Template-based attentional guidance and generic procedural learning in contextual guided visual search: Evidence from reduced response time variability

Hongyu Yang*

Center for Cognition and Brain Disorders,
Affiliated Hospital of Hangzhou Normal University,
Hangzhou Normal University, Hangzhou, China
Zhejiang Key Laboratory for Research in Assessment of
Cognitive Impairments, Hangzhou, China
Zhejiang Philosophy and Social Science Laboratory for
Research in Early Development and Child Care,
Hangzhou Normal University, Hangzhou, China

Shasha Zhu*

Center for Cognition and Brain Disorders,
Affiliated Hospital of Hangzhou Normal University,
Hangzhou Normal University, Hangzhou, China
Zhejiang Key Laboratory for Research in Assessment of
Cognitive Impairments, Hangzhou, China
Zhejiang Philosophy and Social Science Laboratory for
Research in Early Development and Child Care,
Hangzhou Normal University, Hangzhou, China

Senlin Liu

Center for Cognition and Brain Disorders,
Affiliated Hospital of Hangzhou Normal University,
Hangzhou Normal University, Hangzhou, China
Zhejiang Key Laboratory for Research in Assessment of
Cognitive Impairments, Hangzhou, China
Zhejiang Philosophy and Social Science Laboratory for
Research in Early Development and Child Care,
Hangzhou Normal University, Hangzhou, China

Lixia Yuan

Center for Cognition and Brain Disorders,
Affiliated Hospital of Hangzhou Normal University,
Hangzhou Normal University, Hangzhou, China
Zhejiang Key Laboratory for Research in Assessment of
Cognitive Impairments, Hangzhou, China
Zhejiang Philosophy and Social Science Laboratory for
Research in Early Development and Child Care,
Hangzhou Normal University, Hangzhou, China

Xiaowei Xie

Department of Psychology, Southwest University,
Chongqing, China

Citation: Yang, H., Zhu, S., Liu, S., Yuan, L., Xie, X., & Zang, X. (2025). Template-based attentional guidance and generic procedural learning in contextual guided visual search: Evidence from reduced response time variability. *Journal of Vision*, *25*(4):1, 1–19, https://doi.org/10.1167/jov.25.4.1.

https://doi.org/10.1167/jov.25.4.1 Received September 20, 2024; published April 1, 2025 ISSN 1534-7362

Xuelian Zang

Center for Cognition and Brain Disorders, Affiliated Hospital of Hangzhou Normal University, Hangzhou Normal University, Hangzhou, China
Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China
Zhejiang Philosophy and Social Science Laboratory for Research in Early Development and Child Care, Hangzhou Normal University, Hangzhou, China

The contextual cueing effect—where participants search repeated displays faster than novel ones—is often explained by the "attention guidance" account, which posits that repeated exposure helps individuals learn the context and attend to the likely target locations. Alternatively, the "generic procedural learning" account suggests that a general search strategy is developed for all displays, although repeated contexts play a higher weight in optimizing the strategy due to their higher presented frequency. This makes responses faster for repeated displays than novel displays. The current study examined these two mechanisms using a varied contextual cueing paradigm to analyze response time (RT) variability with the coefficient of variation (CV) and time–frequency analysis of RTs. Experiment 1 involved uninterrupted training with repeated and novel displays presented separately, followed by a test with randomly interleaved repeated and novel displays. Experiment 2 used interleaved displays for training before an uninterrupted test phase. Both experiments revealed faster RTs and reduced template-based variability for repeated displays early in the training, supporting attentional guidance. However, generic procedural learning, indicated by a late onset of lower cross-display variability for repeated displays, required more time and training to validate the cueing effect. These findings suggest that attentional guidance dominates early learning, but both mechanisms contribute to the contextual cueing effect overall.

Introduction

The ability to accurately identify and locate targets in complex environments is crucial for human survival and adaptation. This skill has practical applications in various domains, such as animal foraging and disaster rescue response. Humans have developed advanced cognitive processes to enhance search efficiency in these scenarios. Notably, the *procedural learning effect* (Fitts, 1964), where practice of the same task leads to faster performance, and the *contextual cueing effect* (Chun & Jiang, 1998), where search is faster in familiar contexts compared with unfamiliar ones, are both well-researched phenomena with significant implications for optimizing search performance.

The contextual cueing effect was first investigated by Chun and Jiang (1998), who had participants search for a target "T" among distractor "Ls" in both repeated contexts (with stable T–L associations) and novel contexts (where distractor locations varied unpredictably, disrupting T–L associations). They found that participants responded faster in the repeated context, demonstrating that consistent T–L associations facilitated target searching. The dominant explanation for this effect is the attentional guidance account (Sisk, Remington, & Jiang, 2019; Wolfe, 2021; Wolfe, Cave, & Franzel, 1989), such that participants possess the capability to learn spatial associations that are repeatedly presented, forming memory templates of the specific target–distractor associations, thereby allowing this acquired template-based knowledge to direct human attention toward probable target locations within a repeated context. Consequently, this process accelerates the processing of visual search tasks.

Evidence supporting the template-based attentional guidance account comes from electroencephalogram (EEG) studies that have shown increased N2pc (Johnson, Woodman, Braun, & Luck, 2007; Schankin & Schubö, 2009) and N1pc (Zinchenko, Conci, Töllner, Müller, & Geyer, 2020) components in repeated displays compared with novel ones. The N2pc indicates spatial attention allocation toward the target location, whereas the N1pc suggests an automatic spatial bias from contextual memory templates. Notably, both N1pc and N2pc components remain associated with original target locations, even when the target is shifted to the opposite side in subsequent test phases (Zinchenko, Conci, Hauser, Müller, & Geyer, 2020). Furthermore, in instances where participants encountered a repeated context paired with two possible target locations (Zinchenko et al., 2024), the N2pc component was more strongly related to the dominant target than the minor target, exhibiting greater facilitation from contextual cueing and a larger N2pc amplitude. This underscores the robustness of attentional guidance.

In addition to EEG studies, eye-tracking investigations have also provided supportive evidence for attentional guidance. For example, the eye-movement literature consistently reports that fewer fixations and saccades are required to locate the target within a repeated context compared with a novel context (Harris & Remington, 2017; Kroell, Schlagbauer, Zinchenko, Müller, & Geyer, 2019; Manginelli & Pollmann, 2009; Peterson & Kramer, 2001; Tseng & Li, 2004). Zhao et al. (2012) and Zang, Jia, Müller, and Shi (2015) reported optimal scan-path patterns in repeated displays compared with novel displays. These findings suggest that contextual cueing facilitates visual search in terms of fixations and saccades. Moreover, there is a consistent association between the contextual cueing response time (RT) effect and the fixation/saccade effect, further reinforcing the role of contextual cues in guiding attention.

Despite the robust finding of fewer fixations in repeated contexts, whether a specific template-based eye movement pattern exists for repeated displays remains unclear. On the one hand, Peterson and Krammer (2001) reported that the initial fixation after display onset is more likely to land on the target location in repeated displays than in novel displays, suggesting that contextual learning could promote specific eye-movement patterns. However, Tseng and Li (2004) challenged this and found that participants could not complete the search with only one fixation. They reported that only the number of saccades, not other parameters such as distance of the first fixation from the target or time from the last fixation to button press, showed significant interactions for display type (repeated vs. novel) and learning set (pre-learning vs. post-learning). They proposed that contextual cueing does not immediately and exclusively guide attention to the target. Similarly, Zhao et al. (2012) found that the distance of the first fixation to the target location was comparable between repeated and novel displays. Although not reported in our previous study (Zang et al., 2015), we did analyze different properties of the first and last fixations to the target and found no significant differences between repeated and novel displays. These findings raise an interesting question: Although learning in repeated contexts reduces the number of saccades and fixations required, previous studies have rarely observed any constant eye movement pattern for a specific repeated or a learned context template. This calls into question the template-based attentional guidance hypothesis in contextual learning.

Why was no consistent template-based eye movement pattern for repeated displays observed in previous studies? Seitz, Zinchenko, Müller, and Geyer (2023) tackled this question by suggesting that generic procedural learning plays a role in contextual cueing. Specifically, participants may develop a generic search strategy applicable to various displays, regardless of whether they are repeated or novel. As a result, this approach leads to relatively consistent eye movement patterns when searching through both repeated and novel displays. Using a contextual cueing search task, Seitz and colleagues analyzed participants' oculomotor scan paths employing several innovative methods: dynamic time warping (which measures similarity between fixational series of different lengths through temporal alignment), area between curves (which assesses scan-path similarity based on the area between curves), and discrete Fréchet distance (which accommodates time series of varying lengths). Seitz et al. hypothesized that, if participants developed a specific search strategy for each repeated display (see attentional guidance account), there would be greater oculomotor variation among different repeated displays and less variation for multiple presentations of the same displays. In contrast, if a generic search strategy were applied across different displays, oculomotor variation would decrease for all displays, but the decrease would be more pronounced for repeated displays. This is because repeated displays, occurring more frequently, have a stronger influence on optimizing the generic search strategy. This hypothesis was supported by Seitz et al. (2023), whose experiments demonstrated decreased oculomotor variation over experimental blocks, and the decreases were larger for repeated displays than novel displays (i.e., contextual cueing effect).

The generic procedural learning account is compelling because it connects the contextual cueing effect (faster search in repeated contexts) and the general procedural learning effect (faster performance with practice), both observed simultaneously in previous contextual cueing studies (Goujon, Didierjean, & Thorpe, 2015; Sisk et al., 2019). It also explains several previously difficult-to-understand phenomena, as discussed by Seitz et al. (2023). For example, participants may adapt their oculomotor scanning behavior based on the overall display statistics, regardless of whether or not the display is repeated. This approach saves cognitive effort: Instead of memorizing specific distractor–target configurations and checking each display against stored representations, participants develop a strategy optimized for the general statistical search environment. Consequently, participants might exhibit inefficient (indirect) oculomotor scanning behavior during the early stages of search, resulting in similar oculomotor patterns for both repeated and novel displays. This could explain why previous eye movement studies rarely observed distinct template-based eye movement patterns for a specific repeated display (Harris & Remington, 2017; Kroell et al., 2019; Manginelli & Pollmann, 2009; Peterson & Kramer, 2001; Tseng & Li, 2004).

Seitz et al. (2023) also mentioned that the generic procedural learning could also contribute to the implicit nature of contextual learning, where participants often cannot easily distinguish between repeated and novel displays. This difficulty may arise because a general search strategy is developed for both types of displays, making it more difficult to differentiate between them. Despite the compelling nature of the generic procedural learning account, it was only recently proposed by Seitz et al. (2023) based on their eye movement study, and there is not enough evidence to fully support this claim. More studies, especially those using traditional behavior paradigms, are needed to validate this account and to compare it to existing literature.

To obtain evidence for generic procedural learning using the traditional contextual cueing behavior paradigm, one possible approach is to analyze participants' RT variability. If the template-based attentional guidance account explains contextual learning, then participants' RT variability for a specific repeated template or display should decrease, as they are relying on the same or a similar search strategy for the learned template. In contrast, if the generic procedural learning account is the reason for contextual learning, then participants' RT variability among different displays (both repeated and novel) should decrease with contextual learning, because repeated displays have a greater weight in optimizing the search strategy. It is important to note that these two accounts are not mutually exclusive and may both contribute to contextual cueing. Although template-based guidance facilitates early attentional allocation, generic procedural learning may play a role in refining search strategies over extended training. In short, by analyzing the template-based and cross-display RT variability, we could gather further evidence to examine the roles of template-based learning and generic procedural learning in contextual cueing.

Building on the previous arguments, this study conducted two experiments to systematically investigate human contextual learning behavior and evaluate the role of generic procedural learning in the contextual cueing effect, with a particular focus on RT variability. To achieve this, we examined traditional mean RTs and error rates to confirm the classic contextual cueing effects, as well as the coefficients of variation (CVs)—the ratio of the standard deviation (*SD*) of RT to the mean RT—both for specific configurations across different blocks (template-based CVs) and for different displays within a block (cross-display CVs). We chose the CV over *SD* to measure RT variability, as *SD* often covaries with RT. According to the instance theory of Logan (1988), each exposure to a stimulus creates a distinct memory "instance," and repeated presentations of a display result in the accumulation of more instances. As the number of instances increases, memory retrieval becomes more efficient, leading to faster RTs and reduced RT variability (e.g., *SD*). However, Logan's theory predicts that both mean RT and SD should decrease at similar rates over time, as they both reflect the same underlying memory traces. In this context, the CV provides a more precise measure of contextual learning, as it is less influenced by the overall progression of the experiment because the decreases in RT and SD cancel out each other.

With regard to the cross-display CV, separating repeated and novel displays into distinct blocks can provide clearer results, as changes in display type might affect variation. Therefore, in addition to the traditional contextual cueing paradigm with repeated and novel displays mixed within each experimental block, we also used a variant of the contextual cueing paradigm, instructing participants to search for a "T"-shaped target among "L"-shaped distractors in sets with either purely repeated or purely novel displays. These uninterrupted sets were presented either at the beginning of the experiment (without prior contextual learning) in Experiment 1 or after a prior standard contextual learning phase with intermixed repeated and novel trials in Experiment 2. This design allowed us to examine contextual cueing behavior both during the early learning stage and after it. An additional advantage of separate repeated and novel displays in different sets is that each set contains only one experimental condition, enabling us to treat the experimental trials in each set as continuous events. With this assumption, we could perform frequency analysis using fast Fourier transform (FFT) on participants' responses, providing further evidence not only from the time domain (across-display CV) but also from the frequency domain (Adamo, Martino et al., 2014; Adamo, Huo et al., 2014; Di Martino et al., 2008; Johnson, et al., 2007; Liu et al., 2017; Machida, Murias, & Johnson, 2019).

In conclusion, our study aimed to evaluate the role of template-based learning and generic procedural learning in contextual cueing. Results can be linked to these two types of learning in several ways. If only template-based learning contributes to the contextual cueing effect, we would expect overall lower template-based CVs for repeated displays compared with novel displays, with no change in cross-display CVs and FFT results. If only generic procedural learning contributes, we would expect significant facilitation in cross-display CVs and FFT results for repeated displays but no effect on template-based CVs. If both accounts contribute, reductions would be observed in both template-based and cross-display CVs, as well as FFT results for repeated displays. Finally, if no difference is observed in either CV type or FFT results, this would suggest that neither account contributes to the contextual cueing effect.

Experiment 1

Methods

Participants

Forty naive participants with normal or corrected-to-normal visual acuity (36 females; mean age, 20.93 ± 1.98 years; all right-handed) were recruited from Hangzhou Normal University. The sample size was estimated to be 38 using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), given an $f(U)$ effect size of 0.25 with 85% power (Fan et al., 2024), an alpha level of 0.05, and a non-sphericity correction of 1 (groups = 2, number of measurements = 2). The study was approved by the ethics committee of the Institutes of Psychological Sciences at Hangzhou Normal University (no. 2022[E2]-KS-080). Participants provided informed consent and were given monetary compensation for participation.

Apparatus and stimuli

The experiment was conducted in a dimly lit room (ambient light, <1 cd/m^2). Visual stimuli were presented on a 27-inch LCD monitor with a resolution of 1920 × 1080 pixels and a refresh rate of 120 Hz. MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) were used for stimulus presentation and response collection. During the experiment, participants sat approximately 57 cm away from the computer screen, and their head position was fixed by a chin rest. The background color of the screen was gray with a luminance of 11.58 cd/m^2. Each search display contained one "T"-shaped target and 11 "L"-shaped distractors, each subtending 0.80° × 0.80° of visual angle and presented in white (43.34 cd/m^2). The positions of the 11 distractors were randomly selected from the available 10 × 10 invisible matrix grid locations as the stimulus presentation area (subtending a visual angle of 12.02° × 12.02°). Note, however, that the target never appeared in the central 2 × 2 units or in the four corners (each containing six locations), resulting in a total of 28 grid locations being avoided (Figure 1). This was done to prevent participants from spotting it immediately after the display onset or to avoid making the search task excessively difficult. In terms of orientations of the search items, each of the "L"-shaped distractors could be presented in four different rotations (0°, 90°, 180°, or 270°), and the "T"-shaped target had two possible orientations, rotated either 90° to the left or to the right.

Design and procedure

The experiment consisted of two main phases: a training phase and a test phase (see Figure 2B). The training phase consisted of two sets: a repeated set and a novel set. Each set contained 25 blocks, with either 12 repeated configurations or 12 novel configurations presented in each block. The sequence of these two sets was randomly determined and counterbalanced among participants (see Supplementary Materials for details). Each set lasted 12.5 minutes, with no breaks in between.

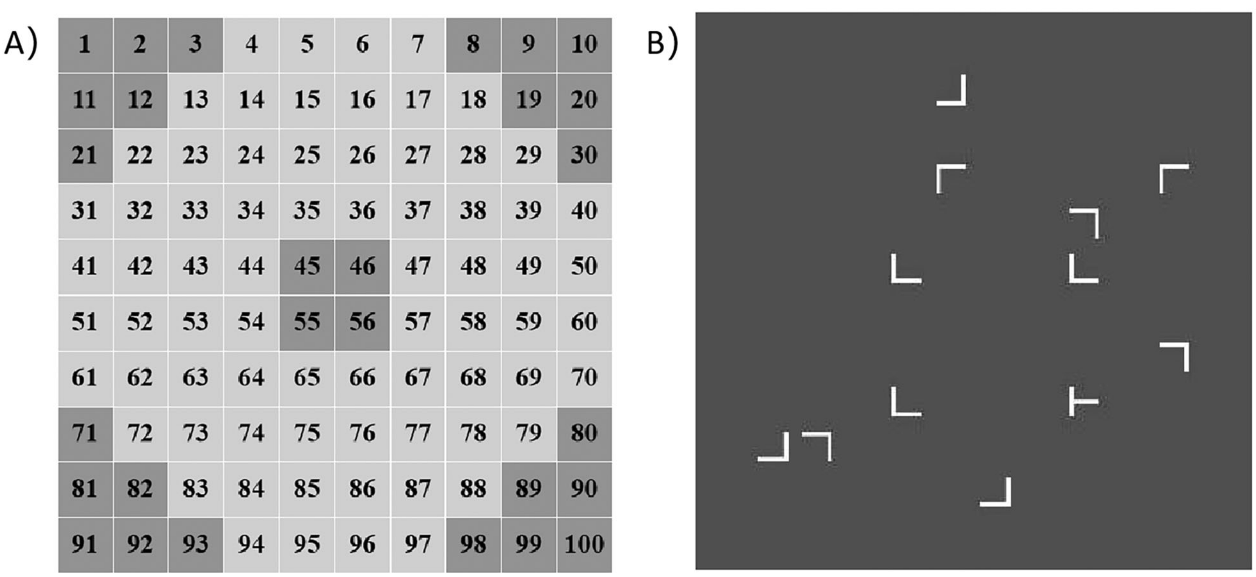

Figure 1. (**A**) Potential item locations within the grid layout. Distractors can appear in any of the 100 grid locations, but the target could not appear in the central 2 × 2 unit or the four corner areas (shaded squares). (**B**) Schematic illustration of a visual search display. Each display contains one "T"-shaped target and 11 "L"-shaped distractors. The target could be oriented 90° to the left or right, whereas the distractors could appear in one of four orientations (0°, 90°, 180°, or 270°).

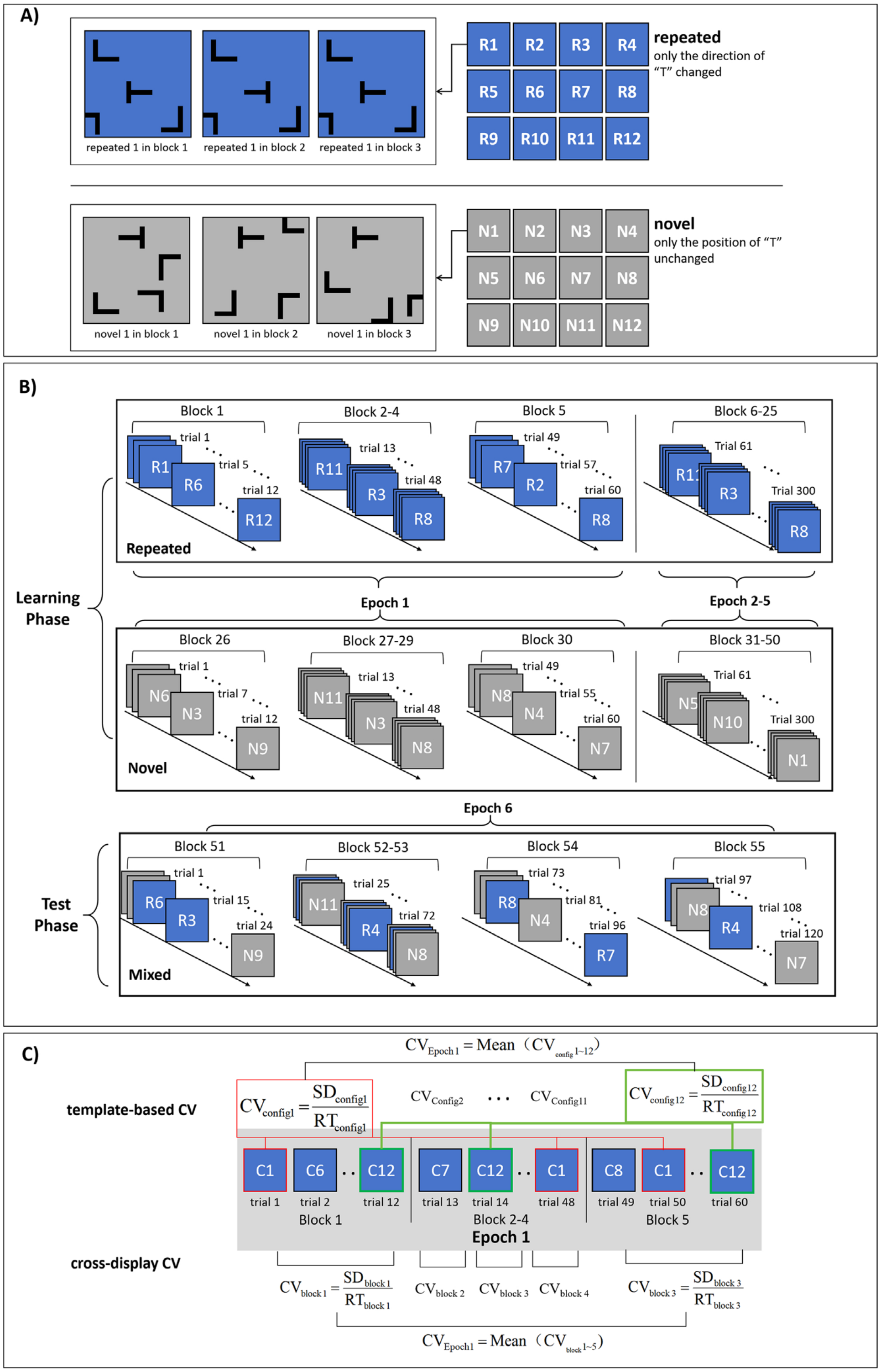

Figure 2. (**A**) Schematic illustration of the design for repeated and novel configurations. The repeated contexts consisted of 12 configurations (R1–R12), each defined by 12 unique positions of the target ("T"). Across different blocks, the positions of the "T" and distractors ("Ls"), as well as the direction of the "Ls" remained constant, whereas only the orientation of "T" varied. The novel contexts were also comprised of 12 configurations (N1–N12), each defined by 12 unique positions of "T." Across different blocks, only the position of "T" remained constant, whereas the positions and orientations of the "L" and the direction of the "T" were varied. Note that the number of items in the figure has been reduced to three for better visual clarity. In the actual experiment, each display contained 12 search items, with 11 distractors and one target, in both repeated and novel configurations. (**B**) Schematic illustration of the experimental paradigm in Experiment 1, which included a two-set training phase and a test phase. In the training phase, repeated

→

←
and novel contexts were presented in separate sets (repeated set and novel set). Each set contained 25 blocks, and each block consisted of 12 trials of repeated or novel configurations. In the test phase, repeated and novel contexts were intermixed within the same blocks. Each test block contained a mix of 12 repeated and 12 novel displays, using the same configurations as in the training phase. (**C**) The calculation of CVs. Template-based CVs were calculated for each specific configuration (e.g., C1, C2) across five blocks within an epoch. For repeated configurations, data from the same configuration across five blocks within an epoch were aggregated to compute the CVs, whereas for novel configurations those with the same target location were grouped together for the calculation. Cross-display CVs were calculated across all trials within a single block, regardless of the specific configuration, across all repeated or novel configurations within that block.

This design was intended to allow for continuous hypothesis testing during the search trials, enabling frequency analysis via FFT.

Participants were allowed to take a relatively long break of 2 to 5 hours between the two training sets. This differs from traditional studies, which typically provide only short breaks (ranging from seconds to minutes) between blocks. The long break was designed to allow for sufficient recovery of cognitive resources (Helton & Russell, 2015) after completing a continuous search of an entire set. A second reason for selecting this longer break was to minimize potential recency or primacy effects, which refer to differences in contextual learning patterns when participants first learn blocks of repeated displays and then blocks of novel displays, or vice versa (Jungé, Scholl, & Chun, 2007; Vaskevich & Luria, 2019). A longer break reduces the likelihood of these effects influencing performance across the two sets. However, the break was not so long as to allow participants to engage in extended rest, such as overnight sleep, which could interfere with memory consolidation (Geyer, Mueller, Assumpcao, & Gais, 2013).

For the repeated training set, 12 unique configurations—each consisting of one “T”-shaped target and 11 “L”-shaped distractors—were randomly generated at the beginning of the experiment. These configurations were then presented once per block across all 25 training blocks, with the sequence randomized for each block. The locations of the target and distractors, as well as the orientation of the distractors, remained constant across the block.

For the novel training set, 12 target locations were randomly generated at the beginning of the experiment, with each target location defining one novel configuration. The targets were then paired with 11 newly and randomly generated distractors, with both the locations and orientations of the distractors varying randomly for each block. To restate, for each novel configuration across different blocks, the location of the target remained constant, but the locations and orientations of the distractors varied randomly.

Note that the orientation of the targets, for both repeated and novel configurations, was randomly varied with each presentation to prevent any potential response learning related to the orientation of the target. The locations of the targets were counterbalanced across the four quadrants of the display for repeated and novel of configurations. Additionally, the sequence of the 12 repeated/novel configurations within each block was randomized.

After the training phase, which included one repeated and one novel set, participants completed a five-block test session (Figure 2B). Each test block consisted of 24 trials, with half being repeated and half novel. Participants were permitted a short break after each block, in line with the standard contextual cueing paradigm (Chun & Jiang, 1998). The same generation rules used for the 12 repeated and novel configurations during the training phase were applied throughout the test phase. Specifically, repeated configurations were identical to those used in the training phase, except that the orientation of the target was randomly varied on each trial. Novel configurations retained the same target locations as in the training phase, but both the locations and orientations of the distractors, as well as the orientation of the target, were randomly varied for each presentation. These repeated and novel configurations were randomly intermixed, and their presentation sequence was randomized for each test block. In this case, neither the repeated nor the novel conditions could be regarded as continuous, precluding the application of frequency analysis to the test session.

Prior to the formal experiment, participants undertook a 24-trial practice phase aimed at acquainting them with the experimental task. Configurations used in the practice phase were not reused in the formal experiment. It was a prerequisite for participants to attain a minimum accuracy level of 85% before initiating the official experiment. If an accuracy level fell below that, the participant was instructed to repeat the practice phase. All participants successfully met the criteria within two repetitions of the practice phase.

Trial sequence

Each trial started with an 800-ms display of a central fixation cross (i.e., + sign). Subsequently, a visual search display, comprised of both the target and distractors, appeared and remained visible for a maximum duration of 1200 ms. During this period, participants were instructed to search for the target stimulus "T" and to press the corresponding response key (the left or right arrow key for a left- or right-tilted "T," respectively) as quickly and accurately as possible. The decision to allocate 1200 ms for display presentation was inspired by our pilot study, wherein it was observed that over 90% of trials elicited accurate responses within this time limitation.

After participants' responses or if the display time elapsed, a 500-ms blank screen, referred to as the inter-trial interval (ITI), was presented. Notably, even if participants responded within the 1200-ms window, the display remained visible for the full 1200 ms. This adaptation ensured uniform trial duration, maintaining consistency at 2500 ms per trial.

Data analysis

Participants' responses underwent analysis in both time and frequency domains. The time domain analysis involved assessing mean error rates, mean RTs, and the mean CVs for both training and test phases. In the frequency domain, the FFT analysis (see Castellanos et al., 2005) was exclusively conducted on the training phase wherein the repeated and novel contexts were segregated into separate blocks. For the analysis of variance (ANOVA) analysis, the Greenhouse–Geisser correction was used when Mauchly's test was violated.

Mean error rates and mean RTs: Statistical power was enhanced by grouping every five consecutive blocks into one epoch, resulting in five epochs for both the repeated and novel sets in the training phase. In the test phase, all five blocks were combined and treated as a sixth epoch for analysis. Repeated-measures ANOVA and paired-sample *t*-tests were applied to participants' mean RTs and error rates to assess features of contextual learning.

Mean RT variability: To assess the variability in participants' responses and the impact of contextual learning over the course of the experiment, we calculated the mean CVs for participants' RTs (Figure 2C). The CVs, which provide a normalized measure of variation, are derived by dividing the *SD* of the RT by the mean RT itself (i.e., CV = *SD*/RT). Importantly, two distinct types of CVs were calculated to evaluate two aspects of the participants' learning: template-based CV and cross-display CV.

In the computation of the template-based CVs, analogous to the RT analysis, blocks were aggregated into epochs: epochs 1 to 5 for the training session and epoch 6 for the test session. The CVs for each specific configuration within each epoch were then calculated, spanning five consecutive blocks (Figure 2C). Subsequently, the CVs for the 12 repeated and 12 novel configurations were averaged separately, yielding two distinct mean CV values. One mean CV value represents the average RT variability for repeated configurations, and the other reflects the variability of novel configurations, each within its respective epoch.

For the cross-display CVs, we measured participants' RT variability within each block containing 12 repeated or novel configurations. This was done by first determining the mean *SD* and mean RT for each type of display within each block. Subsequently, the CVs were computed using the formula CV = *SD*/mean RT for the repeated and novel contexts in each block. Finally, in the training session, the CVs from five consecutive blocks were collapsed into a single epoch (Figure 2C). Similarly, the test blocks were combined into one epoch. This aggregation was performed to enhance statistical power and to clearly define the learning epochs 1 to 5 during the training session and epoch 6 during the test session. Repeated-measures ANOVA and paired-sample *t*-tests were used to analyze both template-based CVs and cross-display CVs, aiming to assess the role of template-based learning and generic procedural learning in the contextual cueing effect.

RT frequency analysis: Frequency analysis of RTs was also conducted during the training phase using FFT to determine whether any specific patterns of RT variability existed. To maintain continuity in the collected RT data, missing trials with invalid responses (5.23%) were restored using linear interpolation, based on the RTs preceding and following the missing trial. The trials with incorrect responses (2.10%) were not omitted from the analysis of the mean RTs; instead, these error trials were treated on par with correct trials.

Recall that each training set was comprised of 300 trials, each lasting 2.5 seconds ($n = 300$, $\Delta t = 2.5$ seconds). Consequently, the valid frequency band for FFT analysis lies from 0.0027 Hz($1/[n\Delta t]/2$) = 1/750 second) to 0.2 Hz ($1/(2\Delta t) = 1/5$ second; see a similar calculation in Di Martino et al., 2008). This band can be divided into slow frequency variability and fast frequency variability (FFV) defined by the frequency of block reparations (0.033 Hz($1/[12\Delta t]$ = 1/30 second), with the former representing a gradual change in RT variation throughout the duration of the task and the latter representing moment-to-moment variation (Lewis, Reeve, Kelly, & Johnson, 2017). Our analysis focused only on FFV within the frequency range of 0.033 to 0.2 Hz, which was selected to examine trial-to-trial variations in cognitive processing within each experimental block. By concentrating on FFV, we aimed to gain a more refined understanding of the momentary fluctuations in attentional allocation and response efficiency throughout the task.

Prior to conducting the frequency analysis, it was essential to preprocess the RT data to avoid potential biases from the RT difference under different conditions. To this end, we implemented a two-step procedure that involved the normalization and detrending of RTs. Normalization was employed to account for scale disparities within the RT data. Specifically, for each participant, the RT of every trial—whether featuring repeated or novel displays—was divided by the overall mean RT of the corresponding display type in the training phase. This normalization process resulted in an equivalent normalized mean RT for both repeated and novel displays, thereby standardizing the data and neutralizing the influence of mean RT discrepancies between repeated and novel conditions. Detrending, on the other hand, aimed to remove the systematic changes; for example, RTs became faster as the experiment progressed, ensuring the integrity of the frequency analysis.

Following the normalization and detrending procedures, the response series from the 300 consecutive trials within each of the two training sets were analyzed for each participant. To process the data, we utilized Welch's averaged, modified periodogram method, which is a technique known to mitigate the noise typically present in the power spectral density estimates derived from FFT analysis (see similar methods in Helps, Broyd, Bitsakou, & Sonuga-Barke, 2011; Johnson, 2022). In more detail, the 300 trials, which constituted our dataset for FFT analysis, were segmented into five equal sub-segments, each comprised of 96 data points, with 48 data points being overlapped between the two consecutive segments. Each segment was then processed with the Hamming window and zero-padded to length 256, followed by FFT calculation. The segmentation rationale was to enable the detection of RT variability as it progressed throughout the experiment. After segmentation, the area under the spectrum (AUS) within our target frequency (0.033–0.2 Hz) was calculated and served as the final result, reflecting trial-to-trial variability. Repeated-measures ANOVA was conducted on the AUS, with Segment and Context as factors, to assess the effects of generic procedural learning.

Results

Error rates

Anticipatory responses (RT < 200 ms) were deemed outliers and excluded from the analysis, although none was observed in the current experiment. Trials with incorrect responses (1.83%) and without responses (4.49%) were considered error trials. Participants' overall mean error rates were 7.33% and 5.31% in the training and test phases, respectively (Figure 3A). The error rates observed in our study were relatively higher

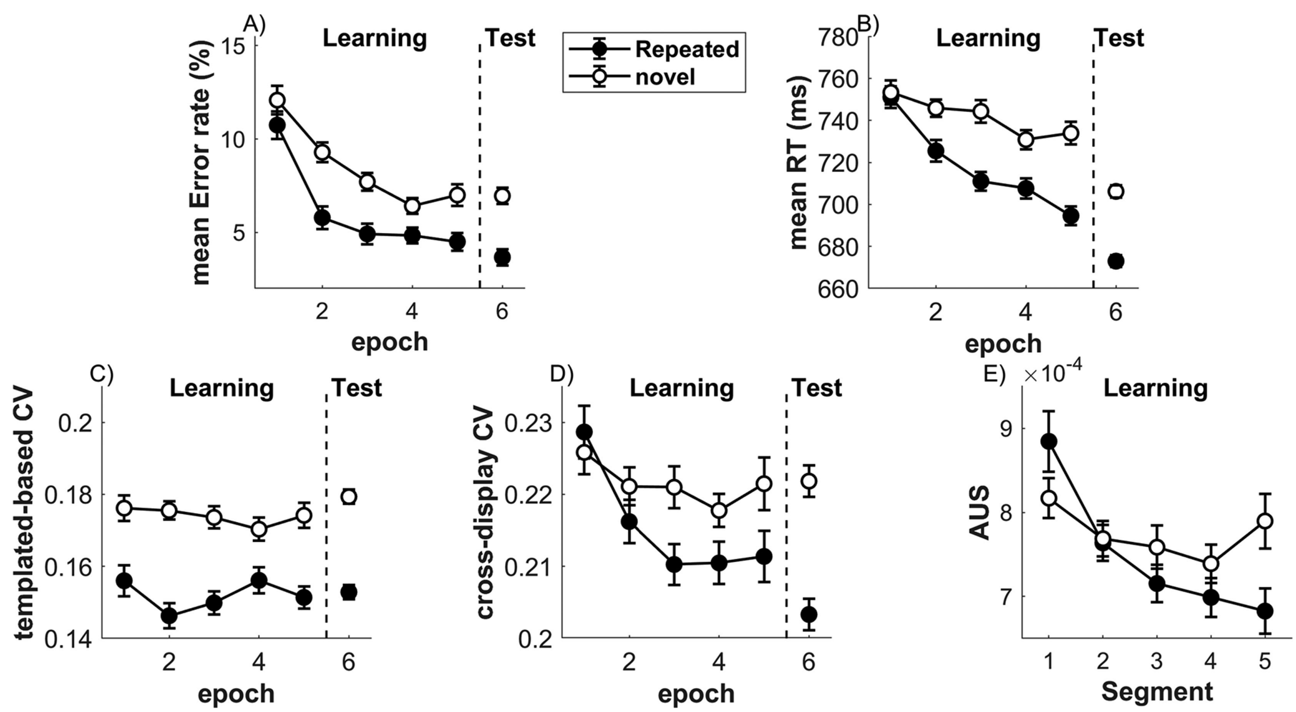

Figure 3. Results for Experiment 1. (**A**–**D**) Mean error rates, mean RTs, template-based CVs, and cross-display CVs as a function of Epoch (1–6) and Context (repeated, novel). (**E**) The AUS of the FFT analysis as a function of Segment (1–5) and Context (repeated, novel). The error bars depict the standard error of the mean within subjects. The open circle line indicates the novel context, and the filled circle lines represent the condition of the repeated context.

than those reported in previous studies (cf. Zang, Shi, Müller, & Conci, 2017; Zang, Zinchenko, Jia, Assumpção, & Li, 2018). This difference can be ascribed to the brief presentation duration of 1.2 seconds for the search displays.

For the training phase, repeated-measures ANOVA for the mean error rates with Context (repeated vs. novel) and Epoch (1–5) as factors revealed significant main effects for Context, $F(1, 39) = 13.556$, $p = 0.001$, $\eta_p^2 = 0.258$, with means of 6.16% and 8.50% for repeated and novel contexts, respectively (mean difference, 2.34%), and for Epoch, $F(3.232, 126.064) = 37.186$, $p < 0.001$, $\eta_p^2 = 0.488$. The mean error rates dropped from 11.42% in the first Epoch to 5.75% in the last Epoch (mean difference, 5.67%). Context × Epoch was not significant, $F(1, 39) = 1.601$, $p = 0.177$, $\eta_p^2 = 0.039$.

For the test phase, paired sample *t*-tests showed significantly lower error rates for the repeated context (3.67%) than for the novel context (6.96%) in the test phase, $t(39) = 6.578$, $p < 0.001$, Cohen's $d = 1.040$. These findings, taken together, indicate that participants reduced search errors over time, attributable to both contextual learning and practice effect.

Mean RT

Error trials were excluded for RT analysis, and results with Context and Epoch as factors are shown in Figure 3B. For the training phase, repeated-measures ANOVA considering Epoch (1–5) and Context (repeated, novel) as factors showed a significant main effect of Context, $F(1, 39) = 15.986$, $p < 0.001$, $\eta_p^2 = 0.291$. Mean contextual cueing effect of 23.69 ms (mean RTs of 717.93 ms and 741.62 ms for repeated and novel displays, respectively). The main effect of Epoch was also significant, $F(3.263, 127.267) = 18.362$, $p < 0.001$, $\eta_p^2 = 0.320$. Performances were 37.94 ms faster in Epoch 5 compared with Epoch 1, with mean RTs of 714.24 ms for Epoch 5 and 752.18 ms for Epoch 1. This finding suggests an RT improvement in performance through practice. The Context × Epoch interaction was also significant, $F(4, 156) = 6.839$, $p < 0.001$, $\eta_p^2 = 0.149$. Further post hoc analysis showed that the RT differences between repeated and novel contexts were significant from Epoch 3 onward when using a stringent criterion of $p < 0.01$ to account for multiple comparisons ($t > 3.139$, $p < 0.003$, Cohen's $d > 0.496$) but not for Epochs 1 and 2 ($t < 2.575$, $p > 0.014$, Cohen's $d < 0.413$). These results suggest that the contextual cueing effect developed during progression of the experiment.

In the test phase, a paired sample *t*-test with Context (repeated vs. novel) as the factor demonstrated a significant effect, $t(39) = 5.845$, $p < 0.001$, Cohen's $d = 0.924$. The mean RTs were 672.97 ms for repeated contexts and 706.18 ms for novel contexts, indicating a mean cueing effect of 33.21 ms.

RT variability in terms of CVs

The results of participants' template-based CVs and cross-display CVs are shown in Figures 3C and 3D. Repeated-measures ANOVA and paired sample *t*-tests were applied for the statistical analysis. For the template-based CVs, a significant main effect of Context was observed during the training phase, $F(1, 39) = 71.000$, $p < 0.001$, $\eta_p^2 = 0.645$, with mean CVs of 0.152 and 0.174 for the repeated and novel displays, respectively. These results indicate that participants were using a more stable and consistent search strategy for each repeated displays, resulting in less variability in their RTs. However, the main effect of Epoch was not significant, $F(4, 156) = 0.652$, $p = 0.626$, $\eta_p^2 = 0.016$. Recall that the CV is calculated by dividing the standard deviation by the mean (*SD*/mean), which normalizes the measure of dispersion by the magnitude of the dataset. The absence of a significant Epoch effect on CVs suggests that the trend of change in participants' *SD*s is congruent with that of their RTs (see further analysis in Supplementary Figure S1). This trend is consistent with the instance theory raised by Logan (1988).

The Context × Epoch interaction also was not significant, $F(1, 39) = 1.205$, $p = 0.311$, $\eta_p^2 = 0.030$, suggesting an early onset of contextual facilitation on the template-based CVs. During the test phase, a significant main effect of Context was also detected, $t(39) = 6.689$, $p < 0.001$, Cohen's $d = 1.058$, with mean CVs of 0.153 and 0.179 for the repeated and novel contexts, respectively, indicating a mean cueing effect of 0.026. This finding of overall lower template-based CVs for repeated displays supports template-based learning, suggesting that repeated displays are learned as templates that guide attention more efficiently.

For the cross-display CVs in the training phase, the main effect of Context was significant, $F(1, 39) = 4.245$, $p = 0.046$; however, the effect size was relatively modest, with a η_p^2 value of 0.098. The cross-display CV values were 0.216 for the repeated context and 0.222 for the novel context. To further investigate the impact of context in each training epoch, paired-sample *t*-tests were conducted. However, no significant differences were observed when applying a stringent criterion of $p < 0.01$ to account for multiple comparisons (all $t < 2.596$, all $p > 0.013$). This means that, although the main effect of Context was statistically significant, the magnitude of the effect was relatively small and requires careful interpretation. A significant main effect of Epoch was also observed, $F(3.568, 139.133) = 5.338$, $p = 0.001$, $\eta_p^2 = 0.120$, and the mean decreased from 0.228 in Epoch 1 to 0.216 in Epoch 5. The Context

$\times$ Epoch interaction was not significant, $F(4, 156) = 1.808$, $p = 0.130$, $\eta_p^2 = 0.044$. During the test phase, a significant main effect of Context was also observed, $t(39) = 4.119$, $p < 0.001$, Cohen's $d = 0.990$, with mean CVs of 0.208 and 0.225 for the repeated and novel contexts, respectively, and a mean cueing effect of 0.018. These results suggest a greater decrease in cross-display CVs for repeated displays, aligning with the predictions of the generic procedural learning account.

Synthesizing all results, both the template-based and cross-display CVs were lower in the repeated than the novel displays. However, the template-based CVs showed a significant difference earlier from the first training epoch, with strong statistical significance emerging early in the training phase. In contrast, the cross-display CVs demonstrated a small effect size in the training epochs and only reached strong significance during the test epoch.

RT frequency analysis

The results of the RT frequency analysis are shown in Figure 3E. A significant main effect of Segment was observed, $F(4, 156) = 7.876$, $p < 0.001$, $\eta_p^2 = 0.168$. Recall that the term Segment refers to the division of the 300 trials into five equal sub-segments. Each segment is comprised of 96 data points, with an overlap of 48 data points between consecutive segments. The average AUS exhibited a decreasing trend with progression of the experiment, as evidenced by the values in the first segment (8.51×10^{-4}) and the fifth segment (7.36×10^{-4}). This decrease in AUS corroborates the hypothesis that RT variability was reduced as the experiment progressed. Although no significant difference was observed for Context, $F(1, 39) = 1.429$, $p = 0.239$, $\eta_p^2 = 0.035$, the Context $\times$ Segment interaction was significant, $F(3.242, 126.443) = 2.892$, $p = 0.034$, $\eta_p^2 = 0.069$. Further post hoc analysis showed that the difference in the AUS between repeated and novel conditions was relatively weak: 1.071×10^{-4} in the fifth segment, $t(39) = 2.305$, $p = 0.027$ (note that the criteria considered multiple comparison should be 0.01), Cohen's $d = 0.364$. This result aligns with the cross-display CV analysis, which revealed a statistically significant yet modest context effect, reflecting the existence of generic procedural learning.

Interim discussion

In Experiment 1, participants underwent initial training phases with two uninterrupted sets, featuring either purely repeated or novel trials, followed by a test phase using the standard contextual cueing search paradigm. We assessed RT variability during the training phase using not only CV but also FFT frequency analysis to explore the role of template-based learning and generic procedural learning in contextual cueing. Specifically, template-based CVs were used to reflect template-based learning, and cross-display CVs and FFTs were used to assess generic procedural learning.

Significant contextual facilitation was observed during both the training and test phases, as evidenced by marked reductions in mean error rates and RTs. Furthermore, the template-based CVs were notably lower for repeated contexts compared with novel ones, suggesting that part of the contextual cueing effect may arise from template-based learning. This finding supports the traditional template-based attentional guidance account proposed in earlier studies (similar results have also been reported in previous research; see Fan et al., 2024; Yao, Luo, Fan, Qian, & Zang, 2024).

When considering cross-display CVs, a significant decrease in the repeated contexts compared with the novel ones during both the training and test phases was also observed, reflecting the engagement of generic procedural learning. Additionally, lower AUSs for the FFTs (0.033–0.2 Hz) for repeated contexts further supported this finding. These two pieces of evidence provide support for the proposed generic procedural learning account. Based on these results, we conclude that both template-based and generic procedural learning contribute to the contextual cueing effect.

However, the effects of cross-display CVs and AUSs in the training phase were relatively weak, which can be explained in two different ways: First, our study employed a modified version of the contextual cueing paradigm, where repeated and novel contexts were presented in separate sets. This differed from prior research that has typically intermixed repeated and novel trials, which may engender distinct patterns of contextual learning and lead to weak variability improvement. In other words, although general procedural learning remains an important component of contextual learning, it may manifest only when repeated and novel contexts are intermixed. Second, generic procedural learning did contribute to contextual learning, but it requires more learning time to become evident, potentially after several epochs of training. This is in contrast with template-based learning, where significant variability reduction was observed much earlier, such as from the first epoch. To differentiate these two possibilities, our subsequent Experiment 2 was designed with a standard contextual cueing paradigm first presented during the training set which was then followed by an uninterrupted test phase to assess the pattern of participants' search behavior after they had effectively established the contextual cueing effect.

Experiment 2

Methods

Participants first completed a 30-block training phase, with each block containing 12 repeated and 12 novel trials, where repeated and novel contexts were intermixed (i.e., same as the standard contextual cueing paradigm). Afterward, participants transitioned to the test phase with an uninterrupted contextual cueing search task. This test phase contains two sets: one consisting of repeated contexts and the other of novel contexts, each containing 300 trials (similar to the training phases in Experiment 1) (Figure 4). The remaining experimental settings closely mirrored those used in Experiment 1.

A new group of 40 naïve participants (36 females; mean age, 20.93 ± 1.98 years) were recruited in Experiment 2. As in Experiment 1, these participants provided informed consent and were given monetary compensation for participation. Half of the participants started the test phase with the repeated set (repeated first, 20 participants), and the other half began with the novel set (novel first, 20 participants) (see Supplementary Materials for details).

Results

Similar to Experiment 1, blocks within the training phase and test phase were aggregated into epochs, with every five blocks forming one epoch, resulting in six training epochs and five test epochs. The results are shown in Figure 5.

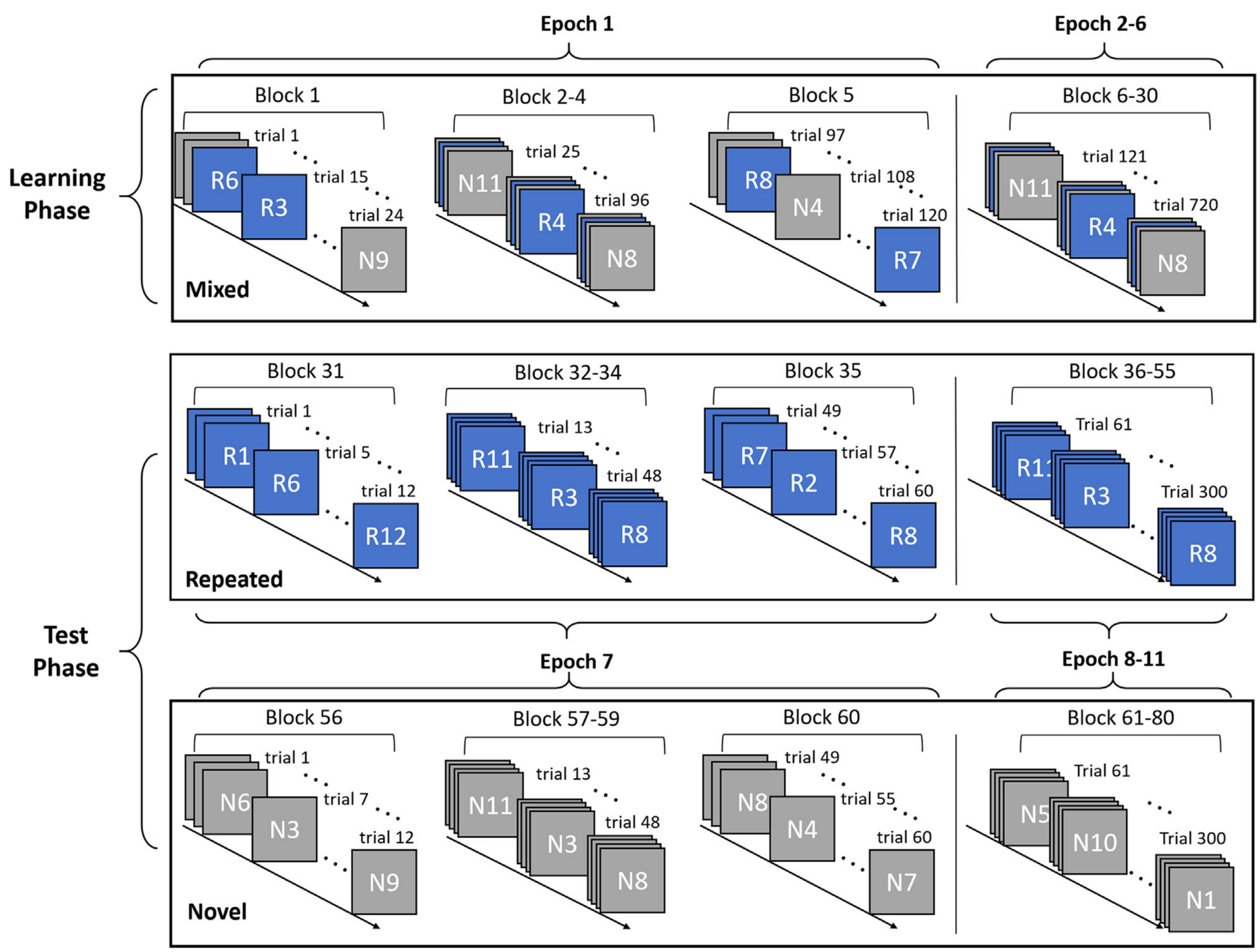

Figure 4. Schematic illustration of the experimental paradigm in Experiment 2. In the training phase, repeated and novel contexts were intermixed within the same blocks. There were 30 blocks, and each block contained 24 trials that were equally divided into two types of contexts: 12 repeated displays and 12 novel displays, presented in a randomized sequence. In the test phase, repeated and novel contexts were presented in separate sets (repeated or novel). Each set contained 25 blocks, and each block consisted of 12 trials of repeated or novel configurations. The same configurations were used as in the training phase.

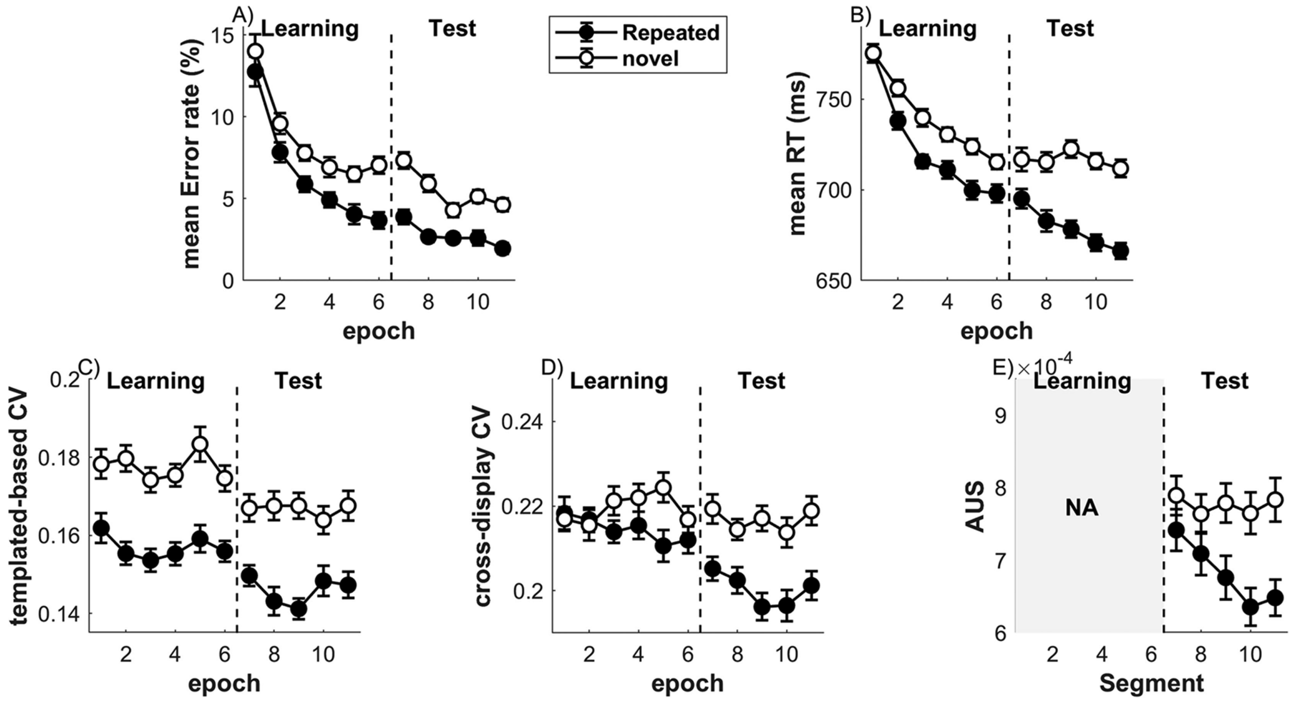

Figure 5. Results for Experiment 2. (**A**–**D**) Mean error rates, mean RTs, template-based CVs, and cross-display CVs as a function of Epoch (1–11) and Context (repeated, novel). (**E**) The AUS of the FFT analysis as a function of Segment (1–5) and Context (repeated, novel) in the test phase. The error bars depict the standard error of the mean within subjects. The open circle line indicates the novel context, and the filled circle lines represent the condition of the repeated context.

Error rates

Participants' mean error rates (wrong and missing trials) with Context and Epoch as factors are shown in Figure 5A. The mean error rates were 7.35% during the training phase and 4.01% during the test phase. Repeated-measures ANOVA for the error rates in the training phase with Context (repeated, novel) and Epoch (1–6) as factors revealed significant main effects for both Context, $F(1, 39) = 13.268$, $p = 0.001$, $\eta_p^2 = 0.254$, with means of 6.51% and 8.64% for repeated and novel contexts, respectively (mean difference, 2.13%), and Epoch, $F(2.379, 92.787) = 37.716$, $p < 0.001$, $\eta_p^2 = 0.492$. The error rates dropped from 13.38% in Epoch 1 to 5.35% in Epoch 6 (mean difference, 8.03%). Context × Epoch was not significant, $F(3.884, 151.468) = 1.110$, $p = 0.356$, $\eta_p^2 = 0.028$. These results suggest that participants were able to improve their search accuracy through both contextual learning and practice.

In the test phase, repeated ANOVA also revealed significant main effects for Context, $F(1, 39) = 43.264$, $p < 0.001$, $\eta_p^2 = 0.526$, within the repeated context (2.73%) compared with those within the novel context (5.46%), resulting in a mean difference of 2.73%, and for Epoch, $F(4, 156) = 12.049$, $p < 0.001$, $\eta_p^2 = 0.236$. The error rates decreased from 5.60% in Epoch 7 to 3.29% in Epoch 11. The Context × Epoch interaction was not significant, $F(4, 156) = 1.324$, $p = 0.264$, $\eta_p^2 = 0.033$, mainly attributable to the robust contextual cueing effect observed early in the test phase.

Response times

Error trials (5.83%) and trials shorter than 200 ms (0.08% of trials) were considered outliers for RT analysis (see results in Figure 5B). For the mean RT in the training phase, repeated-measures ANOVA that considered Epoch (1–6) and Context (repeated, novel) as factors showed significant main effects for Context, $F(1, 39) = 10.862$, $p = 0.002$, $\eta_p^2 = 0.218$, revealing a mean cueing effect of 17.21 ms (mean RTs of 722.95 ms and 740.16 ms for the repeated and novel displays, respectively), and for Epoch, $F(3.855, 150.363) = 54.648$, $p < 0.001$, $\eta_p^2 = 0.584$. These results indicate a practice effect, wherein Epoch 6 exhibited a 68-ms faster RT compared with Epoch 1 (706.70 ms and 775.29 ms for Epoch 6 and Epoch 1, respectively). The Context × Epoch interaction was also significant, $F(5, 195) = 4.223$, $p = 0.001$, $\eta_p^2 = 0.098$, suggesting that the contextual cueing effect developed with progression of the experiment. Using a stringent criterion of $p < 0.008$ to account for multiple comparisons, subsequent post hoc analysis showed that the contextual cueing effect became significant from Epoch 3 onward (all $t > 3.233$, all $p < 0.002$, Cohen's $d > 0.511$), whereas it was not evident in Epoch 1, $t(39) = 0.038$, $p = 0.970$, Cohen's $d = 0.006$, and was modest in Epoch 2 ($t = 2.499$, $p = 0.017$, Cohen's $d = 0.395$, and in Epoch 6 ($t = 2.517$, $p = 0.016$, Cohen's $d = 0.398$).

For the test phase (i.e., Epochs 7–11), repeated-measures ANOVA also revealed significant main effects for Context, $F(1, 39) = 30.262$, $p < 0.001$, $\eta_p^2 = 0.440$, with mean RTs of 678.48 ms and 716.43 ms for the repeated and novel contexts, respectively (with a mean cueing effect of 37.85 ms), and for Epoch, $F(2.593, 101.144) = 4.417$, $p = 0.008$, $\eta_p^2 = 0.102$, with mean RTs decreasing from 705.90 ms in Epoch 7 to 689.04 ms in Epoch 11. The Context × Epoch interaction was also significant, $F(1, 39) = 3.405$, $p = 0.011$, $\eta_p^2 = 0.080$, suggesting that the contextual cueing effect continued to develop throughout the test phase.

RT variability in terms of CVs

Both template-based CVs and cross-display CVs were calculated for the training and test phase to assess participants' variability. These CVs were further analyzed by repeated-measures ANOVA in the training phase and the test phase. The results are shown in Figures 5C and 5D. For the template-based CVs in the training phase, a significant main effect of Context was observed during the training phase, $F(1, 39) = 110.657$, $p < 0.001$, $\eta_p^2 = 0.739$, with mean CVs of 0.157 and 0.178 for the repeated and novel displays, respectively. However, the main effects of Epoch and Context × Epoch were not significant: Epoch, $F(5, 195) = 1.286$, $p = 0.271$, $\eta_p^2 = 0.032$; Context × Epoch, $F(5, 195) = 0.438$, $p = 0.822$, $\eta_p^2 = 0.011$. The lack of Epoch and Context × Epoch effects suggests that the template-based decreasing rates of the mean RTs and mean *SD*s of RTs were comparable; hence, the CVs, calculated with *SD*s normalized by RTs, remained relatively constant during progression of the experiment. The smaller template-based CVs in repeated contexts indicate that participants quickly adapted to using familiar templates, supporting the hypothesis of template-based learning.

As for cross-display CVs in the training phase, no significant main effect of Context was observed, $F(1, 39) = 3.873$, $p = 0.056$, $\eta_p^2 = 0.090$, with means of 0.214 and 0.219 for the repeated and novel displays, respectively. Neither the main effect of Epoch or Context × Epoch interaction reached significance: Epoch, $F(5, 195) = 0.350$, $p = 0.882$, $\eta_p^2 = 0.009$; Context × Epoch, $F(5, 195) = 1.757$, $p = 0.124$, $\eta_p^2 = 0.043$. These results suggest that contextual facilitation in cross-display CVs does not occur during the processing, indicating a delayed onset of generic procedural learning.

For the test phase (Epochs 7–11), repeated-measures ANOVA revealed a significant main effect of Context for the template-based CVs, $F(1, 39) = 35.102$, $p < 0.001$, $\eta_p^2 = 0.474$, with mean CVs of 0.16 for the repeated context and 0.18 for the novel context, but not for the main effects of Epoch, $F(1, 39) = 0.526$, $p > 0.717$, $\eta_p^2 < 0.012$, or Context × Epoch, $F(1, 39) = 0.855$, $p = 0.493$, $\eta_p^2 = 0.021$. There is a similar pattern for the cross-display CVs: Context: $F(1, 39) = 20.416$, $p < 0.001$, $\eta_p^2 = 0.344$, with mean CVs of 0.20 for the repeated context and 0.22 for the novel context; Epoch, $F(1, 39) = 1.665$, $p = 0.161$, $\eta_p^2 = 0.041$; Context × Epoch, $F(1, 39) = 0.713$, $p = 0.584$, $\eta_p^2 = 0.018$. These results suggest that both template-based and cross-display variability reduced significantly and remained stable after extensive learning of the initial training phase.

Results of frequency analysis during the test phase

For the frequency analysis, anticipatory trials (0.08%) and missing trials without responses (1.29%) in test phases were restored via linear interpolation to ensure a continuous time series. Error trials with incorrect responses (2.81%) were treated identically to correct trials without any specialized processing. Similar to Experiment 1, the series of RTs were normalized and detrended across participants before analysis. Subsequently, the data were divided into five segments of 96 data points, each overlapping by 48 data points, and subjected to Hamming windowing and zero-padding to a length of 256, followed by FFT calculation. Then repeated-measures ANOVA with Context (repeated vs. novel) and Segment (1–5) as factors was applied to the AUS during the test phase.

Repeated-measures ANOVA showed a significant main effect of Context for the AUS, ranging from 0.033 to 0.2 Hz, $F(1, 39) = 10.910$, $p = 0.002$, $\eta_p^2 = 0.219$, with $6.82 \pm 1.94 \times 10^{-4}$ in the repeated context and $7.76 \pm 1.39 \times 10^{-4}$ in the novel context. There were no significant main effects of Segment or Context × Segment, all $F < 1.242$, all $p > 0.196$, all $\eta_p^2 < 0.038$. This result is consistent with the cross-display CV analysis from the test phase, which showed a statistically significant context effect. This pattern aligns with the cross-display CV analysis, indicating late onset of general procedural learning that remained stable throughout the test phase.

Interim discussion

In Experiment 2, participants first underwent a training phase using the standard contextual cueing paradigm, followed by a test phase with two uninterrupted sets featuring either repeated or novel trials. The results indicated a robust contextual cueing effect during both the training and test phases, with faster search times and fewer errors. Furthermore, lower template-based CVs for repeated displays were observed compared with novel ones, supporting the role of template-based learning.

Additionally, although cross-display CVs in the training phase did not reach significance between repeated and novel displays, both CVs and AUSs from the FFT analysis in the test phase showed a reduction in repeated compared with novel ones, reflecting a decrease with extensive training (for more details on comparisons between phases, see Supplementary Table S3). These results suggest that participants' variability across different configurations was also reduced, supporting the role of generic procedural learning in contextual cueing. In other words, a generic search strategy could be developed for a visual search task, regardless of whether the context is repeated or novel. However, because repeated contexts appear more frequently than novel ones, the generic search strategy becomes more effective in repeated contexts, leading to less variability in searching repeated displays compared with novel ones.

It is important to note that, in Experiment 1, a significant but modest reduction in cross-display CVs and AUSs was observed during the early training phase. However, in the test phase of Experiment 2, after participants received thorough training, stronger facilitation in cross-display variation was observed. These results suggest that the reduction of cross-display variation takes time.

General discussion

This study investigated whether the contextual cueing effect is driven by template-based learning or generic procedural learning, using a variant of the contextual cueing paradigm. We analyzed not only traditional error rates and RTs—which reflect the acquisition of contextual cueing—but also RT variability, using CV and time–frequency analyses of RTs across two experiments. These new analyses could provide empirical evidence for template-based learning or generic procedural learning in contextual cueing. In Experiment 1, participants underwent an uninterrupted training phase that separated the repeated and novel displays into two distinct sets, followed by a standard test phase with mixed displays. In Experiment 2, participants first learned the context using the standard paradigm (i.e., mixed displays) and were then tested with the uninterrupted design.

Both experiments revealed significant contextual cueing effects, with faster RTs for repeated displays compared with novel displays, confirming the existence of the classic contextual cueing effect (Liu, Ma, Zhao, & Sun, 2024; Vadillo, Giménez-Fernández, Beesley, Shanks, & Luque, 2021; Xie, Chen, & Zang, 2020; Zang et al., 2022). Interestingly, both template-based CVs and cross-display CVs were smaller for repeated displays compared with novel displays. The FFT analysis further revealed reduced RT variability at higher frequency ranges (0.033–0.2 Hz) in repeated display, reflecting variation across all display types within the same block. Based on the RT variability analysis, we found supportive evidence for both template-based learning and procedural learning, which we will discuss in more detail below.

Regarding the template-based CV results, it was significantly lower for repeated displays than for novel displays in both experiments, across both the training and test phases. Similar findings have been documented in previous studies (Fan et al., 2024; Yao et al., 2024). However, these studies did not refer to this approach as template-based CVs, instead using the term "RT variability." These studies found that RT variability (calculated as *SD*/RT for each specific configuration) was smaller for repeated displays than for novel displays for both college students and older adults (60+ years) (Yao et al., 2024), but not for young children (8–9 years) (Fan et al., 2024). These results, combined with our current findings, suggest that template-based CVs are a reliable indicator of the contextual cueing effect, at least for adult participants.

In the current study, we linked template-based CVs to the classic template-based attentional guidance account, which we refer to as template-based learning. We hypothesize that repeated exposure to a specific display or configuration enables participants to develop a template-based search strategy. When the same display is encountered again, this strategy is retrieved and activated, enabling participants to direct their attention more efficiently to the target location, thereby enhancing search performance. Furthermore, because the same search strategy is applied to repeated displays, we would expect RT variability to be lower for each repeated display compared with novel one, where different search strategies are used for each presentation. Based on this reasoning, the observed reduction in template-based CVs for repeated displays provides empirical evidence supporting the role of template-based learning in the contextual cueing effect.

Previous studies have reported on template-based CV results (Fan et al., 2024; Yao et al., 2024), but, to our knowledge, no research has investigated cross-display variability (i.e., RT variability across different repeated/novel displays within an experiment). Analyzing cross-display variability is crucial for understanding the role of generic procedural learning in contextual cueing: If a general search strategy is developed for all of the displays, whether they are repeated or novel, cross-display CVs should decrease for repeated displays due to their higher occurrence frequency and greater impact on optimization of the search strategy. In our current study, we found that the cross-display CVs for repeated displays was smaller compared with novel displays in both experiments. This was further supported by frequency analysis (FFT), which reduced power (albeit moderately during early training phases) for repeated displays in the higher frequency range (0.033–0.2 Hz), This suggests that contextual learning reduces RT variability across different displays, providing support for the role of generic procedural learning in contextual cueing.

Together, the overall lower template-based and cross-display CVs in repeated displays (as well as FFT results) suggest that both template-based and generic procedural learning contribute to the contextual cueing effect, although they may operate at different stages of the search process. Specifically, we propose that template-based learning plays a dominant role early in training, whereas generic procedural learning requires more extensive training to become significant. This conclusion is supported by our findings: The template-based variation, linked to template-based learning, showed a significant difference between repeated and novel displays as early as the first epoch of the training phase. In contrast, the cross-display variation, associated with generic procedural learning, remained relatively weak throughout the training phase and only became robust during the subsequent test phase. Notably, unlike the study by Seitz et al. (2023), which focused primarily on generic procedural learning in contextual cueing while largely overlooking template-based learning, our study examined the contributions of both factors, underscoring their distinct roles in the cueing effect.

The assumption that participants first benefit from template-based learning and later shift to procedural learning is reasonable. Initially, participants may quickly learn some of the repeated templates, leading to the development of a preliminary search strategy. This early strategy enhances search efficiency by allowing participants to recognize and respond more quickly to familiar configurations. Indeed, some studies observed the onset of the contextual cueing effect as early as the second block when participants encounter the search display for the second time (Bergmann & Schubö, 2021; Zang, Huang, Zhu, Müller, & Shi, 2020). However, developing a more generalized search strategy that applies to a broader range of displays requires more time and exposure. Over time, participants refine a search strategy that can be applied across various contexts, not just those with familiar templates. Therefore, procedural learning typically requires extended practice and exposure to diverse trials (Anderson, 1982; Beaunieux et al., 2006). The transition from template-based learning to procedural learning reflects a natural progression in the acquisition of expertise.

It is worth noting that some may question certain aspects of the experimental design and consider them as potential factors influencing the results and conclusions. Two points in particular deserve mention. First, the stimulus presentation duration was fixed at 1200 ms. Second, the sequence of presenting the repeated set first followed by the novel set, or vice versa, could also be a concern. The fixed display presentation duration of 1200 ms was selected to maintain a consistent trial duration of 2.5 seconds, which included the fixation duration and ITI. This design ensured a uniform sampling rate (one RT data point every 2.5 seconds), a necessary prerequisite for FFT frequency analysis (see similar settings in Di Martino et al., 2008). This design in our study differed significantly from the longer durations typically used in classic contextual cueing studies, which often exceed 2.5 seconds (e.g., Liu et al., 2024; Vadillo et al., 2021). This discrepancy could potentially limit our conclusions, particularly regarding whether template-based and procedural learning mechanisms can operate effectively under the 1200-ms condition. Moreover, the relatively short duration might introduce an artificial floor effect, potentially influencing the results. However, we consider this unlikely, as our findings demonstrate a robust contextual cueing effect, evidenced by faster search times and lower error rates in repeated displays compared with novel displays. Additionally, prior research has shown that contextual learning can occur with much shorter display durations—for example, as brief as 300 ms (Xie et al., 2020) or even 100 ms (Kobayashi & Ogawa, 2020). Nonetheless, to fully resolve this issue, future studies employing unlimited presentation durations are warranted.

With regard to the presentation sequence, one might question whether the sequence in which contexts are presented—repeated first or novel first—could influence contextual learning behavior, particularly given the established primacy and recency effects. For example, Jungé et al. (2007) reported a significant contextual cueing effect when participants first encountered repeated displays in the initial trials, followed by novel displays. In contrast, when participants were exposed to novel displays first and then repeated displays, no contextual cueing effect was observed. From these findings, Jungé et al. (2007) concluded that a primacy effect plays a role in contextual cueing. However, Vaskevich and Luria (2019) challenged this notion by comparing conditions where participants first encountered purely random displays (random-first condition) with those where they were first exposed to a mixture of random and repeated displays (mixed-first condition). In the latter test phase, a comparable contextual cueing effect was observed, suggesting that encountering random displays first does not inhibit contextual learning, thereby conflicting with the primacy effect proposed by Jungé et al. (2007). Luhmann (2011) also questioned the primacy effect by associating repeated contexts with one target location in the first half of the experiment and then with a different target location in the second half. They found a significant contextual cueing effect for the second (but not the first) target location in the post-test phase, leading them to suggest a recency effect in contextual learning. In the uninterrupted phase of our current study, we did not observe any significant difference between the repeated-first and novel-first sets (see Supplementary Materials). This finding runs counter to both the primacy and recency effects, probably due to the relatively long break time (2–5 hours) between the repeated and novel sets. Taken together, these studies suggest that both primacy and recency effects can occur, with the specific outcome likely depending on factors such as the nature of the displays, the presence or absence of repeated signals, the pairing relationship between the target and distractors, and the break duration of exposure to different conditions. Future research should continue to explore the temporal dynamics of contextual cueing to further clarify the sequence effects in this phenomenon.

Conclusions

This study explored the roles of template-based attention guidance and generic procedural learning in contextual cueing through uninterrupted visual search tasks conducted across two experiments. By analyzing participants' RT variability from multiple perspectives—including template-based CVs, cross-display CVs, and FFT analysis—we found evidence supporting both mechanisms. Notably, template-based learning, associated with attention guidance account, was found to play a more prominent role during the early stages of learning, whereas generic procedural learning required extended practice to contribute to the improvements in search speed.

Keywords: contextual cueing, visual search, response time variability, fast fourier transform, procedural learning

Acknowledgments

Supported by the National Natural Science Foundation of China (NSFC 32071042). The funding bodies played no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and codes: The data and data processing codes (based on MATLAB) for the experiments are available via the Open Science Framework: https://osf.io/wt9d6/?view_only=12ddf75e1f5945f6b4344c3b292226f7. None of the experiments was preregistered.

Commercial relationships: none.
Corresponding author: Xuelian Zang.
Email: zangxuelian@gmail.com.
Address: Center for Cognition and Brain Disorders, Affiliated Hospital of Hangzhou Normal University, Hangzhou Normal University, Hangzhou, Zhejiang 310004, China.

*HY and SZ contributed equally to this work.

References

Adamo, N., Di Martino, A., Esu, L., Petkova, E., Johnson, K., Kelly, S., . . . Zuddas, A. (2014). Increased response-time variability across different cognitive tasks in children with ADHD. *Journal of Attention Disorders, 18*(5), 434–446, https://doi.org/10.1177/1087054712439419.

Adamo, N., Huo, L., Adelsberg, S., Petkova, E., Castellanos, F. X., & Martino, A. D. (2014). Response time intra-subject variability: Commonalities between children with autism spectrum disorders and children with ADHD. *European Child & Adolescent Psychiatry, 23*(2), 69, https://doi.org/10.1007/s00787-013-0428-4.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369–406, https://doi.org/10.1037/0033-295X.89.4.369.

Beaunieux, H., Hubert, V., Witkowski, T., Pitel, A.-L., Rossi, S., Danion, J.-M., . . . Eustache, F. (2006). Which processes are involved in cognitive procedural learning? *Memory, 14*(5), 521–539, https://doi.org/10.1080/09658210500477766.

Bergmann, N., & Schubö, A. (2021). Local and global context repetitions in contextual cueing. *Journal of Vision, 21*(10):9, 1–17, https://doi.org/10.1167/jov.21.10.9.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436, https://doi.org/10.1163/156856897X00357.

Castellanos, F. X., Sonuga-Barke, E. J., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry, 57*(11), 1416–1423, https://doi.org/10.1016/j.biopsych.2004.12.005.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71, https://doi.org/10.1006/cogp.1998.0681.

Di Martino, A., Ghaffari, M., Curchack, J., Reiss, P., Hyde, C., Vannucci, M., . . . Castellanos, F. X. (2008). Decomposing intra-subject variability in children with attention-deficit/hyperactivity disorder. *Biological Psychiatry, 64*(7), 607–614, https://doi.org/10.1016/j.biopsych.2008.03.008.

Fan, C., Zinchenko, A., Chen, L., Wu, J., Qian, Y., & Zang, X. (2024). Invariant contexts reduce response time variability in visual search in an age-specific way: A comparison of children, teenagers, and adults. *Attention, Perception, & Psychophysics, 86*(6), 1974–1988, https://doi.org/10.3758/s13414-024-02926-2.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191, https://doi.org/10.3758/BF03193146.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. Melton (Ed.), *Categories of Human Learning* (pp. 243–285). Amsterdam: Elsevier, https://doi.org/10.1016/B978-1-4832-3145-7.50016-9.

Geyer, T., Mueller, H. J., Assumpcao, L., & Gais, S. (2013). Sleep-effects on implicit and explicit memory in repeated visual search. *PLoS One, 8*(8), e69953, https://doi.org/10.1371/journal.pone.0069953.

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences, 19*(9), 524–533, https://doi.org/10.1016/j.tics.2015.07.009.

Harris, A. M., & Remington, R. W. (2017). Contextual cueing improves attentional guidance, even when guidance is supposedly optimal. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 926–940, https://doi.org/10.1037/xhp0000394.

Helps, S. K., Broyd, S. J., Bitsakou, P., & Sonuga-Barke, E. J. S. (2011). Identifying a distinctive familial frequency band in reaction time fluctuations in ADHD. *Neuropsychology, 25*(6), 711–719, https://doi.org/10.1037/a0024479.

Helton, W. S., & Russell, P. N. (2015). Rest is best: The role of rest and task interruptions on vigilance. *Cognition, 134*, 165–173, https://doi.org/10.1016/j.cognition.2014.10.001.

Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin & Review, 14*(5), 834–839, https://doi.org/10.3758/BF03194108.

Johnson, K. (2022). *FFT and ExGaussian Matlab scripts*. Retrieved from https://doi.org/10.17605/OSF.IO/NTWY7.

Johnson, K. A., Kelly, S. P., Bellgrove, M. A., Barry, E., Cox, M., Gill, M., . . . Robertson, I. H. (2007). Response variability in attention deficit hyperactivity disorder: Evidence for neuropsychological heterogeneity. *Neuropsychologia, 45*(4), 630–638, https://doi.org/10.1016/j.neuropsychologia.2006.03.034.

Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition, 15*(1), 1–11, https://doi.org/10.1080/13506280600859706.

Kobayashi, H., & Ogawa, H. (2020). Contextual cueing facilitation arises early in the time course of visual search: An investigation with the ‘speed-accuracy tradeoff task. *Attention, Perception, & Psychophysics, 82*(6), 2851–2861, https://doi.org/10.3758/s13414-020-02028-9.

Kroell, L. M., Schlagbauer, B., Zinchenko, A., Müller, H. J., & Geyer, T. (2019). Behavioural evidence for a single memory system in contextual cueing. *Visual Cognition, 27*(5–8), 551–562, https://doi.org/10.1080/13506285.2019.1648347.

Lewis, F. C., Reeve, R. A., Kelly, S. P., & Johnson, K. A. (2017). Sustained attention to a predictable, unengaging Go/No-Go task shows ongoing development between 6 and 11 years. *Attention, Perception, & Psychophysics, 79*(6), 1726–1741, https://doi.org/10.3758/s13414-017-1351-4.

Liu, X., Ma, J., Zhao, G., & Sun, H.-J. (2024). The effect of gaze information associated with the search items on contextual cueing effect. *Attention, Perception, & Psychophysics, 86*(1), 84–94, https://doi.org/10.3758/s13414-023-02817-y.

Liu, Z.-Z., Qu, H.-J., Tian, Z.-L., Han, M.-J., Fan, Y., Ge, L.-Z., . . . Zhang, H. (2017). Reproducibility of frequency-dependent low frequency fluctuations in reaction time over time and across tasks. *PLoS One, 12*(9), e0184476, https://doi.org/10.1371/journal.pone.0184476.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527, https://doi.org/10.1037/0033-295X.95.4.492.

Luhmann, C. C. (2011). Integrating spatial context learning over contradictory signals: Recency effects in contextual cueing. *Visual Cognition, 19*(7), 846–862, https://doi.org/10.1080/13506285.2011.586653.

Machida, K., Murias, M., & Johnson, K. A. (2019). Electrophysiological correlates of response time variability during a sustained attention task. *Frontiers in Human Neuroscience, 13*, 363, https://doi.org/10.3389/fnhum.2019.00363.

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search? *Psychological Research, 73*(2), 212–221, https://doi.org/10.1007/s00426-008-0211-1.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442, https://doi.org/10.1163/156856897X00366.

Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics, 63*(7), 1239–1249, https://doi.org/10.3758/BF03194537.

Schankin, A., & Schubö, A. (2009). Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. *Psychophysiology, 46*(3), 668–679, https://doi.org/10.1111/j.1469-8986.2009.00807.x.

Seitz, W., Zinchenko, A., Müller, H. J., & Geyer, T. (2023). Contextual cueing of visual search reflects the acquisition of an optimal, one-for-all oculomotor scanning strategy. *Communications Psychology, 1*(1), 20, https://doi.org/10.1038/s44271-023-00019-8.

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, & Psychophysics, 81*(8), 2571–2589, https://doi.org/10.3758/s13414-019-01832-2.

Tseng, Y.-C., & Li, C.-S. R. (2004). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics, 66*(8), 1363–1378, https://doi.org/10.3758/BF03195004.

Vadillo, M. A., Giménez-Fernández, T., Beesley, T., Shanks, D. R., & Luque, D. (2021). There is more to contextual cueing than meets the eye: Improving visual search without attentional guidance toward predictable target locations. *Journal of Experimental Psychology: Human Perception and Performance, 47*(1), 116–120, https://doi.org/10.1037/xhp0000780.

Vaskevich, A., & Luria, R. (2019). Statistical learning in visual search is easier after experience with noise than overcoming previous learning. *Visual Cognition, 27*(5–8), 537–550, https://doi.org/10.1080/13506285.2019.1615022.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review, 28*(4), 1060–1092, https://doi.org/10.3758/s13423-020-01859-9.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 419–433, https://doi.org/10.1037/0096-1523.15.3.419.

Xie, X., Chen, S., & Zang, X. (2020). Contextual cueing effect under rapid presentation. *Frontiers in Psychology, 11*, 603520, https://doi.org/10.3389/fpsyg.2020.603520.

Yao, Y., Luo, R., Fan, C., Qian, Y., & Zang, X. (2024). Age-related contextual cueing features are more evident in reaction variability than in reaction time. *Quarterly Journal of Experimental Psychology, 78*(3), 604–618, https://doi.org/10.1177/17470218241241954.

Zang, X., Huang, L., Zhu, X., Müller, H. J., & Shi, Z. (2020). Influences of luminance contrast and ambient lighting on visual context learning and retrieval. *Attention, Perception, & Psychophysics, 82*(8), 4007–4024, https://doi.org/10.3758/s13414-020-02106-y.

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*(3), 807–819, https://doi.org/10.1037/xlm0000060.

Zang, X., Shi, Z., Müller, H. J., & Conci, M. (2017). Contextual cueing in 3D visual search depends on representations in planar-, not depth-defined space. *Journal of Vision, 17*(5):17, 1–11, https://doi.org/10.1167/17.5.17.

Zang, X., Zinchenko, A., Jia, L., Assumpção, L., & Li, H. (2018). Global repetition influences contextual cueing. *Frontiers in Psychology, 9*, 402, https://doi.org/10.3389/fpsyg.2018.00402.

Zang, X., Zinchenko, A., Wu, J., Zhu, X., Fang, F., & Shi, Z. (2022). Contextual cueing in co-active visual search: Joint action allows acquisition of task-irrelevant context. *Attention, Perception, & Psychophysics, 84*(4), 1114–1129, https://doi.org/10.3758/s13414-022-02470-x.

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*(6):11, 1–13, https://doi.org/10.1167/12.6.11.

Zinchenko, A., Conci, M., Hauser, J., Müller, H. J., & Geyer, T. (2020). Distributed attention beats the down-side of statistical context learning in visual search. *Journal of Vision, 20*(7):, 1–14, https://doi.org/10.1167/jov.20.7.4.

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020). Automatic guidance (and misguidance) of visuospatial attention by acquired scene memory: Evidence from an N1pc polarity reversal. *Psychological Science, 31*(12), 1531–1543, https://doi.org/10.1177/0956797620954815.

Zinchenko, A., Geyer, T., Zang, X., Shi, Z., Müller, H. J., & Conci, M. (2024). When experience with scenes foils attentional orienting: ERP evidence against flexible target-context mapping in visual search. *Cortex, 175*, 41–53, https://doi.org/10.1016/j.cortex.2024.04.001.