. (1994), we found an object-based attention effect when the objects were defined top-down as Chinese words. When the cue was invalid, RTs were shorter when the target was in the same word as the cue than when it was in the other word. This result provides new insight into the question of what defines an object in object-based attention. The displays were constructed so that Gestalt grouping principles could not be used to organize the display into words. If there were only bottom-up grouping, no word advantage would have been observed. Thus, our results show that top-down factors can define objects and constrain the deployment of attention.

Note that the object effect found in the present study was not due to bottom-up spatial constraints on attention. First, we did not find any evidence that the object effect was affected by the location of the target or by the orientation of the words. The crucial comparison occurred between the invalid same-word condition and the invalid different-word condition. Although there was an interaction between target location and word orientation, this interaction did not compromise the comparison between the invalid same-word and invalid different-word conditions. Hence, the invalid different-word condition was a sufficient baseline for the object-based effect in this study. Second, Pratt and Sekuler (2001) showed that four rectangles located at the four corners of an imaginary square did not introduce an object-based effect, which suggests that the object-based effect observed in the present study was not caused by spatial constraints.

Two possible mechanisms could explain the object-based effect observed in the present study. First, subjects might have grouped the characters into words because of the memory task. One fourth of the experimental trials were followed by explicit, intentional memory tests. This memory task might have made it necessary to group the characters into words. Second, the word-segmentation and word-recognition systems might have led subjects to group the characters into objects, regardless of the explicit memory task. In either case, lexical knowledge played an important role in grouping the characters into words, which constrained deployment of attention. It will be interesting to explore whether the lexical knowledge can be applied without a memory task.

The results of the present study raise some questions for future research. In the present study, we presented the stimuli at least 1 sec before the cue was presented, as did Egly et al. (1994), so subjects had enough time to segment the display into objects. It would be interesting to explore the time course of object segmentation by reducing the exposure duration before the presentation of the cue. Another possibility for future research concerns the interaction between top-down and bottom-up segmentation. In the present study, using bottom-up grouping, we segmented the stimuli into four objects (characters), whereas in top-down grouping, the stimuli are segmented into two objects (words). It would be interesting to pit top-down organization against bottom-up organization by increasing the spacing between characters that form words (grouping by proximity) or presenting characters in different words on the same background (common region) to reveal the relative strengths of various grouping processes.

AUTHOR NOTE

This research was supported by Grant R01-MH073879-01 from the National Institute of Mental Health and by Grant FA9550-07-1-0192 from the Air Force Office of Scientific Research. We thank Darryl Schneider for his helpful suggestions, and Zhe Chen, Jay Pratt, and two anonymous reviewers for their advice and helpful suggestions on an earlier draft of the article. Correspondence concerning this article should be addressed to X. Li, Department of Psychology, Vanderbilt University, Nashville, TN 37203 (e-mail: xingshan.li@vanderbilt.edu).

REFERENCES

- AWH, E., DHALIWAL, H., CHRISTENSEN, S., & MATSUKURA, M. (2001). Evidence for two components of object-based selection. *Psychological Science*, *12*, 329-334.
- CHEN, Z. (1998). Switching attention within and between objects: The role of subjective organization. *Canadian Journal of Experimental Psychology*, *52*, 7-17.
- DODD, M. D., & PRATT, J. (2005). Allocating visual attention to grouped objects. *European Journal of Cognitive Psychology*, *17*, 481-498.
- DOWNING, C. J., & PINKER, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 171-168). Hillsdale, NJ: Erlbaum.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501-517.
- EGLY, R., DRIVER, J., & RAFAL, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161-177.
- ERIKSEN, B. A., & ERIKSEN, C. W. (1974). Effects of noise letters upon identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143-149.
- ERIKSEN, C. W., & ST. JAMES, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, 225-240.
- KRAMER, A. F., WEBER, T. A., & WATSON, S. E. (1997). Object-based attentional selection—Grouped arrays or spatially invariant representations? Comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General*, *126*, 3-13.
- LEE, D., & CHUN, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, *63*, 253-257.
- LOGAN, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603-649.
- MOORE, C. M., YANTIS, S., & VAUGHAN, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, *9*, 104-110.
- NATIONAL LANGUAGES COMMITTEE (1997). Chinese dictionary: Character and word frequency statistics report. Taipei: Author.
- POSNER, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 2-25.
- PRATT, J., & SEKULER, A. B. (2001). The effects of occlusion and past experience on the allocation of object-based attention. *Psychonomic Bulletin & Review*, *8*, 721-727.
- ROBERTSON, L. C., & TREISMAN, A. (2006). Attending to space within and between objects: Implications from a patient with Balint's syndrome. *Cognitive Neuropsychology*, *23*, 448-462.
- VECERA, S. P. (1994). Grouped locations and object-based attention: Comment on Egly, Driver, and Rafal (1994). *Journal of Experimental Psychology: General*, *123*, 316-320.
- VECERA, S. P. (1997). Grouped arrays versus object-based representations: Reply to Kramer et al. (1997). *Journal of Experimental Psychology: General*, *126*, 14-18.
- VECERA, S. P., & FARAH, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*, 146-160.
- ZEMEL, R. S., BEHRMANN, M., MOZER, M. C., & BAVELIER, D. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 202-217.

NOTE

1. The source for word frequency was the Chinese Dictionary (National Languages Committee, 1997), in which the word frequency count is based on a corpus of 1,116,417 words.

(Manuscript received January 5, 2008;
revision accepted for publication April 4, 2008.)