Interaction between scene-based and array-based contextual cueing

Gail M. Rosenbaum · Yuhong V. Jiang

Published online: 10 April 2013
© Psychonomic Society, Inc. 2013

Abstract *Contextual cueing* refers to the cueing of spatial attention by repeated spatial context. Previous studies have demonstrated distinctive properties of contextual cueing by background scenes and by an array of search items. Whereas scene-based contextual cueing reflects explicit learning of the scene–target association, array-based contextual cueing is supported primarily by implicit learning. In this study, we investigated the interaction between scene-based and array-based contextual cueing. Participants searched for a target that was predicted by both the background scene and the locations of distractor items. We tested three possible patterns of interaction: (1) The scene and the array could be learned independently, in which case cueing should be expressed even when only one cue was preserved; (2) the scene and array could be learned jointly, in which case cueing should occur only when both cues were preserved; (3) overshadowing might occur, in which case learning of the stronger cue should preclude learning of the weaker cue. In several experiments, we manipulated the nature of the contextual cues present during training and testing. We also tested explicit awareness of scenes, scene–target associations, and arrays. The results supported the overshadowing account: Specifically, scene-based contextual cueing precluded array-based contextual cueing when both were predictive of the location of a search target. We suggest that explicit, endogenous cues dominate over implicit cues in guiding spatial attention.

Keywords Attention · Visual search · Implicit/explicit learning · Contextual cueing

Visual attention is affected not only by salient sensory input and an observer's goals, but also by one's previous experience with the environment. Locations that were important in the past tend to receive attentional priority in future encounters (Torralba, 2003), a phenomenon known as *contextual cueing* (Chun, 2000; Chun & Jiang, 1998; Oliva & Torralba, 2007). At least two types of contextual cueing have been discovered, reflecting explicit and implicit learning of visual contexts. In studies of scene-based contextual cueing, participants search for a target embedded in a natural scene. When the scene is shown repeatedly, participants acquire explicit knowledge about the target's location in the scene and deploy attention accordingly (Brockmole, Castelano, & Henderson, 2006; Brockmole & Henderson, 2006a, b; Ehinger & Brockmole, 2008). In studies on array-based contextual cueing, participants search for a target among an array of distractors. When the array locations are repeatedly presented, participants find the target faster on repeated arrays than on unrepeated arrays, even though they are unaware of the array repetition (Chun & Jiang, 1998). Both scene- and array-based contextual cueing may be important in shaping visual attention in the real world. For example, the task of finding one's favorite cereal in a supermarket involves not only a background scene but also distractor objects that may be mistaken for the cereal. However, most previous studies on scene- and array-based contextual cueing have focused on one type or the other. The purpose of this study is to examine the relationship between these two types of contextual cueing.

The main question that we investigated is whether explicit learning of the scene–target association occurs in parallel to implicit learning of the array–target association when both cues are present. In our experiments, participants

G. M. Rosenbaum · Y. V. Jiang
Department of Psychology, University of Minnesota, Minneapolis,
MN, USA

Y. V. Jiang
e-mail: jiang166@umn.edu

G. M. Rosenbaum (✉)
Department of Psychology, Temple University, 714 Weiss Hall,
1701 North 13th Street,
Philadelphia, PA 19122, USA
e-mail: gailrosenbaum@gmail.com

searched for a T target among L distractors displayed against a natural scene (Fig. 1). To mimic real-world situations in which both the scene and distractor locations tend to repeat over time, we repeated both the scenes and the arrays several times, such that both the background scene and the search array in a particular display were predictive of the target's location. Three possibilities existed: The scene and the array might be learned independently, they might be learned jointly, or learning of one cue might prevent learning of the other.

Brooks, Rasmussen, and Hollingworth (2010) presented suggestive evidence that the scene and array may be learned jointly. In their study, Brooks et al. examined the interaction between local and global spatial context on cueing. Using computer-rendered 3-D images, they placed search items on a large table, which was situated in a global scene. The items on the table—the array—were the local context, whereas the peripheral region of the scene was the global context. To examine whether people could learn the local and global contexts independently, Brooks et al. repeated both the array and the scene in a training phase. Subsequently, they tested participants on displays in which the scene was random but the array was predictive, or in which the array was random but the scene was predictive. They found no transfer in either case, suggesting that the scene and the array were learned jointly: Learning could not be expressed unless both cues were present.

Although Brooks et al.'s (2010) study was consistent with the joint-learning hypothesis, it is unclear how

generalizable the results were. Because Brooks et al. were interested in comparing local and global context learning, they limited the array exclusively to local regions and the scene exclusively to peripheral regions. As a result, the scene context might have been substantially weakened, both because it was far from the target and because it occupied a small part of the display. Because contextual cueing is weaker for far than for near context (Brady & Chun, 2007; Olson & Chun, 2001), displacing the array and the scene to local and global regions might have changed how the two types of cues interacted. In addition, because Brooks et al.'s study was not intended to examine the interaction between implicit and explicit cueing, no information about participants' awareness of the cues was provided. For these reasons, it is unclear whether explicit, scene-based cues and implicit, array-based cues are always learned jointly.

Our study was adapted from Brooks et al. (2010), with the following modifications. First, the array was placed directly on the scene, so both the array and the scene provided local and global context for the target. Second, we administered a recognition test to confirm the explicit nature of scene-based cueing and the implicit nature of array-based cueing.

Experiment 1: Scene+array predictive during training, scene predictive during testing

The purpose of Experiment 1 was to dissociate the joint-learning account from an independent-learning account and an overshadowing account. The participants conducted visual search for a T among an array of Ls placed against a background scene (Fig. 1). In the training phase, the target's location was consistently associated with the same scene and the same search array (*scene+array* predictive). Instead of restricting the array to a local context and the scene to a global context, as in Brooks et al. (2010), we placed search items in many locations over a natural scene. In a subsequent testing phase, we replaced the predictive array with a random, unpredictable array, leaving the scene as the sole predictive search cue (*scene-predictive* display).

If the scene and array are learned jointly, cueing should not transfer to displays with predictive scenes but random arrays (Brooks et al., 2010). The independent-learning hypothesis, in contrast, predicts that cueing should partially transfer to scene-predictive displays, although cueing might be weakened by the lack of predictive arrays. This hypothesis has been supported by studies on human memory, which have shown that exposure to experimental materials, such as a list of words, leads to both explicit knowledge and implicit memory of the materials (Schacter, 1996). Given that scene-based and array-based contextual cueing employ explicit and implicit learning,

Fig. 1 A sample search display used in this study. Participants searched for a T among Ls and pressed an arrow key to report the orientation of the T

respectively, it is possible that the two types of learning may develop independently.

Overshadowing is the third possible account. Scene-based contextual cueing (in the absence of search arrays) usually leads to a response time (RT) gain of 1–2 s (Brockmole & Henderson, 2006a, b), which is about ten times greater than array-based contextual cueing (Chun & Jiang, 1998). In addition, because scene-based contextual cueing is explicit, it may be more effective at guiding visual attention than is array-based implicit learning. In fact, two previous studies have suggested that spatial attention is more effectively guided by endogenous than by implicit cues (Jiang, Swallow, & Rosenbaum, 2013; Kunar, Flusberg, & Wolfe, 2006). If scene-based contextual cueing overshadows array-based contextual cueing, then only scene-based cueing should be observed when both cues are present. Consequently, learning should fully transfer from scene+array predictive to scene-predictive displays.

Thus, by examining the degree of transfer from scene+array predictive displays to scene-predictive displays, we would be able to dissociate the joint-learning hypothesis (no transfer) from the independent-learning hypothesis (transfer) or the overshadowing hypothesis (transfer).

Method

Participants

Students from the University of Minnesota were participants in all of the experiments reported in this study. They all had normal or corrected-to-normal visual acuity and were naïve as to the purpose of the study. Their ages ranged from 18 to 35 years. We obtained informed consent prior to the experiment, and participants were compensated \$10/h or extra course credits for their time.

Sixteen participants (four males and 12 females, mean age 20 years) completed Experiment 1.

Materials and equipment

The participants completed the experiment individually in a room with normal interior lighting. Viewing distance was unconstrained but was approximately 57 cm. Stimuli were presented on a 17-in. CRT monitor (1,024 × 768 pixels, 75-Hz vertical refresh rate) controlled by the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB (www.mathworks.com).

Participants searched for a 90°-rotated black T among 11 rotated black Ls (0°, 90°, 180°, or 270°). The offset between the two segments of the Ls was 0.15°. Each item subtended 0.56° × 0.56° and was placed against a solid white circle (diameter 0.8°). The items were positioned at randomly selected locations in an invisible 10 × 10 matrix (20° × 20°).

The search items were displayed on top of one of 128 photos (20° × 20°) of indoor and outdoor natural scenes that had been obtained through an online search. The orientation of the target (left or right rotation) and the orientation of the distractors were randomly selected for each trial. Figure 1 shows a sample display.

Procedure

Each trial started with a white central fixation square (0.5° × 0.5°) for 500 ms, after which an array of one T and 11 Ls was displayed, with three items lying in each quadrant of the screen. Participants were asked to find the T and to report whether the stem was pointing to the left or to the right by pressing an arrow key. They were told to respond as quickly and accurately as possible. The display was presented until a response was made. A correct response was followed by three rising tones that lasted for a total of 300 ms. An incorrect response was followed by a buzz (for 200 ms) and a blank timeout period of 2,000 ms.

Design

The participants completed ten practice trials in which the T and Ls were randomly positioned. Then they were tested in a training phase (20 blocks of 16 trials each) and a testing phase (five blocks of 16 trials each). Finally, explicit memory for the scenes and the scene–target associations was tested.

Training phase Each of the 20 training blocks contained 16 trials. Before the first block, 16 scenes were randomly selected for each participant. Eight of the 16 scenes were assigned to the “scene+array predictive” condition, and the other eight were assigned to the “unpredictive” condition. Each participant was also randomly assigned eight target locations, constrained to two target locations per quadrant. The same target locations were used for the scene+array predictive and unpredictable conditions, equating learning of the targets’ possible locations.

In the scene+array predictive condition, a given target location was associated with the same scene and the same repeated distractor array across 20 repetitions (once per block). In the unpredictable condition, the distractor locations were randomly generated, so the search array did not predict the target’s location. In addition, the pairings between the eight unpredictable scenes and the eight target locations were shuffled across the 20 training blocks, resulting in lack of a consistent association between the scene and the target.

Testing phase To examine whether participants had acquired scene-based contextual cueing, two conditions were tested in the testing phase. Each block contained eight trials in the unpredictable condition, which was the same as in the

training phase (including use of the same eight unpredictable scenes). Additionally, eight “scene-predictive” trials were presented per block. This condition used scenes and target locations from the scene+array predictive training condition, except that the distractor locations were randomized. Consequently, the search array was no longer predictive of the target’s location, but the background scene remained predictive.

Recognition At the completion of the experiment, we tested explicit memory for the repeated search displays. In the *scene-recognition* block, participants were shown the 16 training scenes (including the eight predictive and eight unpredictable scenes) and 16 novel scenes. No search items were shown, and participants were asked to report whether the scene was old or new. For each scene, the computer also assumed a target location (not displayed). For predictive scenes, the assumed target location was the one that had been associated with the scenes previously. For unpredictable scenes and new scenes, the assumed target location was randomly selected from the eight possible target locations. Regardless of their old/new judgment for a given scene, participants were asked to click on the target’s location in the scene. We calculated the Euclidean distance between the mouse click and the assumed target location in units of pixels. No feedback was given in the recognition phase.

Results

In all experiments reported in this study, the overall accuracy for visual search was high (over 97 %) and was unaffected by the experimental manipulations (all p s > .20). We calculated mean RTs for correct trials, excluding trials with RTs longer than 10 s. Figure 2 shows the search data from Experiment 1, separately for the predictive and unpredictable conditions across the 20 training blocks and five testing blocks.

Training phase

We conducted a repeated measures analysis of variance (ANOVA) using Training Condition and Block (1–20) as

factors. This analysis showed a significant contextual-cueing effect in the training phase, as RTs were faster in the scene+array predictive condition than in the unpredictable condition, $F(1, 15) = 49.25$, $p < .001$, $\eta_p^2 = .77$. RTs also improved as training progressed, leading to a significant main effect of block, $F(19, 285) = 10.06$, $p < .001$, $\eta_p^2 = .40$. Furthermore, we found a significant interaction between display condition (scene+array predictive or unpredictable) and block, $F(19, 285) = 3.544$, $p < .001$, $\eta_p^2 = .19$: The two display conditions had comparable RTs at the beginning of the experiment ($p > .50$ in Block 1), but RTs started to diverge after a few repetitions. The first block in which contextual cueing became significant was Block 5, $t(15) = 4.34$, $p < .001$, but cueing was not consistently significant until after Block 9.

Testing phase

Contextual cueing remained significant in the testing phase, even though the search array was unpredictable of the target’s location. An ANOVA on display condition and block (21–25) showed a significant main effect of display condition, $F(1, 15) = 52.67$, $p < .001$, $\eta_p^2 = .78$, which did not interact with testing block, $F(4, 60) = 1.03$, $p > .30$. To compare the sizes of contextual cueing in the testing and training phases, we conducted an ANOVA contrasting the last five training blocks with the five testing blocks. Three factors were entered into the ANOVA: Phase (training or testing), Condition (predictive or unpredictable), and Block (five in each phase). This analysis revealed no interaction between condition and phase, $F < 1$, suggesting that contextual cueing from the scene+array predictive displays fully transferred to the scene-predictive displays.

Recognition

Figure 3 displays the recognition data from Experiment 1. Participants were able to explicitly recognize scenes used in the training phase: Their percentages of “old” responses were significantly influenced by scene type, $F(2, 30) = 290.55$, $p < .001$, $\eta_p^2 = .95$. In addition, participants had explicit memory of the target’s location in the scene: The distance between their selected target location and the actual

Fig. 2 Response time (RT) results from Experiment 1. Error bars show 95 % confidence intervals of the difference between the predictive and unpredictable conditions

Fig. 3 Scene recognition results from Experiment 1. Error bars show 95 % confidence intervals within each condition

one was shorter in the scene+array predictive condition than in either the unpredictable scenes, $t(15) = 6.11$, $p < .001$, or the foil scenes, $t(15) = 6.64$, $p < .001$.

Discussion

In Experiment 1, we trained participants to search for a target among repeated distractor arrays presented against repeated scenes. Search RTs were faster when the target's location was consistently associated with specific background scenes and search arrays than when the pairing was inconsistent. Furthermore, contextual cueing fully transferred to displays with random arrays placed on predictive scenes. Unlike in Brooks et al. (2010), random arrays did not disrupt transfer of cueing to scene-predictive displays; the scene and array were not learned as a joint cue.

The inconsistency between our study and Brooks et al.'s (2010) may be attributable to the extent of the background scenes and the nature of the experimental materials. Brooks et al. generated the search arrays such that they were 3-D structures within a global scene. This integration may have contributed to joint learning of the two types of cues. In contrast, the search items in our study were 2-D images superimposed on the scene, increasing their perceptual segregation. In addition, the scene occupied the entire display in our study, providing both local and global context for the target. In Brooks et al.'s study, a large table on which the array was presented occluded the background scene, leaving just the peripheral region of the scene on the display. This occlusion may have reduced the utility of the scene for contextual cueing. In fact, one of Brooks et al.'s experiments provided evidence that the scene by itself was insufficient to produce contextual cueing. In that study, Brooks et al. trained participants to search from scene-predictive displays with random arrays, and the participants showed no contextual cueing. In contrast, in a preliminary study, we were able to obtain significant contextual cueing when participants were trained on scene-predictive, but array-random, displays. The increased scene utility, and the greater perceptual segregation between the array and the scene, may have been the reason why contextual cueing transferred from scene+array predictive displays to scene-predictive displays in our study.

Experiment 2: Scene+array predictive during training, array predictive during testing

Experiment 1 had shown that the scene and the array were not learned jointly when both were predictive of the target's location. Instead, we found that the scene–target association was learned independently of the array. These findings raise the question of whether the array may also be learned independently of the scene. A crucial difference between the independent-learning and the overshadowing hypotheses lies in the nature of array learning. The independent-learning account predicts that when both scene and array cues are available, both cues are learned. Alternatively, overshadowing maintains that the stronger cue should prevent learning of the weaker cue. Because scene-based contextual cueing on its own is about ten times greater than array-based contextual cueing, the scene cue might overshadow learning of the array cue.

In Experiment 2, we tested the independent-learning and overshadowing hypotheses. Training was the same as in Experiment 1. During testing, we presented old arrays against completely novel scenes. Thus, the scenes were no longer predictive of the target's location. If the array cue was learned, contextual cueing from scene+array predictive displays should transfer, at least partially, to array-predictive displays. Alternatively, if scene-based cueing overshadowed learning of the array–target association, cueing should not transfer to array-predictive testing displays.

Method

Participants Sixteen participants completed Experiment 2, including four males and 12 females, with a mean age of 20 years.

Procedure and design This experiment was modified from Experiment 1. The training phase was identical to that of Experiment 1. In each block, participants were exposed to eight displays in the scene+array predictive condition and eight other displays in the unpredictable condition. Just as in Experiment 1, for the unpredictable condition we used eight scenes and eight target

locations across the 20 training blocks, but without a consistent association between target locations and specific distractor arrays or scenes. In the scene+array predictive trials, the same eight target locations were consistently paired with eight unique scenes, along with an assigned distractor array for each display.

During a subsequent testing phase, we removed the scene predictability but retained the array predictability. The 16 trials of each block contained eight old search arrays learned in the training phase, as well as eight new arrays. Novel background scenes were used in the testing phase, without any repetitions.

Participants then completed a scene recognition block in which they saw 32 new scenes, the 16 scenes presented during the training phase along with the 16 scenes presented during the testing phase. As in Experiment 1, their task was to judge each scene as old or new and to click on the target's likely location.

Finally, participants completed an *array-recognition* block, in which they were shown 16 search arrays. Half of these arrays were the predictive arrays that had been used during training and testing, whereas the other half were newly generated. No scenes were presented in the background. Participants made an “old/new” response to each array. Other aspects of the experiment were similar to those in Experiment 1.

Results

Training phase Figure 4 plots the mean RTs for the predictive and unpredictable conditions as a function of training block. An ANOVA using Condition (scene+array predictive vs. unpredictable) and Block (1–20) as within-subjects factors revealed that RTs were significantly faster in the predictive than in the unpredictable condition, $F(1, 15) = 20.78$, $p < .001$, $\eta_p^2 = .58$. RTs also became faster as training progressed, $F(19, 285) = 11.58$, $p < .001$, $\eta_p^2 = .44$. These two factors showed a significant interaction, $F(19, 285) = 2.55$, $p < .001$, $\eta_p^2 = .15$. Follow-up t tests showed that while the cueing effect was marginally significant during some of the first nine blocks, it was not consistently significant until Block 10, $t(15) = 3.31$, $p < .005$.

Testing phase When old arrays were presented against novel scenes, contextual cueing was disrupted. An ANOVA on testing block (21–25) and condition (array predictive or unpredictable) revealed no effect of condition, $F(1, 15) = 1.40$, $p < .25$, and no interaction between condition and block, $F(4, 60) = 1.74$, $p > .10$. None of the other experimental effects were significant.

To further quantify the lack of a significant cueing effect in the testing phase, we conducted an additional ANOVA contrasting the last five training blocks with the five testing blocks, using Phase, Condition, and Block as factors. This analysis revealed a significant interaction between condition and phase, $F(1, 15) = 28.79$, $p < .001$, $\eta_p^2 = .66$: Whereas contextual cueing was significant in the last five blocks of the training phase, it was disrupted in the testing phase. Repeated arrays, in the absence of previously associated scenes, were insufficient for contextual cueing.

Recognition Participants were highly accurate in recognizing the scenes they had seen in the experiment (Fig. 5). Their percentages of “old” responses were significantly influenced by scene type, $F(3, 45) = 86.48$, $p < .001$, $\eta_p^2 = .85$. In addition, participants had explicit memory of the associated target location on scenes from scene+array predictive trials. The distance between the selected and actual target locations was significantly influenced by scene type, $F(3, 45) = 43.60$, $p < .001$, $\eta_p^2 = .74$. Follow-up tests showed that the distance was shorter for scene+array predictive scenes than for the other types of scenes, $ps < .001$.

In contrast to explicit knowledge about the predictive scenes, participants were at chance in recognizing repeated search arrays. The hit rate for recognizing old arrays was not significantly higher than the false alarm rate on random arrays, $t(15) = 1.57$, $p > .14$.

Discussion

In Experiment 2, we trained participants to search for a target among repeated search arrays presented against repeated scenes. Search RTs were faster when the target's location was consistently associated with specific

Fig. 4 Response time (RT) results from Experiment 2's visual search task. Error bars indicate 95 % confidence intervals of the difference between the scene+array predictive and unpredictable trials

Fig. 5 Recognition results from Experiment 2. Error bars show 95 % confidence intervals within each condition

background scenes and search arrays than when the pairing was inconsistent. If contextual cueing were independently established for the scene–target association and the array–target association, there would have been some transfer to testing trials in which the repeated arrays were displayed against novel scenes. However, we did not find a significant contextual-cueing effect in the testing phase. These data are inconsistent with the independent-learning hypothesis. Although they are consistent with Brooks et al.’s (2010) findings of joint learning, the data from Experiment 1 suggested that the scene and arrays were not learned as a joint cue. Together, the first two experiments support the overshadowing hypothesis, according to which only the stronger cue—the scene—is learned.

Experiments 3A and 3B: High perceptual load

The first two experiments showed that when both the scene and the array are predictive in visual search, scene-based cueing overshadows array-based cueing. However, it is unclear whether this pattern of results is generalizable. In our experiments, the background scene was irrelevant to the search task until its predictability was noticed. The array, on the other hand, was relevant to the task: Participants had to inspect the items in the array to find the target. According to the perceptual-load theory (Lavie, 2005), the degree to which irrelevant information (the scenes) is processed depends on the perceptual load of the primary task (the T/L search). Increasing the perceptual load of the T/L search task should reduce the amount of attention that is available for processing the irrelevant background scene. In fact, a previous study has shown that contextual cueing is modulated by perceptual load (Jiang & Chun, 2001): Ignored items yield a contextual-cueing effect only when the perceptual load of the primary search task is low. One issue that we wanted to address in Experiment 3 was whether the pattern of

overshadowing would change when the perceptual load of the T/L search task increased.

In previous research, perceptual load has been manipulated in several ways, the most common of which has been to increase the number of distractors on the display (Lavie, 1995). However, this manipulation is open to alternative interpretations, since additional items, whether relevant or irrelevant to the task, may dilute the impact of late selection (known as the “dilution theory”; see Lavie & Torralbo, 2010; Tsai & Benoni, 2010; Wilson, Muroi, & MacLeod, 2011). Our study relied on a manipulation of target–distractor similarity (Duncan & Humphreys, 1989), which should affect the demand for perceptual attention without introducing dilution. To this end, we increased the similarity between the T and Ls by increasing the offset of the two segments of the L to nine pixels, rather than five as in Experiment 2. This increased the perceptual load of the task, and thus reduced the amount of attention available for the background scenes. We examined whether this manipulation would change the relative strengths of scene-based and array-based contextual cueing.

In addition to manipulating perceptual load, in Experiment 3 we also tested the possibility that array-based cueing could be acquired, but its expression was disrupted by the presence of new scenes during testing. To this end, we ran two versions of the experiment. Experiment 3A was the same as Experiment 2, except that the perceptual load of the T/L search task was increased. Following training, the predictive arrays were displayed against new scenes that could not have produced scene-based cueing. Experiment 3B was the same as Experiment 3A, except that in the testing phase, the predictive arrays were displayed against a blank screen. If scene-based cueing overshadows array-based cueing, contextual cueing should be eliminated in the testing phase, regardless of whether the predictive array was shown against a blank background or against a new scene.

Method

Participants A total of 32 participants took part in Experiment 3: 16 in Experiment 3A (ten males and six females, mean age 21) and 16 in Experiment 3B (three males and 13 females, mean age 21).

Materials, procedure, and design Experiment 3A was exactly the same as Experiment 2, except that the L distractor stimuli looked more similar to the target T: The offset between the two segments of the L was 0.28° instead of 0.15° . This manipulation increased the perceptual load of the array-based search. In the testing phase, the predictive arrays were displayed against new scenes.

Experiment 3B was the same as Experiment 3A, but background scenes were not shown during the testing phase. Instead, participants only saw the array of one T and 11 Ls against a solid gray background and were asked to find the target. Since no scenes were displayed in this phase, array-based cueing could not have been disrupted by the new scenes. The recognition test in Experiment 3B included 32 scenes (16 from the training phase and 16 novel scenes).

Results

Training phase Figure 6 plots the RT results separately for Experiments 3A and 3B. Because the training phases were identical in the two versions of Experiment 3, we combined

data from all 32 participants and performed an ANOVA using Condition (scene+array predictive vs. unpredictable) and Block (1–20) as within-subjects factors, and Version (3A or 3B) as a between-subjects factor. This analysis revealed no main effect or interaction effects of version, all $ps > .10$. RTs were significantly faster in the scene+array predictive than in the unpredictable condition, $F(1, 30) = 25.08$, $p < .001$, $\eta_p^2 = .46$. RTs also became faster as training progressed, $F(19, 570) = 10.85$, $p < .001$, $\eta_p^2 = .27$. These two factors showed a significant interaction, $F(19, 570) = 4.00$, $p < .001$, $\eta_p^2 = .12$: Contextual cueing was absent in early blocks, but became consistently significant by Block 14, $t(31) = 3.20$, $p < .003$.

Testing phase When old arrays were presented against novel scenes (Exp. 3A) or against a blank background (Exp. 3B), contextual cueing was disrupted. An ANOVA on experimental version (3A or 3B) and condition (array predictive or unpredictable) revealed no effect of condition, $F < 1$, and no interaction between condition and experimental version, $F(1, 30) = 1.09$, $p > .30$. To further quantify the negligible cueing effect in the testing phase, we conducted an additional ANOVA contrasting the last five training blocks with the five testing blocks, using Phase, Block, and Condition as factors. This analysis revealed a significant interaction between condition and phase, $F(1, 31) = 23.41$, $p < .001$, $\eta_p^2 = .43$. This interaction was significant in both Experiments 3A, $F(1, 15) = 15.58$, $p < .001$, $\eta_p^2 = .51$, and 3B, $F(1, 15) = 21.16$, $p < .001$, $\eta_p^2 = .59$. Repeated arrays, in the absence of previously associated scenes, produced negligible contextual cueing.

Fig. 6 Response time (RT) results from **a** Experiment 3A and **b** Experiment 3B. Error bars show 95 % confidence intervals of the difference between the predictive and unpredictable conditions

Recognition Participants were highly accurate in recognizing scenes that they had seen in the experiment (Fig. 7). Their percentages of “old” responses were significantly influenced by scene type: $F(3, 45) = 37.86, p < .001, \eta_p^2 = .72$, for Experiment 3A, and $F(2, 30) = 74.04, p < .001, \eta_p^2 = .83$, for Experiment 3B. In addition, participants had explicit memory of the associated target location on scenes from scene+array predictive trials. The distance between the selected target location and the target’s actual location was significantly influenced by scene type: $F(3, 45) = 22.64, p < .001, \eta_p^2 = .60$, for Experiment 3A, and $F(2, 30) = 9.533, p < .001, \eta_p^2 = .39$, for Experiment 3B. Follow-up tests showed that the distance was shorter for scene+array predictive scenes than for the other types of scenes, $ps < .001$ in Experiment 3A and $ps < .05$ in Experiment 3B.

While array recognition had not been significant in Experiment 2, the array recognition hit rate was significantly higher than the false alarm rate in Experiments 3A and 3B combined, $F(1, 31) = 6.812, p < .01, \eta_p^2 = .18$. To examine whether this recognition accuracy correlated with search performance, we calculated contextual cueing for the last five blocks of the training phase and for the five testing blocks. We also calculated recognition accuracy as the hit rate minus false alarm rate. Pearson correlations showed that recognition accuracy did not significantly correlate with contextual cueing during training, $r = .28, p > .10$, or during testing, $r = -.21, p > .20$. Thus, although increased perceptual load for the T/L search task led to greater recognition accuracy of the predictive arrays, explicit awareness of the array did not contribute to contextual cueing.

Discussion

Experiment 3 generalized the results from Experiment 2. Increasing perceptual load for the T/L search task delayed the onset of contextual cueing in the scene+array displays.

However, it did not change the pattern of learning. Following training with the scene- and array-predictive displays, participants showed no transfer of learning to array-predictive displays. These results were unaffected by whether the predictive arrays were shown against a blank background (Exp. 3B) or against novel scenes (Exp. 3A). Scene-based contextual cueing overshadowed array-based contextual cueing.

Experiment 4: Array-predictive displays with random scenes

Experiment 4 was designed to test a corollary of the overshadowing hypothesis: Specifically, if array-based contextual cueing is overshadowed by scene-based contextual cueing, it should be possible to establish array-based cueing when the background scenes are not predictive of the target location. To this end, we trained participants on displays with repeated arrays placed against unpredictable scenes. Because no consistent scene–target pairing was used during training, the overshadowing hypothesis predicts that we should observe significant contextual cueing for array-predictive displays.

Method

Participants Sixteen college students (five males and 11 females) completed Experiment 4. Their mean age was 20 years. Three additional participants completed the experiment, but their data were not included due to a computer error during the recognition phase.

Materials, procedure, and design This was a standard array-based contextual-cueing experiment, except that unpredictable scenes were presented in the background of predictive search arrays. The distractor L stimuli were the

Fig. 7 Recognition results for Experiments 3A (gray bars) and 3B (white bars). Error bars show 95 % confidence intervals within each condition

Fig. 8 Response time (RT) results from Experiment 4. Error bars show 95 % confidence intervals of the difference between the predictive and unpredictable conditions

same as those used in Experiments 1 and 2, with an offset of 0.15° between the two segments of the L. Participants were tested in 24 training blocks, each comprising 16 trials. In each block, eight trials involved random combinations of eight scenes, eight target locations, and random distractor locations (unpredictive). The other eight trials involved random pairings between eight other scenes and eight target locations, but repeated distractor locations for a given target location (array predictive). The sets of scenes used for the unpredictable and array-predictive trials did not overlap. All scenes were shown once per block, a total of 24 times. The recognition phase involved scene recognition and array recognition, similar to the previous experiments.

Results

Visual search Due to the small number of trials per block, we combined the data from six adjacent blocks to produce four experimental epochs (see also Chun & Jiang, 1998). Figure 8 plots the RT data by epochs. An ANOVA on condition (array predictive vs. unpredictable) and epoch revealed a significant main effect of condition, $F(1, 15) = 8.56$, $p < .01$, $\eta_p^2 = .36$. RTs also improved as the experiment progressed, $F(3, 45) = 22.50$, $p < .001$, $\eta_p^2 = .60$. Critically, these two factors showed a

significant interaction, $F(3, 45) = 3.29$, $p < .029$, $\eta_p^2 = .18$. Follow-up t tests confirmed that contextual cueing was absent in the first two epochs, $t(15)s < 1$, but was significant in Epoch 3, $t(15) = 2.67$, $p < .018$, and Epoch 4, $t(15) = 3.49$, $p < .003$.

Recognition Figure 9 shows the recognition data from Experiment 4. Participants were highly accurate in recognizing scenes used in the experiment, suggesting that they had explicitly processed these scenes, $F(2, 30) = 131.61$, $p < .001$, $\eta_p^2 = .90$. However, as in Experiment 2 and previous studies (e.g., Chun & Jiang, 2003), participants were unable to recognize repeated arrays: $t(15) = -0.79$, $p > .40$, when comparing the hit and false alarm rates during array recognition.

Discussion

Experiment 4 demonstrated that in the absence of a salient scene cue, it was possible to develop contextual cueing for repeated arrays. These data are consistent with Brooks et al. (2010, Exp. 3), which showed that the array could be learned in the absence of a predictive scene. Even though the scenes in our study were spatially extensive, they did not disrupt learning of the array–target association as long as they were unpredictable of the search target. Experiments 1–3 showed that when a salient scene-based cue was available, it overshadowed learning of the array-based cueing. These data together support the overshadowing hypothesis over either the joint-learning or independent-learning hypothesis.

The size of the array-based contextual cueing in the second half of the experiment was about 222 ms in Experiment 4, which was smaller than the size of scene-based contextual cueing (about 650 ms in Exp. 2), but the 10 % improvement in RTs was comparable to results from previous studies on array-based contextual cueing (Chun & Jiang, 1998; Jiang, Song, & Rigas, 2005). In addition, array-based contextual cueing was largely implicit, whereas scene-based contextual cueing led to explicit memory of the scenes and the scene–target associations.

Fig. 9 Recognition data from Experiment 4. Error bars show 95 % confidence intervals within each condition

General discussion

In everyday visual search, targets are often found in the context of distractors and a background scene. Consistent pairing of the target's location and the spatial context of the target leads to enhanced visual search, a phenomenon known as *contextual cueing*. In this study, we investigated the interaction between two sources of contextual cues: explicit, scene-based contextual cueing, and implicit, array-based contextual cueing. We showed that when both the scene and the array were predictive of the target's location, only the scene–target association was acquired, overshadowing the array-based contextual cueing. This occurred even when the perceptual load of the T/L search was high. In the absence of predictive scenes, however, it was possible to observe array-based contextual cueing. These data do not support two alternative hypotheses. First, they suggest that scene- and array-based cueing do not simultaneously and independently occur, but instead that the more salient cue dominates learning. Second, they argue against the idea that the scene and the array are learned jointly.

The finding that overshadowing occurs, rather than joint learning, is inconsistent with Brooks et al.'s (2010) study, in which they examined local and global context learning. Unlike in our study, Brooks et al. found that training with scene+array predictive displays led to joint learning of the two cues, such that cueing was disrupted when either the scene or the array became unpredictable. The two studies differed in the arrangements of their scenes and arrays and in the nature of the experimental materials. Brooks et al. constrained the array to a local context in the center of the screen, which occluded a large portion of the scene. This arrangement might have limited the utility of the scene-based contextual cueing. In contrast, the scene subtended the entire search space in our study, increasing the likelihood that it could be used for cueing. In addition, the 3-D rendering technique used by Brooks et al. increased the integration of the arrays and scenes, which might have contributed to the joint learning observed in that study. Finally, because Brooks et al. were interested in comparing local and global learning, they did not assess participants' explicit awareness of the scenes and arrays. It is possible that in their study, the global scene and local array contexts did not map clearly onto explicit and implicit learning, respectively. Any of these differences might have led to more integrated learning of the scene and the array in Brooks et al.'s study than in ours.

Unlike Brooks et al.'s (2010) study, in which the scene and the array were mapped to global and local contexts, our study was designed to map the scene and the array to explicit learning and implicit learning. The recognition data suggested that our experimental manipulation was effective in inducing explicit, scene-based learning and implicit,

array-based learning. Several previous studies have also endorsed the idea that array-based cueing is implicit in nature (e.g., Chaumon, Drouet, & Tallon-Baudry, 2008; Chaumon, Schwartz, & Tallon-Baudry, 2009; Chun & Jiang, 2003).

The present study has significant theoretical implications for the interaction between attentional cueing by explicit and implicit learning. Explicit cueing may be considered a form of endogenous cueing, in which participants intentionally use the cue (background scene) to allocate spatial attention. This form of cueing may affect spatial attention by modulating the priority map (Fecteau & Munoz, 2006; Wolfe, 2007), increasing priority weights for the associated target location. In contrast, implicit cueing lacks the intentional component of top-down control. Several studies have demonstrated characteristic differences in attentional guidance by implicit cueing and endogenous cueing. For example, Kunar et al. (2006) suggested that once participants become aware of the association between a background image and the target's location, attentional guidance becomes more effective. Jiang et al. (2013) showed that when the target's location probability was cued by both a central arrow and implicit learning of the target's likely locations, endogenous cueing dominated performance. The present study adds to the growing literature by showing that spatial attention is more effectively guided by an explicit, endogenous cue than by an implicit cue. In fact, the concurrent presence of explicit and implicit cues overshadowed learning of the implicit cue.

To conclude, by training participants with both predictive scenes and predictive arrays, we have demonstrated overshadowing in contextual cueing. In our study, scene-based contextual cueing overshadowed array-based cueing. Our data are inconsistent with the idea that explicit and implicit contextual cueing occur simultaneously and independently. They also suggest that the two types of cues may not be acquired in conjunction to form a joint cue. Our study supports the growing literature on the dominance of endogenous cueing over implicit cueing.

Author note We thank Alex Fleming, Andrew Mekhail, Heather Sigstad, Liwei Sun, and Josh Tisssell for help with data collection, and Khena Swallow for comments and suggestions.

References

- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 798–815. doi:10.1037/0096-1523.33.4.798
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. doi:10.1163/156856897X00357

- Brockmole, J. R., Castelano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 699–706. doi:10.1037/0278-7393.32.4.699
- Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, 59, 1177–1187. doi:10.1080/17470210600665996
- Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues during search. *Visual Cognition*, 13, 99–108. doi:10.1080/13506280500165188
- Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1406–1418. doi:10.1037/a0019257
- Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision*, 8(3), 10, 1–10. doi:10.1167/8.3.10
- Chaumon, M., Schwartz, D., & Tallon-Baudry, C. (2009). Unconscious learning versus visual perception: Dissociable roles for gamma oscillations revealed in MEG. *Journal of Cognitive Neuroscience*, 21, 2287–2299. doi:10.1162/jocn.2008.21155
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170–178. doi:10.1016/S1364-6613(00)01476-5
- Chun, M. M., & Jiang, Y. V. (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. M., & Jiang, Y. V. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 224–234. doi:10.1037/0278-7393.29.2.224
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458. doi:10.1037/0033-295X.96.3.433
- Ehinger, K. A., & Brockmole, J. R. (2008). The role of color in visual search in real-world scenes: Evidence from contextual cuing. *Perception & Psychophysics*, 70, 1366–1378. doi:10.3758/PP.70.7.1366
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, 10, 382–390. doi:10.1016/j.tics.2006.06.011
- Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology*, 54A, 1105–1124. doi:10.1080/02724980042000516
- Jiang, Y., Song, J.-H., & Rigas, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review*, 12, 524–529. doi:10.3758/BF03193799
- Jiang, Y. V., Swallow, K. M., & Rosenbaum, G. M. (2013). Guidance of spatial attention by incidental learning and endogenous cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 285–297. doi:10.1037/a0028022
- Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2006). Contextual cuing by global features. *Perception & Psychophysics*, 68, 1204–1216. doi:10.3758/BF03193721
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451–468. doi:10.1037/0096-1523.21.3.451
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, 9, 75–82. doi:10.1016/j.tics.2004.12.004
- Lavie, N., & Torralba, A. (2010). Dilution: A theoretical burden or just load? A reply to Tsai and Benoni (2010). *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1657–1664. doi:10.1037/a0020733
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11, 520–527. doi:10.1016/j.tics.2007.09.009
- Olson, I. R., & Chun, M. M. (2001). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, 9, 273–302. doi:10.1080/13506280042000162
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. doi:10.1163/156856897X00366
- Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past*. New York, NY: Basic Books.
- Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision*, 53, 169–191.
- Tsai, Y., & Benoni, H. (2010). Diluting the burden of load: Perceptual load effects are simply dilution effects. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1645–1656. doi:10.1037/a0018172
- Wilson, D. E., Muroi, M., & MacLeod, C. M. (2011). Dilution, not load, affects distractor processing. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 319–335. doi:10.1037/a0021433
- Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York, NY: Oxford University Press.