

How Attentional Guidance and Response Selection Boost Contextual Learning: Evidence from Eye Movement

Chao Wang¹, Hanna Haponenko², Xingze Liu³, Hongjin Sun², and Guang Zhao¹

¹ Faculty of Psychology, Tianjin Normal University, Tianjin, Tianjin, China, 300387

² Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

³ Medical Psychological Center, Second Xiangya Hospital of Central South University, Hunan, China, 410011

ABSTRACT

The contextual cueing effect (CCE) refers to the learned association between predictive configuration and target location, speeding up response times for targets. Previous studies have examined the underlying processes (initial perceptual process, attentional guidance, and response selection) of CCE but have not reached a general consensus on their contributions to CCE. In the present study, we used eye tracking to address this question by analyzing the oculomotor correlates of context-guided learning in visual search and eliminating indefinite response factors during response priming. The results show that both attentional guidance and response selection contribute to contextual learning.

KEYWORDS

contextual cueing effect
eye movement
attentional guidance
response selection

INTRODUCTION

Visual contexts contain abundant visual information. This information could guide visual attention towards relevant targets embedded in the visual field. Recently, empirical evidence has shown that individuals respond faster to a target presented in a predictive (i.e., repeated) contextual configuration compared to a random configuration, even when they are unaware of the configuration repetition (Chun & Jiang, 1998; Jiang & Chun, 2001; Olson & Chun, 2001; Zang, Jia, Müller, & Shi, 2015; Jiang & Sisk, 2019). This phenomenon is called the contextual cueing effect (CCE). The possible underlying mechanism for this effect is that participants learned the association between predictive configurations and target locations (Chun & Jiang, 1998), and between distractors within the search context (Beesley, Vadillo, Pearson, & Shanks, 2015).

It is still unclear in what stage of visual processing—initial perceptual processing, attentional guidance, or response selection—the CCE occurs. To explore this issue, several behavioral studies have compared the search slope and intercept as a function of distractor set size on

target search reaction time (RT × set size function, Wolfe, 1998). The search slope indexes search efficiency (i.e., attentional guidance) and the intercept indexes non-search factors (i.e., initial perceptual processing and response selection, Wolfe, 1998; Wolfe, Vo, Evans, & Greene, 2011). Based on this definition, the CCE could be driven by an improved search efficiency, as evidenced by a shallower slope for predictive contexts compared to random contexts (Chun & Jiang, 1998). However, by measuring the intercept, researchers also found non-search factors (e.g., response selection) to be an important source of the CCE (Kunar, Flusberg, Horowitz, & Wolfe, 2007). In their study, significant differences were found between predictive and random configurations for intercepts instead of slopes. Predictive configurations enhanced priming for response selection. Moreover, the CCE still occurred in the single feature search task, where the target “popped out”

Corresponding author: Guang Zhao, Faculty of Psychology, Tianjin Normal University, Tianjin, Tianjin, China, 300387 E-mail: zhaoguang721@163.com

(i.e., was immediately identifiable) and the attentional guidance was already maximal suggesting that attentional guidance was not responsible (Kunar et al, 2007, Kunar, Flusberg, & Wolfe, 2008). Moreover, Schankin and Schubö (2009, 2010) used ERP recordings to distinguish between attentional guidance and response-related processes. Their results found that the behavioral CCE correlated significantly with the LRP component (an indicator of response-related motor processes). They concluded that both a more efficient attentional selection and faster response-related processes, as opposed to only attentional guidance, contributed to the CCE. Even so, there remains a debate on what specific mechanism underlies the CCE and at what stage of attentional processing.

Zhao et al. (2012) used eye tracking to index the precise processing stage of the CCE. By observing the substages of eye movements during a visual search task, the target search RT was partitioned into three consecutive phases: (a) the early phase, corresponding to initial perceptual processing; (b) the middle phase, corresponding to attentional guidance, and (c) the late phase, corresponding to response selection factors. Significant RT differences were found between predictive and random configurations in both the middle and the late phase, suggesting that attentional guidance and response selection contributed jointly to the CCE. However, Zhao et al.'s (2012) study had limitations, which we try to address with our study.

First, Zhao et al. (2012) analyzed the distribution of different types of eye saccades—ineffective saccades, effective saccades, and total saccades (i.e., the sum of ineffective and effective saccades)—in individual trials to examine how context guided attention. For each trial, effective saccades bring fixations increasingly closer to the target, while ineffective saccades direct eyes away from the target and “waste” effective saccades, as defined by Tseng and Li (2004). By continually integrating information across saccades, the indexes of different saccades were considered to indicate attentional guidance during search. Eye movement patterns related to attentional guidance are primarily evident in the number of saccades (or fixations) made (Peterson & Kramer, 2001). However, Zhao et al. (2012) failed to find a significant difference in slope between predictive/random configurations for both effective saccades and total saccades. The population data also revealed that there was a low statistical difference between predictive and random configurations contributed by effective saccades to the guidance of attention.

Second, Tseng and Li (2004) found that a reduction in ineffective saccades was the only factor that contributed to CCE. However, a higher proportion of trials (over 40%) in Zhao et al. (2012) showed the number of ineffective saccades centred at a distribution peak of zero. Their statistical analysis of ineffective saccades thus lacks power. The cause for the reduction in ineffective saccades may have reflected a general practice effect for learning random configurations instead of predictive configurations (i.e., rather than reflecting an advantage in learning predictive configurations). Therefore, the data analysis in Zhao et al. (2012) was not conducive for thoroughly explaining the extent of the CCE found in the middle phase.

To address this limitation, we used the iMap toolbox, packaged with inferential statistical resources, to analyze the fine-tuned oculo-

motor correlates of context-guided learning in visual search before and after contextual learning. iMap can generate a heatmap showing the distribution of participants' fixations and can calculate statistical differences between experimental conditions (Caldara & Miellet, 2011). Traditional analyses of eye movements rely on predefined regions of interest (ROIs), which results in discarding other eye movements data outside of the ROIs. In contrast, iMap does not require a priori segmentation of digital images to define ROIs and can incorporate all fixations (Caldara & Miellet, 2011; Lao, Miellet, Pernet, Sokhn, & Caldara, 2017). Furthermore, iMap provides a useful method of difference maps, which allows for comparisons, between different conditions or groups, of eye movement indices such as the number of fixations and fixation duration. For example, a difference map on fixation distributions between pre-learning of the CCE and post-learning of the CCE can be generated. Another advantage of iMap lies in its ability to avoid multiple comparison errors associated with the analysis of a large pixel space. For these reasons, the present study uses the iMAP toolbox to compare the eye fixation map in a data-driven way, before and after learning the CCE, to reveal how contextual learning affected fixation distributions.

The process of speeding up response priming should be triggered after the learning of contextual information. However, in Zhao et al. (2012), the late phase (response selection) differences between predictive and random configurations were significant even at the beginning of the learning phase. This counters the assumptions about the CCE because there should not have been any significant differences in the prelearning phase. One possible reason for this difference, however, might be that indefinite responses had not been removed in Zhao et al. (2012) when estimating the CCE from the last phase. Indefinite responses occur when observers fixate on the target during the search phase without explicitly identifying the target. During an indefinite response, a participant's eyes move farther away from the target to search for a nearby item, subsequently comparing the items to confirm original target identity. Indefinite responses involve extra eye fixations, which add noise to the estimation of time between the last eye fixation to the button press (TLFtoBP) during the late phase. The last eye fixation is defined as the fixation that is spatially close to the target, and is assumed to occur during the response selection phase (Tseng & Li, 2004; Zhao et al., 2012). Generally in Zhao et al. (2012), an eye fixation lasted around 200 ms, but the size of the CCE obtained from the late phase was only about 70 ms (Zhao et al., 2012). Therefore, adding an extra eye fixation to TLFtoBP because of indefinite responses confounded the contribution of TLFtoBP (i.e., response selection factors) to the estimation of CCE. Indefinite responses should thus be filtered out to ensure that the analysis of TLFtoBF in the CCE is free of noise caused by indefinite responses.

In summary, we used eye movement recordings to investigate the role of attentional guidance and response selection in the CCE. We first analyzed the oculomotor correlates of context-guided learning in visual search and then eliminated the potential influence of indefinite responses. Three oculomotor indexes were used to define the visual search stages: (a) initial latency before the first eye saccade was made

corresponded to the initial perceptual process, (b) eye saccades corresponded to attentional guidance, and (c) TLFToBP corresponded to response selection. Previous studies debate the processing stages underlying the CCE. This study provides insight into the stages of attentional processing where the CCE happens by comparing the three search substages defined by the oculomotor index, indicating how a visual context is learned by examining the fixation iMAP. Through the analysis of eye movement in prelearning and postlearning, we compared the fixation maps between predictive and random configurations to shed light on how contextual learning facilitates visual search processes. If attentional guidance is the source of the CCE, we should observe fewer fixations and saccades in predictive compared to random configurations (Tseng & Li, 2004; Zhao et al, 2012). Specifically, after a predictive context was learned, we expected the increased centralization of eye fixations around the target location in predictive compared to random configurations. This is because we expected participants to learn the association between predictive configurations and target locations. In addition, if response selection is also one of the contributors to CCE, TLFToBP should display a larger magnitude of downward trend for predictive configurations compared to random configurations.

METHODS

Participants

Twenty-seven undergraduate students (13 males), between the ages of 18 and 22, participated in the experiment. They were compensated with monetary rewards. All had normal or corrected-to-normal vision acuity. All participants were naïve to the purpose of the study. The research was approved by the Research Ethics Board. All participants were given informed consent prior to the experiment.

Apparatus and Stimuli

The stimuli were presented on a 19 in. CRT monitor (85 Hz sample rate) with a resolution of 1024×768 pixels and equipped with an EyeLink tracker (EyeLink 1000, SR Research, Toronto, 1000 Hz temporal resolution) that recorded the participants' eye movement while performing the task. The participants sat 60 cm away from the monitor. They were required to search for and report the orientation of a target, which was a T-shaped stimulus rotated 90° to the left or right. The distractor stimuli were L-shaped letters, rotated randomly in one of four orientations (0° , 90° , 180° , and 270° ; see Figure 1), ensuring heterogeneity among distractor stimuli. All stimuli consisted of two lines of equal length. For the search display, three black concentric circles with diameters of 9.5° , 15.5° , and 25° visual angles, respectively, were presented at the center of the monitor with a gray background. Sixteen black radius lines were presented approximately equidistant from each other and divided these concentric circles into radial lattices (Kunar et al., 2007). Thus, there were 48 conjunctions between the concentric circles and the radiated lines. On each trial, either 9 or 12 (depending on the set size) stimuli were enveloped in circular placeholders that appeared at the conjunctions. Thus, either three or four stimuli were distributed on each of the concentric circles. To determine whether the CCE occurred during the initial perception stage (for the duration of the initial latency), the size of each stimulus was varied according to its eccentricity from the fixation point to match the retina size of the stimuli (Zhao et al., 2012). The diameters of the placeholders within the inner, middle, and outer circle were subtended by a visual angle of 2° , 3.3° , and 5.4° , respectively. The stimulus items presented in the small, medium, and large placeholders were subtended by visual angles of $1^\circ \times 1^\circ$, $1.5^\circ \times 1.5^\circ$, and $2.5^\circ \times 2.5^\circ$, respectively. To rule out location probability learning, all targets appeared with equal frequency at each of the 16 possible locations with regard to quadrants and concentric circles. Half of the target locations were used for predictive configurations, and the other half - for random configurations.

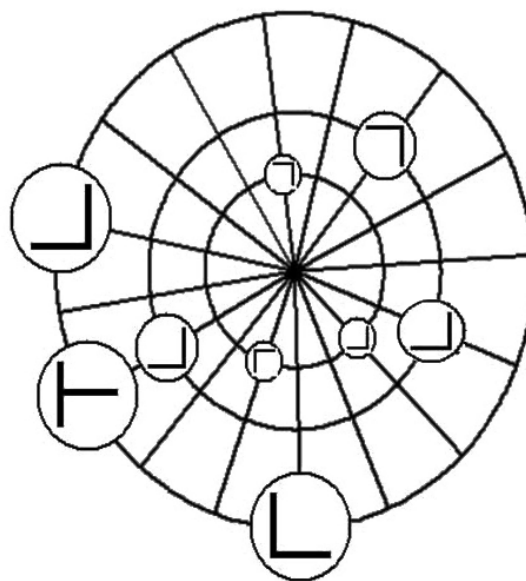


FIGURE 1.

A sample display of the search task.

Design and Procedure

There were three within-subject factors: configuration (predictive vs. random), epoch (Epochs 1-7) and set size (9 vs.12). The predictive configurations were repeated across blocks and random configurations were generated and displayed only once throughout the experiment. The orientation of the target was randomized for each trial. The experiment consisted of 28 blocks of 16 trials. Each block included eight predictive and eight random configurations. Two different set sizes (9 and 12) were randomized within the blocks. To increase statistical power, every four successive blocks were collapsed into one epoch, leading to seven epochs in total.

Each trial began with a fixation display with a duration of 800 to 1100 ms. The experimenter checked the drift correction for each trial. Following the fixation display, the search display was presented. Participants were instructed to respond to the orientation of the target as accurately and quickly as possible by pressing one of two keys: pressing the *F* key if the target was towards left and pressing the *J* key if the target was towards right. Following the response, a grey screen was displayed for 200 ms and a text message reading "next" was presented to indicate the beginning of the next trial.

Data Analysis

Trials with RTs shorter than 200 ms or longer than 4 s were excluded from the analysis. The mean accuracy was above 99%.

Initial latency is the time that elapsed between the display onset and the initiation of the first saccade (Nakatani & Pollatsek, 2004; Rayner, 1998). Saccades are defined as deflections in eye position that are greater than 0.18° , with velocities greater than 30 %/s, and accelerations greater than of 8000 %/s² (Tatler, 2007; Zhao et al., 2012). The TL_{FtoBP} is the time between the last fixation to the button press, in which the last eye fixation is defined as the final fixation that was spatially close to target (Tseng & Li, 2004; Zhao et al., 2012). The criterion to filter out indefinite responses in the late phase was dispersion greater than 2 SDs above the mean of the distance between the last fixation and target, that is, fixations over 9° based on dispersion.

The iMap method applies a Gaussian kernel function to spatially smooth each fixation map. Z scores are then computed for each map to normalize the data (Caldara & Miellet, 2011). To reveal differences in fixation patterns between the predictive and random configurations, we subtracted the saliency map of random from the predictive configuration to obtain the difference map and Z scored the difference map prior to the statistical comparison. All of the 16 predefined targets were panned to the same location at the center of the display, at the coordinates (512, 384). The corresponding fixations for each configuration were also subjected to a translation based on the panned coordinates of each target (see Figure 6). Then, we sorted all visual displays by configuration (predictive/random) and size of target placeholder (large, middle, small) and collapsed the results from all the predictive and random configurations, respectively, according to different target sizes. We predicted that fixations would be more centralized around the target location after learning the contextual information in the predictive compared to random configurations. Therefore, we focused on whether there were significant differences adjacent to the target locations.

RESULTS

Behavioral Responses

Figure 2 shows the mean RTs for the predictive and random configurations across epochs for Set sizes 9 (left panel) and 12 (right panel), respectively. The mean RTs were submitted to a three-way repeated measures analysis of variance (ANOVA, configuration [predictive, random] \times epoch: [1, 2, 3, 4, 5, 6, 7] \times set size [9, 12]), which revealed significant main effects of configuration, $F(1, 26) = 75.88, p < .001, \eta^2 = 0.75$, epoch, $F(6, 156) = 156.71, p < .001, \eta^2 = 0.86$, and set size, $F(1, 26) = 231.91, p < .001, \eta^2 = 0.89$. The interaction of configuration \times epoch was marginally significant, $F(6, 156) = 2.14, p = .052, \eta^2 = 0.08$, indicating a greater downward trend across the epochs for predictive than for random configurations. The interaction of set size \times configuration was not significant, $F(1, 26) = 3.59, p = .069$, indicating that the contextual benefits were equivalent for both set sizes. The set size \times epoch interaction was not significant, $F(6, 156) = 1.74, p = .116$. Finally, the three-way interaction of configuration \times epoch \times set size was also not significant, $F(6, 156) = 0.67, p = .676$. The results indicated that the CCE showed no difference for both set size conditions.

We calculated the search slopes and intercepts. For the slope data, a repeated-measures ANOVA (configuration [predictive, random] \times epoch [1, 2, 3, 4, 5, 6, 7]) revealed no differences between predictive and random configurations. The main effects and the interaction effect were not significant (all $ps > .069$). After collapsing the last three epochs (5 to 7), the difference between predictive and random slopes was not significant, $t(26) = -1.70, p = .101$, with about 15 ms / item greater efficiency for random configurations than for predictive ones. However, for the intercept data, a repeated-measures ANOVA (configuration [predictive, random] \times epoch [1, 2, 3, 4, 5, 6, 7]) revealed a significant main effect of configuration, $F(1, 26) = 12.95, p < .001, \eta^2 = 0.33$. No other effects showed significance (all $ps > .16$). The difference between predictive and random intercepts over last three epochs was significant, $t(26) = 3.32, p = .003, d = 0.64$. Predictive configurations produced a benefit of approximately 38 ms over random configurations.

Eye Movement Results 1: Initial Latency

Figure 3 shows the mean durations of the initial latency for predictive and random configurations across epochs for Set size 9 (left panel) and 12 (right panel). A three-way repeated-measures ANOVA (configuration [predictive, random] \times epoch [1, 2, 3, 4, 5, 6, 7] \times set size [9, 12]) showed that the main effects of configuration, $F(1, 26) = 2.04, p = .165, \eta^2 = 0.07$, set size, $F(1, 26) = 0.32, p = .58, \eta^2 = 0.01$, and epoch, $F(6, 156) = 0.54, p = .78, \eta^2 = 0.02$, were not significant. None of the interactions reached significance (all $ps > .14$). The lack of significance of the main effects during the initial latency phase for predictive and random configurations across suggests that the CCE is probably not driven by initial perceptual processing.

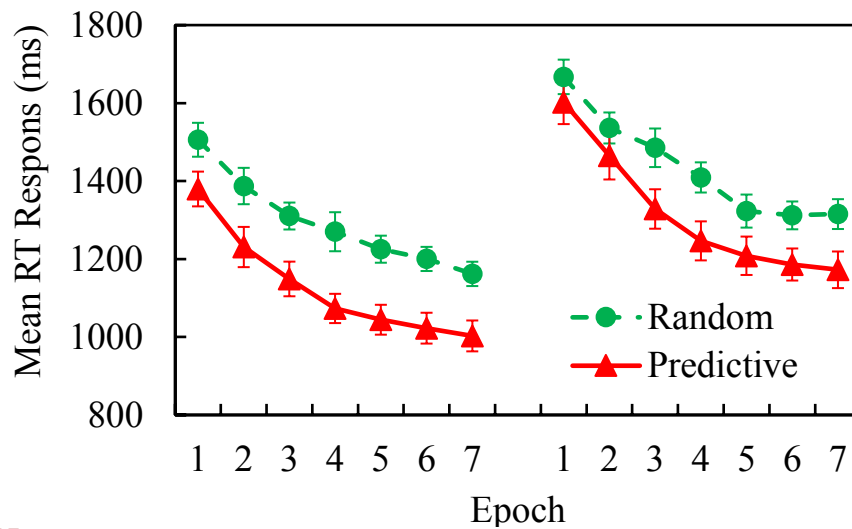


FIGURE 2.

Mean RTs as a function of epoch in predictive and random configurations for Set size 9 (left) and 12 (right).

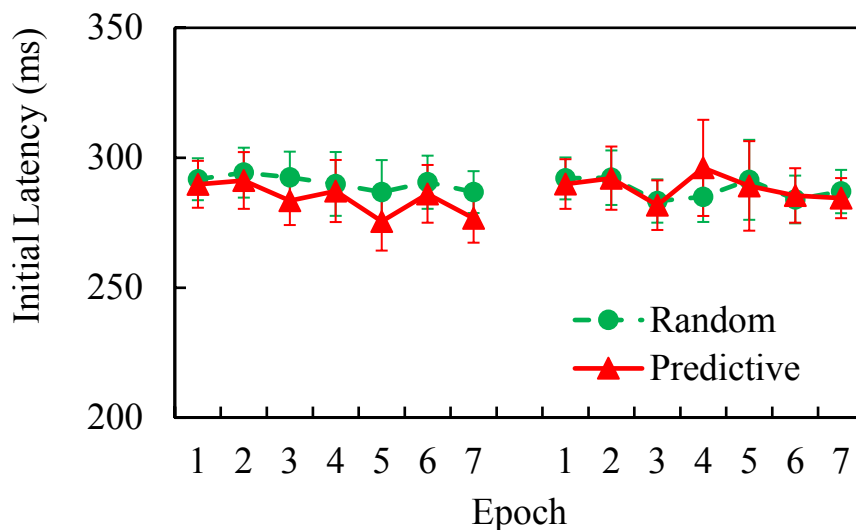


FIGURE 3.

Mean durations of initial latency as a function of epoch in predictive and random configurations for Set size 9 (left) and 12 (right).

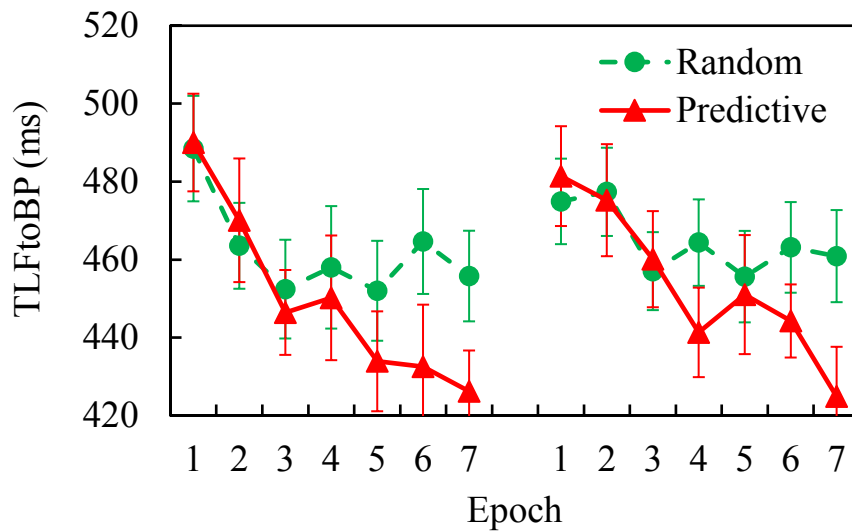
Eye Movement Results 2: Saccade Number

Studies have shown that eye movement patterns related to attentional guidance are primarily evident in the number of saccades (or fixations, Peterson & Kramer, 2001). Therefore, we analyzed the saccade number as a reflection of attentional guidance. Figure 4 shows the mean saccade numbers for predictive and random configurations across epochs for Set size 9 (left panel) and 12 (right panel). A three-way repeated-measures ANOVA (configuration [predictive, random] epoch [1, 2, 3, 4, 5, 6, 7] set size [9, 12]) revealed significant main effects of configuration, $F(1, 26) = 66.55, p < .001, \eta^2 = 0.72$, indicating a more efficient search for predictive than for random contexts. The main effect of epoch, $F(6, 156) = 92.14, p < .001, \eta^2 = 0.78$, and set size, $F(1, 156) =$

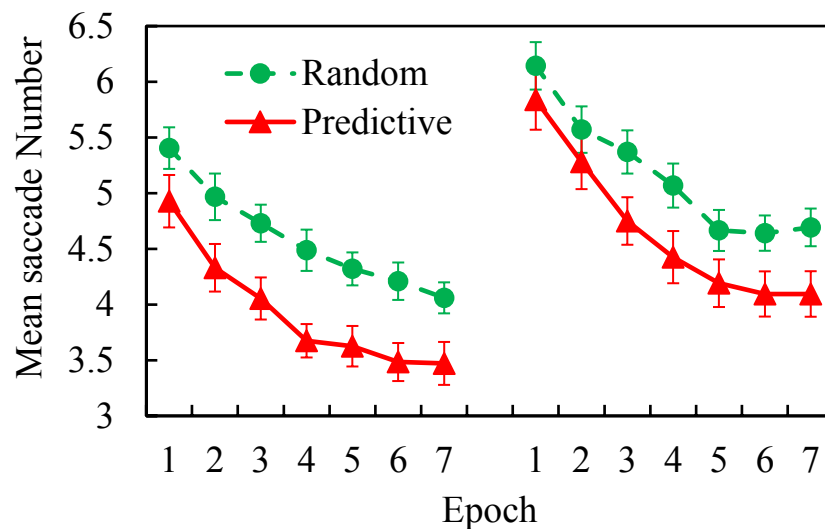
$165.97, p < .001, \eta^2 = 0.87$, also reached significance. Additionally, the interaction of set size \times epoch was significant, $F(6, 156) = 2.73, p = .015, \eta^2 = 0.10$, indicating a more efficient search for the larger set size as a function of epoch. The other interactions of set size \times configuration, configuration \times epoch, and configuration \times epoch \times set size were not significant (all $ps > .11$).

Eye Movement Results 3: The Last Fixation to Button Press

We examined ocular response in the late phase. The TLFToBP indexed response selection relevant factors. Figure 5 shows the mean durations of TLFToBP for the predictive and random configurations across epochs for Set sizes 9 (left panel) and 12 (right panel). A three-way

**FIGURE 4.**

Mean saccade number as a function of epoch in predictive and random configurations for Set size 9 (left) and 12 (right).

**FIGURE 5.**

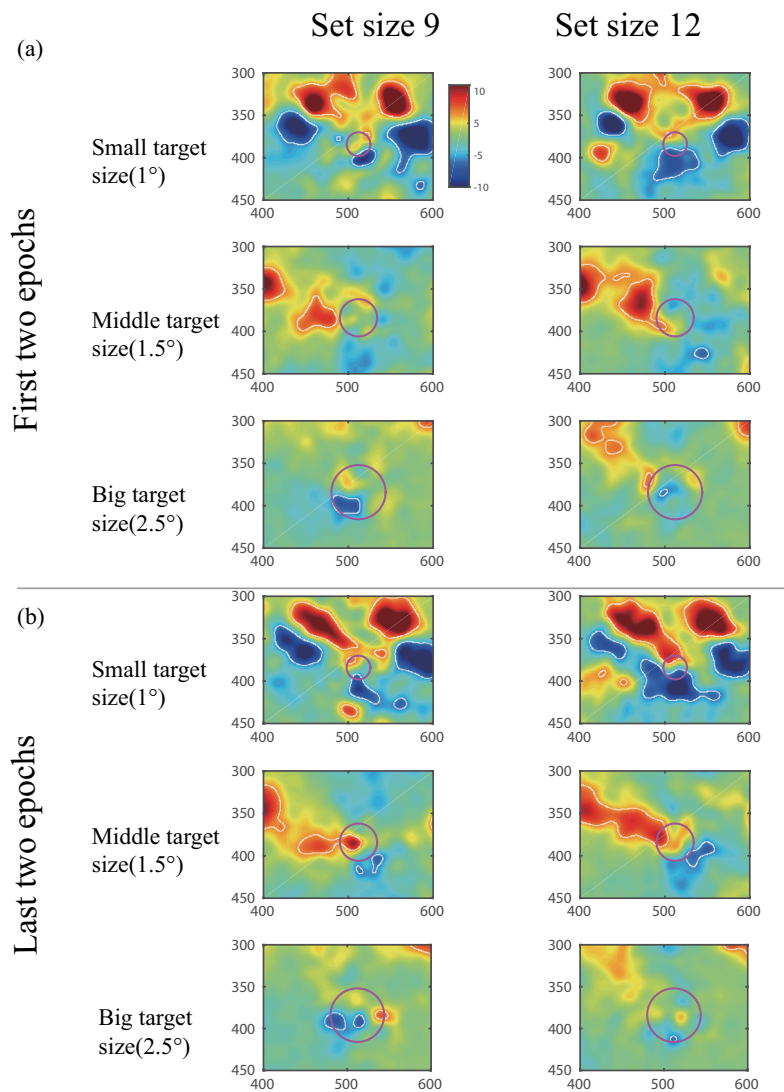
Mean durations of the last fixation to button press as a function of epoch in predictive and random configurations for Set size 9 (left) and 12 (right).

repeated-measures ANOVA (configuration [predictive, random] \times epoch [1, 2, 3, 4, 5, 6, 7] \times set size [9, 12]) revealed significant main effects of configuration, $F(1, 26) = 15.19, p < .001, \eta^2 = 0.37$, and epoch, $F(6, 156) = 7.55, p < .001, \eta^2 = 0.225$. The main effect of set size was not significant, $F(1, 26) = 1.42, p = .245, \eta^2 = 0.052$. Importantly, the interaction was significant for configuration \times epoch, $F(6, 156) = 5.41, p < .001, \eta^2 = 0.172$, indicating that there was a greater downward trend for the predictive compared to the random configurations across the epochs. The two way interactions of set size \times configuration, $F(1, 26) = 0.12, p = .728, \eta^2 = 0.01$, set size \times epoch, $F(6, 156) = 1.27, p = .273, \eta^2 = .047$, and the three way interaction of configuration \times epoch \times set

size, $F(6, 156) = 0.99, p > .433, \eta^2 = 0.04$, were not significant. It is clear from these results that response selection relevant factors were one of the sources of the CCE.

Evidence from iMap

By using the iMap toolbox, we compared the distribution of the fixations before and after the learning of contextual information. The iMap toolbox illustrated the role of attentional guidance in contextual learning. Figure 6 shows the areas of target positions delimited by purple borders according to different target sizes. Areas that show significant fixation difference are delimited by white borders ($p <$

**FIGURE 6.**

The difference salience fixation maps (Predictive - Random) with iMap for set size 9 (left panel) and set size 12 (right panel) in the first two epochs (top three panels) and the last two epochs (bottom three panels). Areas that show significant fixations are delimited by white borders ($p < 0.05$, corrected). Areas that show the target position are delimited by pink borders. Red signals represent fixations in a repeated configuration that were frequent within this region of space, relative to random configuration. Blue signals represent fixations in a repeated configuration that were less frequent within this region of space, relative to random configuration. The X and Y axis is centred and symmetrical around the central pixel values (512, 384).

0.05, corrected). Figure 6 shows the difference salience map, which was generated by subtracting the salience map of random configurations from the salience map of predictive configurations, for Set sizes of 9 (left side) and 12 (right side) for the first two (Figure 6, Panel A, prelearning) and the last two epochs (Figure 6, Panel B, postlearning) respectively. The CCE within the starting block was referred to as CCE before learning and the CCE within the final block (i.e., after training) was referred to as CCE after learning (Chun & Jiang, 1998). Thus, the difference between the first and last two epochs, as seen on the difference salience map, indicated the effect of attentional guidance from contextual learning.

The iMAP toolbox shows that in the first two epochs (prelearning), there was no significant difference in fixation distributions between predictive and random configurations in most of the target size conditions for both set sizes (see Figure 6). Significance was found only for the 2.5° target size in Set size 9. This pattern of results indicated that fixations were distributed evenly around target locations for both predictive and random configurations before learning the contextual information. However, in the last two epochs (postlearning), the fixation distributions were more centralized around the targets for the predictive compared to the random configurations in most target size conditions across set sizes. This suggests contextual information cued

visual attention to targets. Fixations were more centralized around the target locations for predictive displays than the random displays after contextual learning. However, for the 1.0° target size, significance was found only in the edge of the target location in Set size 9. The iMap results showed that before learning, there was no significant difference between fixation distributions, but after learning, the fixation distributions were more centralized around targets for predictive configurations.

Moreover, using a similar analysis to the one proposed by Beesley et al. (2018), we compared the mean distance (degrees of visual angle) between fixations to target locations. This technique allowed us to precisely measure the proximity of fixations to the target across all trials. Figure 7 shows that the mean distance from fixation to target was closer around the target for predictive versus random configurations across learning blocks. A three-way repeated-measures ANOVA (configuration [predictive, random] × epoch [1, 2, 3, 4, 5, 6, 7] × set size [9, 12]) revealed significant main effects of configuration, $F(1, 26) = 50.56$, $p < .001$, $\eta^2 = 0.66$, epoch, $F(6, 156) = 14.27$, $p < .001$, $\eta^2 = 0.35$, and set size, $F(1, 26) = 8.40$, $p = .008$, $\eta^2 = 0.24$. Importantly, the interaction was significant for configuration × epoch, $F(6, 156) = 3.37$, $p = .004$, $\eta^2 = 0.12$, indicating that the fixations were closer to target locations for the predictive compared to the random configurations across epochs. The two-way interactions of set size × configuration, $F(1, 26) = 0.22$, $p = .641$, $\eta^2 = 0.008$, set size × epoch, $F(6, 156) = 1.32$, $p = .254$, $\eta^2 = .048$, and the three-way interaction of configuration × epoch × set size, $F(6, 156) = 1.63$, $p = .142$, $\eta^2 = 0.059$, were not significant. It is clear from these results that fixation distributions were more centralized around targets for predictive configurations, which confirms the iMAP results.

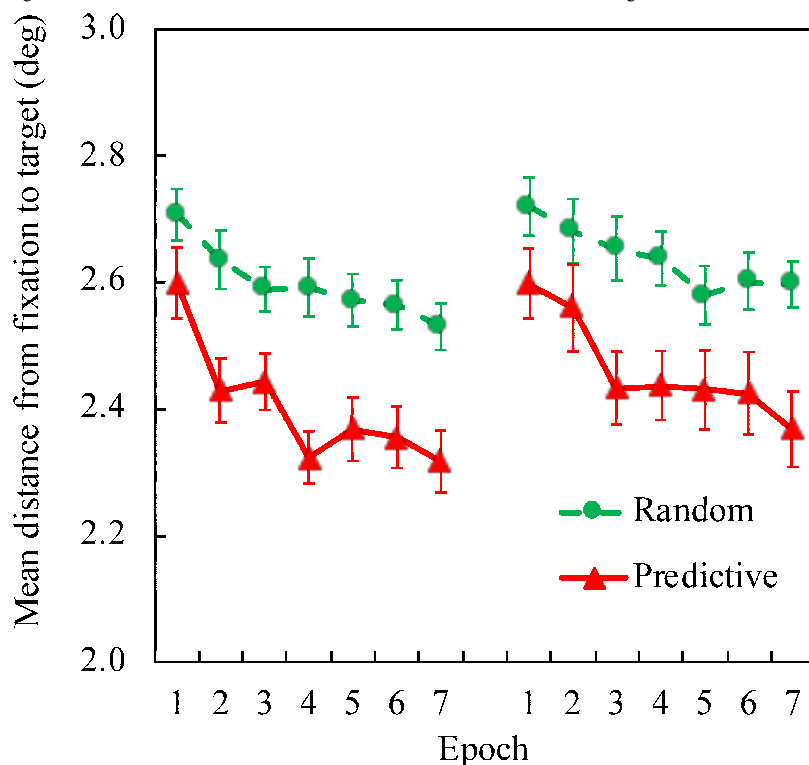


FIGURE 7.

Mean distance from fixation to target as a function of epoch in predictive and random configurations for Set size 9 (left) and 2 (right).

DISCUSSION

In this study, we used eye movement to explore the role of attentional guidance and response selection in contextual learning. The analysis of the manual RT measure revealed a significant CCE in both set sizes, suggesting an implicit context-guided learning in a visual search task. Although contextual information can facilitate visual search, it is important to understand at which processing stage contextual information can be acquired to facilitate the search process. The CCE could be driven by three processes: (a) initial perceptual processing, (b) attention guidance, where visual context matches the perceptual representation stored in the working memory, and (c) decision or response factors. Some studies hold the view that the CCE is only driven by attentional guidance (Chun & Jiang, 1998), whereas others argue that CCE could also be driven by response selection (Kunar et al., 2007; Zhao et al., 2012). To address these inconsistencies, we explored the exact processing stage in which the CCE took place.

Since perceptual and cognitive processes influence latency before the first eye saccade is made (Rayner, 1998), we used initial latency as an index of initial perceptual processing. Consistent with previous studies, we found no difference in the initial latency between the predictive and random configurations, and the RT performance was equivalent between random and predictive configurations. Our findings suggest that the CCE is not driven by the initial perceptual stage.

Our results indicate that attentional guidance is one of the sources of the CCE. By using saccade number as an index of attentional guidance, we found that the patterns of saccade number were nearly the same as for RT measures (see Figures 2 and 5). Our data suggests that the search

facilitation of the CCE was caused by reduced saccade number for predictive configurations. This finding is consistent with previous work by Tseng and Li (2004), who divided saccades into ineffective and effective saccades. In their study, all types of saccades displayed a significant CCE. In our study, the analysis of iMap for fixation density before and after contextual learning revealed that fixations were distributed closer to the target in predictive displays than in random displays. After the context was learned, fixation density was significantly higher in predictive than in random displays. These results suggest that eye movements may guide attention toward the location of targets following implicit learning of associations between target locations and informative configurations (Chun & Jang, 1998). This implicit association causes individuals to search for targets with greater intent and efficiency. These results are in line with previous studies examining attentional guidance—these studies found that visual contexts facilitate search performance (Chun & Jang, 1998; Chun & Jang, 1999). Thus, if contextual information determines the saliency map for a viewed display, then attention and eye movements are deployed to regions of high saliency, which then facilitates the response to objects within that area of the display (Chun, 2000).

Attentional guidance, however, cannot entirely account for the CCE. The slopes of behavioral data did not reveal differences between configurations. Instead, a difference in intercepts signified greater improvement for predictive over random configurations, which suggested that nonsearch factors may have contributed to the CCE. The late response selection was indexed by the TLFToBP. The significant difference in the TLFToBP for predictive and random configurations observed for both set sizes indicated that response selection was also one of the sources of CCE. The difference between predictive and random configurations over the last three epochs were 27 ms and 20ms for Set sizes 9 and 12, respectively. This finding was in line with the study by Kunar et al. (2007), who found a small magnitude of about 30 ms CCE in a single feature search task (no search required). However, when interference was added to the response selection process, the CCE disappeared. Their results raised the question of whether response-related processes can be facilitated during context learning. As we argued, response selection involves a comparison, decision, and response to a target, all of which are nonsearch factors. A quicker response to targets in predictive contexts may be attributed to a response threshold that might be lower when a target appears in a predictive context compared to a random one (Kunar et al., 2007). In the study by Tseng and Li (2004), no effect of context was found on the TLFToBP. In the Zhao et al. (2012) study, a significant difference in configuration was found even in the first epoch. However, Zhao et al. (2012) did not exclude indefinite responses when estimating the contribution of the TLFToBP to the CCE. Thus, we filtered out indefinite fixations in the present study, and the current results show that response selection could be one of the sources of contextual learning. Another explanation for the role of response selection is the familiarity of predictive contexts, which creates a higher certainty of response-related processes that additively contributed to the contextual cueing effect. Under this assumption, Schankin and Schubö (2009, 2010) used ERP recordings to distinguish between attentional guidance and response-related processes using electrophysiological components of N2pc (an indicator of atten-

tion shift) and LRP (an indicator of response-related motor processes). Their results found that the behavioral CCE significantly correlated with both the N2pc and the LRP component. They concluded that both a more efficient attentional selection and faster response-related processes (probably due to certainty in response selection) contribute to the CCE.

Our fixation maps also showed that after contextual information was learned, the fixation density for predictive displays was thicker and centralized in smaller target size conditions. When the target size was small, as measured by a small visual angle, the saliency areas for targets were dense and almost inside the placeholders (ROI; see Figure 6, Panel B). When the target size was larger, the saliency areas for the targets were relatively sparse. For the 2.5° target size in Set size 12, the number of fixations made around target locations in the predictive display was not significantly greater than that made in the random display. This could have occurred for two reasons. First, for a larger target, the detection threshold may have been relatively lower because subjects did not need to devote many attentional resources to detect the target; only fixating near the target could have resulted in its successful detection. Alternatively, the larger targets were distributed at the peripheral circle, and the precision of eye movement recordings might have declined with increased distance from the display center. This could have deviated the observed measure of fixation position from the actual value, resulting in a sparse distribution on the fixation map.

Mapping eye fixations provided us with a visualization of how contextual learning affects fixation distribution. However, we admit that the magnitude difference we observed between prelearning and postlearning was not as great as what we expected for the current experiment. Nevertheless, the analysis of mean distance from fixation to target as a function of epochs confirmed the iMAP results; fixations tended to be closer to the target for predictive configurations compared to random configurations. Adopting an analytic technique from Beesley et al. (2018), we also grouped the fixation data by the number of fixations within a trial. Analyzing data in this way presented a dynamic analysis of fixations across the course of the trials. In line with Beesley et al. (2018), Figure 8, Panel A shows that the pattern of data is very similar between predictive and random configurations across the trials of different length (i.e., different numbers of fixations taken to finish the search task). For trials with more than six fixations, an inefficient search process was followed by an efficient search process (see also Tseng & Li, 2004). Figure 8, Panel B shows the percentage of trials with different numbers of fixations taken to find the target for predictive and random configurations, respectively. Combining Panels A and B in Figure 8, it can be inferred that the benefit of contextual learning of predictive configurations is driven by a greater number of trials having fewer fixations, particularly in the range of 3-5 fixations (see also Beesley et al., 2018).

Overall, by using iMAP, we provided a visualization of how implicit contextual learning affects the fixation distribution, and we confirmed the results by quantifying the mean distance from fixation to target across the learning progress. Thus, we provided qualitative and quantitative evidence that the fixations were closer to target locations after contextual learning.

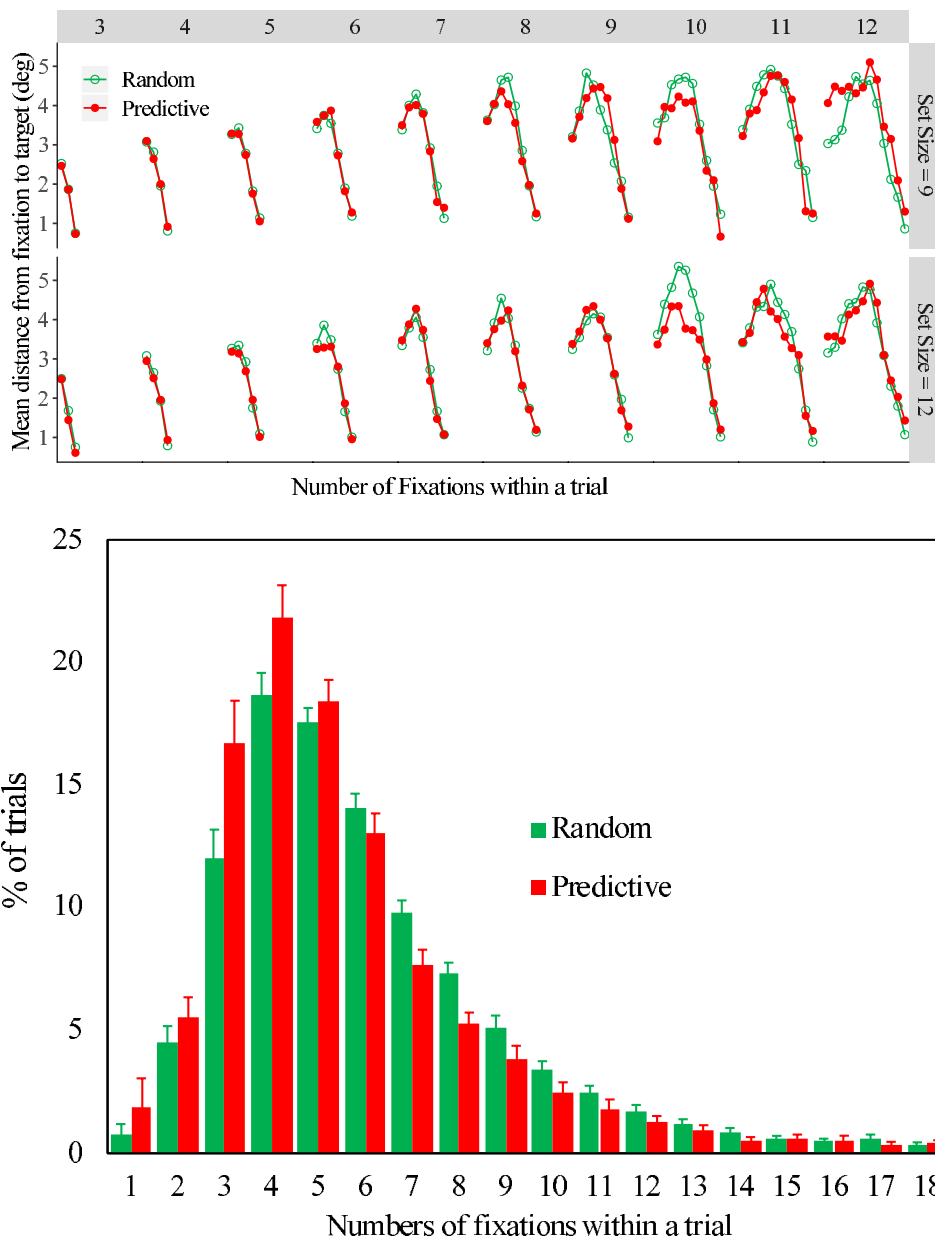


FIGURE 8.

Fixation data grouped according to the number of fixations per trial and separately for predictive and random configurations. Top: mean distance (degree of visual angle) from the target of each fixation in turn from left to right. Bottom: percentage of trials with different numbers of fixations.

CONCLUSIONS

The present study recorded eye movements to explore the mechanism of the CCE by partitioning visual search into three substages. With regard to attentional guidance, the acquisition of associations between target locations and predictive contexts may guide attention to targets after the visual context was learned, speeding up RTs. Even by eliminating the indefinite responses in the response selection process, we found response priming with the facilitation of contextual information. Therefore, both attentional guidance and response selection are sources of the CCE.

REFERENCES

- Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 348–362. doi: 10.1037/xlm0000033
- Beesley, T., Hanafi, G., Vadillo, M. A., Shanks, David, R., & Livesey, E. J. (2018). Overt attention in contextual cuing of visual search is driven by the attentional set, but not by the predictiveness of distractors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 707–721. doi: 10.1037/xlm0000467

- Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, 43, 864–878. doi:10.3758/s13428-011-0092-x
- Chun, M., & Jiang. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28–71. doi: 10.1006/cogp.1998.0681
- Chun, M., & Jiang, Y. H. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360–365. doi: 10.1111/1467-9280.00168
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Science*, 4, 170–178. doi: 10.1016/S1364-6613(00)01476-5
- Jiang, Y. H., & Chun, M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology Section A Human Experimental Psychology*, 54, 1105–1124. doi: 10.1080/713756001
- Jiang, Y. V., & Sisk, C. A. (2019). Contextual Cueing. In: S. Pollman (Ed.), *NeuroMethods: Spatial learning and attentional guidance* (pp. 1–14). Totowa, NJ: Humana Press.
- Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, 33, 816–828. doi: 10.1037/0096-1523.33.4.816
- Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2008). Time to guide: Evidence for delayed attentional guidance in contextual cueing. *Visual Cognition*, 16, 804–825. doi:10.1080/13506280701751224
- Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). iMap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods*, 49, 559–575. doi:10.3758/s13428-016-0737-x
- Nakatani, C., & Pollatsek, A. (2004). An eye movement analysis of “mental rotation” of simple scenes. *Perception & Psychophysics*, 66, 1227–1245. doi: 10.3758/BF03196848
- Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology-Learning Memory and Cognition*, 27, 1299–1313. doi: 10.1037/0278-7393.27.5.1299
- Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics*, 63, 1239–1249. doi: 10.3758/BF03194537
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Schankin, A., & Schubö, A. (2009). Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. *Psychophysiology*, 46, 668–679. doi: 10.1111/j.1469-8986.2009.00807.x
- Schankin, A., & Schubö, A. (2010). Contextual cueing effects despite spatially cued target locations. *Psychophysiology*, 47, 717–727. doi:10.1111/j.1469-8986.2010.00979.x
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7, 1–17. doi: 10.1167/7.14.4
- Tseng, Y.-C., & Li, C.-S. R. (2004). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, 66, 1363–1378. doi: 10.3758/BF03195004
- Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9, 33–39. doi: 10.1111/1467-9280.00006
- Wolfe, J. M., Vo, M. L., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Science*, 15, 77–84. doi:10.1016/j.tics.2010.12.001
- Zang, X., Jia, L., Müller, H. J., & Shi, Z. (2015). Invariant spatial context is learned but not retrieved in gaze-contingent tunnel-view search. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 41, 807–819. doi:10.1037/xlm0000060
- Zhao, G., Liu, Q., Jiao, J., Zhou, P., Li, H., & Sun, H. J. (2012). Dual-state modulation of the contextual cueing effect: Evidence from eye movement recordings. *Journal of Vision*, 12, 11. doi:10.1167/12.6.11

RECEIVED 01.04.2019 | ACCEPTED 09.10.2019