

Contextual cueing in 3D visual search depends on representations in planar-, not depth-defined space

Xuelian Zang

China Centre for Special Economic Zone Research,
College of Psychology and Sociology,
Shenzhen University, China
Experimental Psychology, Department of Psychology,
Ludwig-Maximilian University, Munich, Germany



Zhuanghua Shi

Experimental Psychology, Department of Psychology,
Ludwig-Maximilian University, Munich, Germany



Hermann J. Müller

Experimental Psychology, Department of Psychology,
Ludwig-Maximilian University, Munich, Germany
Department of Psychological Sciences, Birkbeck College,
University of London, London, UK



Markus Conci

Experimental Psychology, Department of Psychology,
Ludwig-Maximilian University, Munich, Germany



Learning of spatial inter-item associations can speed up visual search in everyday life, an effect referred to as contextual cueing (Chun & Jiang, 1998). Whereas previous studies investigated contextual cueing primarily using 2D layouts, the current study examined how 3D depth influences contextual learning in visual search. In two experiments, the search items were presented evenly distributed across front and back planes in an initial training session. In the subsequent test session, the search items were either swapped between the front and back planes (Experiment 1) or between the left and right halves (Experiment 2) of the displays. The results showed that repeated spatial contexts were learned efficiently under 3D viewing conditions, facilitating search in the training sessions, in both experiments. Importantly, contextual cueing remained robust and virtually unaffected following the swap of depth planes in Experiment 1, but it was substantially reduced (to nonsignificant levels) following the left–right side swap in Experiment 2. This result pattern indicates that spatial, but not depth, inter-item variations limit effective contextual guidance. Restated, contextual cueing (even under 3D viewing conditions) is primarily based on 2D inter-item associations, while depth-defined spatial regularities are probably not encoded during contextual learning. Hence, changing the depth relations does not impact the cueing effect.

Introduction

In order to deal with the complex and frequently changing environment surrounding us, we have developed the sophisticated ability to register regularities and learn contextual associations among objects in scenes, facilitating performance in everyday search tasks. In the laboratory, the mechanisms underlying context learning are typically investigated using the contextual cueing paradigm (Chun & Jiang, 1998). In the standard variant of this task, participants are instructed to search for a T-shaped target among a number of L-shaped distractors. Half of the search displays presented over the course of the experiments are *old*, repeated contextual layouts that maintain the spatial relation of the target to the other distractor items (with each old search display repeating once per block), while the other half of the displays are *new*, with distractor locations (relative to a given target location) changing randomly across trials. The main finding is that participants' responses to the target are faster for old, as compared to new, target–distractor configurations, indicating that contextual regularities are learned and come to guide, or cue, visual search.

Several studies have revealed contextual cueing to be a rather robust phenomenon (for review, see Goujon, Didierjean, & Thorpe, 2015), based on the stable

Citation: 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, doi:10.1167/17.5.17.

doi: 10.1167/17.5.17

Received January 19, 2017; published May 26, 2017

ISSN 1534-7362 Copyright 2017 The Authors



learning of contextual regularities. For instance, contextual facilitation has been found to be maintained even when the entire search display was rescaled (Jiang & Wagner, 2004), when the identity of the search items was exchanged (Chun & Jiang, 1999), or with alterations of the global display layout (Shi, Zang, Jia, Geyer, & Müller, 2013; Zang et al., 2016). For instance, Shi et al. (2013) observed reliable contextual cueing even though a global geometric structure of the display changed.

It is commonly agreed that what is learnt in contextual cueing are spatial associations of (the arrangement of) distractors to the target location (Brady & Chun, 2007; Conci & Müller, 2012), which, when automatically retrieved in the response to an old display, provide attentional guidance to a target location. This is a core assumption of Brady and Chun's (2007) connectionist model of contextual cueing, where spatial associations of distractors with the target within the narrower target surround play a dominant role for effective contextual learning. In support of this view, Brady and Chun (2007) showed contextual cueing to be reliable even when only two distractors in the vicinity of the target remained constant, while changes of distractors at more distant locations had little effect on contextual cueing. Similar observations were also reported by Olson and Chun (2002) and Song and Jiang (2005).

In addition to the acquisition of learned target–distractor associations, Beesley, Vadillo, Pearson, and Shanks (2015) showed that constant spatial associations among distractors alone suffice to facilitate visual search to some extent (for comparable findings, see also Schankin & Schubö, 2009a; Schankin & Schubö, 2009b). In general agreement with the crucial role of target–distractor associations in contextual cueing, changes to previously learned associations in a scene usually lead to contextual costs. For example, Conci, Sun, and Müller (2011) found that contextual facilitation was abolished when the target location changed to a previous nontarget location, while relearning of the new target–distractor associations following the change could only be incorporated into contextual memory after a substantial amount of training: more than 80 repetitions were required for contextual facilitation for the relocated target to become manifest (Zellin, Von Mühlennen, Müller, & Conci, 2014). Taken together, these findings support the idea that contextual cueing depends on acquired inter-item associations between the context and the target, whereas changes of the target–distractor associations yield substantial and long-lasting reductions in contextual cueing.

While all of the above-mentioned studies examined item associations in contextual cueing in two-dimensional (2D) visual search, a few other studies also examined three-dimensional (3D) layouts (Chua &

Chun, 2003; Kawahara, 2003). For instance, Chua and Chun (2003) investigated contextual cueing using pseudo-3D scenes (i.e., 2D projections of 3D displays), with the search items inducing apparent-depth information throughout the display. They first trained participants with displays presented at varying viewing angles (0°, 15°, 30°, or 45°) and thereafter examined whether the cueing transferred across depth rotations by presenting all displays at a viewing angle of 0°. Contextual facilitation was found to be systematically decreased with an increasing change in the viewing angle from training to test. Chua and Chun (2003) took this viewpoint dependence of contextual cueing in pseudo-3D search displays to suggest that depth information was incorporated into the underlying contextual memory representations. Note, however, that viewpoint changes in Chua and Chun's (2003) study not only altered the apparent-depth information of the visual search items, but the rotation of the layout also changed the item associations of the 2D projection of the display. Given this, it is not clear whether the reduced cueing effect from training to test occurred due to changes in apparent depth or because of changes of the contextual associations in the display brought about by the rotation of the search layout.

Kawahara (2003) investigated contextual cueing in 3D layouts with the search items presented on two different depth planes: a front and a back plane. Contextual cueing was decreased when the binocular disparity of the distractors in the search display was reversed; that is, when distractors on the back plane moved to the front and distractors on the front plane moved to the back, while the target remained unchanged. These findings led Kawahara to conclude that 3D structure of the layout (including depth information) is encoded in contextual memory. However, these findings potentially again involve a confound: changing only the binocular disparity of the distractors but not of the target (while generating depth variations) would have disrupted previously acquired target–distractor associations. Such associations between the target and the local context of nontarget items (i.e., within a given depth plane) are thought to be a crucial factor for contextual cueing to be maintained (see above). Hence, arguably, the role of depth information in contextual cueing is not clear.

To summarize, despite numerous observations of contextual cueing in 2D visual search, only few studies have explored contextual cueing in 3D layouts. However, these studies not only manipulated 3D aspects of the search display, but also disrupted (or at least weakened) 'invariant' inter-item associations—which are thought to be crucial for contextual cueing. It is, thus, not clear whether contextual cueing is based on 3D memory representations, or whether it primarily depends on 2D projections of spatial target–distractor

relations. On these grounds, we carried out two experiments to examine the type of representation underlying spatial contextual memory across three-dimensional space.

Experiment 1

Experiment 1 was designed to examine whether the learning of a spatial layout in contextual cueing depends on depth information. To this end, 3D search displays were presented with all items distributed across two depth layers: one in the front and one in the back. The experiment consisted of two phases. During an initial training phase, observers learned to associate a given target in an old (repeated) display with the invariant spatial context presented at fixed 3D coordinates for each repeated display (for examples, see Figure 1). In a subsequent transfer phase, the front and back planes were exchanged to test whether contextual learning is sensitive to changes in binocular disparity. Importantly, each item location in the front and back layers was determined so as not to overlap with other items. As a result, the 2D projection (i.e., the 2D retinal image) of the 3D display presented a rather constant pattern of inter-item associations both before and after swaps of the depth planes. Thus, if contextual cueing is indeed sensitive to depth relations (rather than being purely based on a constant pattern of 2D inter-item associations), then a change in binocular disparity should considerably reduce the cueing effect.

Method

Participants

A total of 16 participants (eight women, eight men; mean age: 29.12 ± 4.47 years) with normal or corrected-to-normal visual acuity took part in the experiment, and were paid for their participation. To ensure that all participants were able to perceive depth information, the experimenter asked the participants prior to the experiment about their ability to perceive stereoscopic visual images. In addition, participants were asked at the beginning and after the practice session whether they had clearly seen the 3D display layouts. The formal experiment would continue only when the participant reported that they had seen the 3D structure in the displays. None of the participants were aware of the purpose of the study. The experiment was approved by the ethics committee of the Ludwig-Maximilian University Munich, Department of Psychology, and participants gave informed consent prior to the experiment.

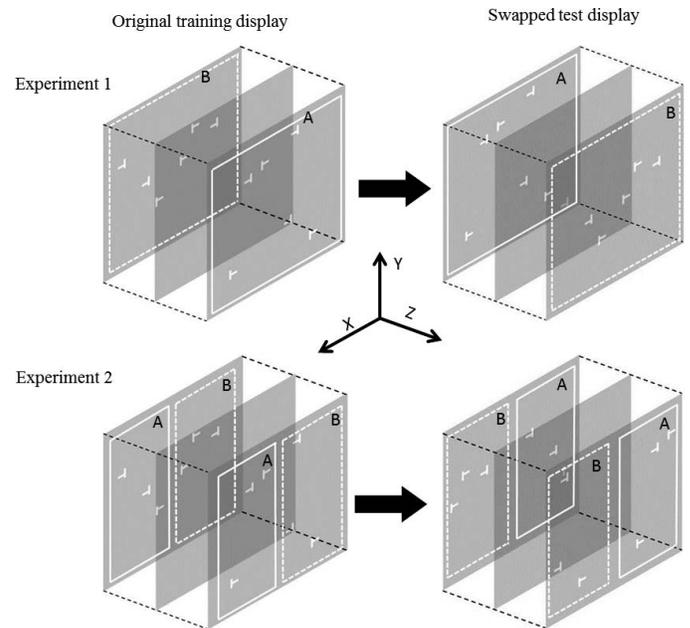


Figure 1. Examples of “original old” and “swapped old” search displays, as used in the experiments. The panels on the left depict the original old displays presented during the training session, and the panels on the right show swapped displays as used in the test session. The spatial context swapped between the front and back layers (i.e., along the Z-axis in Experiment 1; upper row), and between the left and right display halves (i.e., along the X-axis in Experiment 2 (lower row)). Note that the swap of the “new” displays (not shown in the figures) followed the same rule except that distractors in a given display were randomly generated for each presentation. Thus, the target’s location in the new displays swapped between the front and back layers or, respectively, the left and right halves from the training to the test session, while the distractor locations were selected randomly on each trial. The white rectangles and the letters A and B are used here to depict the swapping logic; they were not visible during the actual experiments.

Apparatus and stimuli

The experiment was conducted in a dark cabin (ambient luminance: 0.12 cd/m^2). The search display was presented via a 3D-compatible Optoma projector (HD131Xe; Optoma USA, Fremont, CA) on a white canvas at a refresh rate of 120 Hz. During the experiment, the participant’s viewing distance to the canvas was fixed at 79 cm, with the support of a chin rest. Participants wore a pair of 3D shutter glasses (Optoma ZF2100; Optoma USA), with display frames presented at a rate of 60 Hz to each eye and with alternations of the left- and right-eye shutters (opening and closing) during the presentation. Stimulus presentation and response recording were controlled by Matlab (Mathworks, Natick, MA) using Psychtoolbox extensions (Brainard, 1997; Pelli, 1997).

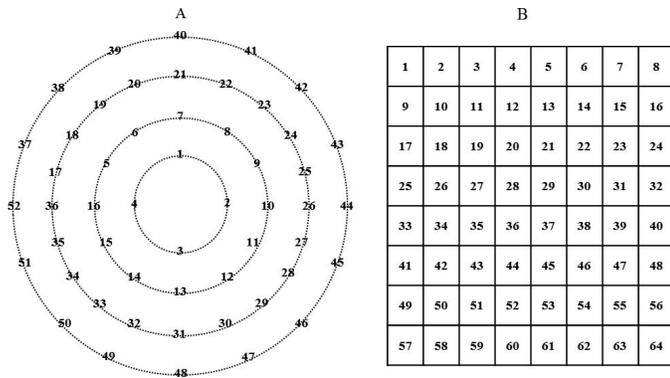


Figure 2. Schematic illustrations of possible item locations. (A) Items were randomly positioned on four concentric circles in Experiment 1 and on a square area in Experiment 2.

Visual stimuli (see Figure 1) were presented to the participants via 3D glasses on a cuboid area with a frontal square subtending $24.5 \text{ cm} \times 24.5 \text{ cm}$ (x - y dimensions) and a depth of 14.7 cm (z dimension), displaying three semi-transparent—front, middle, and back—layers; i.e., effectively, participants viewed the middle and back layers through the front layer. The middle layer presented a fixation cross at the beginning of a trial to equate the distance from central fixation to both front and back layers, which contained the search display items. The depth between adjacent layers was 30% of the edge size of the square area (7.35 cm), so that the viewing distance to the three layers amounted to 71.7 cm , 79.0 cm , and 86.4 cm , respectively, corresponding to a size of the front layer of about $19.4^\circ \times 19.4^\circ$ of visual angle.

At the start of a trial, a fixation cross ($1.2 \text{ cm} \times 1.2 \text{ cm}$) appeared at center of the search display in middle depth layer. Next, the search display was presented, displaying an empty middle layer and six search items in the front and six in the back planes, respectively (one T-shaped target and 11 L-shaped distractors). In both front and back layers, the items were positioned randomly at six of overall 52 possible locations, arranged on four invisible, concentric circles with a diameter of 4.88 cm , 8.71 cm , 12.54 cm , and 16.37 cm , respectively (see Figure 2). The target only appeared at locations on the second and third (invisible) circle. Since the search items were distributed across both the front and back layers, item positioning in the two layers was controlled to prevent occlusions. For instance, if an item was presented at position X of the 52 possible locations in the front layer, then no item would be presented at position X in the back layer. In addition, the size of the items was chosen to be relatively small (around $0.48^\circ \times 0.48^\circ$ of visual angle) compared to the whole visual area, in order to distribute the items rather sparsely across the display. With these controls, the items on the two layers were not overlapping with each

other. The visual search display was followed by an inter-trial interval during which three (i.e., front, middle, and back) blank semitransparent layers were presented to keep participants in a 3D-viewing mode throughout the entire experiment.

Similar to our previous study (Zang, Jia, Müller, & Shi, 2015), both T-shaped and L-shaped stimuli were composed of two equal-length lines—one with horizontal and one with vertical orientation. For the T-shaped stimulus, the lines contact point was at the tip of the vertical line and, respectively, the center point of the horizontal line. The contact point of the L-shaped stimulus was at the tip of the vertical line and at the left side of the horizontal line with an offset of 0.1 cm ; this made the L distractors look more like the target T, increasing the difficulty of the search task. The T-shaped target could be tilted by either 90° or 270° , while the L-shaped distractors could be presented at orientations of 0° , 90° , 180° , or 270° (see Figure 1 for examples).

Design and procedure

Experiment 1 consisted of three sessions: a training session of 20 blocks, a test session of another 20 blocks, and a final recognition session consisting of one block. Each block contained 24 trials with 12 old and 12 new configurations, presented in random order within each block. The depth ordering of items in each configuration was reversed from training to test (i.e., swapping front and back planes). The final recognition session then presented all items as shown in the initial training session.

For the old displays, the positions and orientations of the distractor Ls and the position of the target T remained constant. To avoid learning of target identities (and to conversely trigger spatial learning of the target position), the orientation of the target T varied randomly on each trial. For the new displays, the positions and orientations of the distractors varied randomly on each presentation. The target appeared at one of 12 predefined locations in new displays to equate target location repetitions between old and new displays. After the training session, in the test session, the configurations in the front and back layers were swapped (with a swapping distance of 14.7 cm) for both old and new displays, thus forming swapped-old and swapped-new configurations during test. That is, the search items in the front layer during training were swapped to the back layer during the test session and vice versa (see Figure 1). Note that, essentially, the same swapping logic was also applied for the new displays; thus, the binocular disparity of the target was also swapped during the test session while all distractor items in the new displays varied randomly on each trial, with the constraint that a given display layout presented an equal number of items in the front and the back layers.

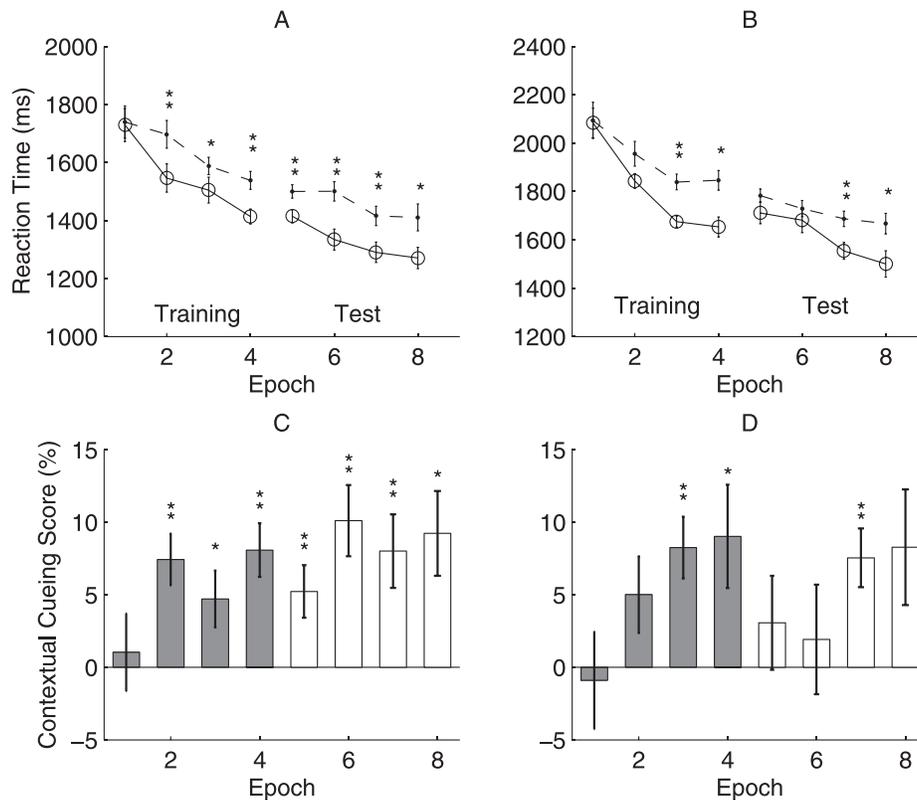


Figure 3. Panels A and B depict the mean reaction times (RTs) with associated standard errors as a function of epoch in Experiments 1 and 2, respectively. Solid lines depict the mean RTs for old context displays, whereas the dashed lines denote new contexts. Panels C and D depict corresponding normalized contextual cueing scores as a function of epoch in Experiments 1 and 2, respectively. Epochs 1–4 represent the training session, Epochs 5–8 represent the test session. * $p < 0.05$, ** $p < 0.01$.

On each trial of the training and test sessions, participants were asked to search for the target letter T and to respond to the orientation (left or right) of the target as fast and accurately as possible by pressing the left or right arrow key on the keyboard using their index fingers. A trial started with the presentation of a fixation cross, which participants were instructed to fixate. The fixation cross disappeared after a random interval of 0.8–1.0 s and was followed by the search display until a response was made, or else it disappeared automatically after 20 s. After a random period of 1.0–1.2 s break, the next trial started. At the end of each block, feedback was provided (presenting the mean correct responses per block), and participants could take a short break.

Following the training and test sessions, the recognition session presented the initial 12 old displays from the training session and another 12 newly generated configurations. Participants were asked to determine whether a given display was old or new by pressing the left or right arrow key, respectively. Displays were presented on the screen until a response was made or else for a maximum of 40 s. Participants were informed that about half of the displays were repeated and the other half were newly generated configurations. Response feedback was not provided.

Prior to the experiment, participants were given a practice session of one block of 24 search trials with random item layouts to become familiar with the task and to get accustomed to the 3D visual display. No layouts presented during the practice session were reused in the subsequent experiment. Observers were asked to aim for (and maintain) a performance level of at least 85% correct responses before they started the experiment proper. If the error rate was too high, participants were given an extra practice block.

To ensure that all participants were able to perceive depth in the presented search displays, the experimenter verbally asked participants at the beginning and after the practice session whether they had clearly seen the 3D display with the search items positioned in the front and back planes. The formal experiment would continue only when the participant reported that they had seen the 3D displays.

Results

Participants' overall mean response times (RTs) are depicted in Figure 3A. To increase statistical power, each five consecutive blocks were grouped into one

epoch, with Epochs 1–4 corresponding to the training session and Epochs 5–8 corresponding to the test session. In each block, the first two trials were discarded from the analysis to ensure that observers were reaccustomed to the 3D layouts after the short breaks between blocks. Trials with erroneous responses or RTs that were faster than 200 ms or slower than 2.5 standard deviations from the individual mean were also excluded from further analysis. The overall mean error rate was only 1.82%, indicating that participants were generally performing highly accurately during the experiment. The overall mean error rates were subjected to a repeated-measures analysis of variance (ANOVA) with the factors epoch (1–8) and context (old or new), which revealed no significant effects: context, $F(1, 15) = 0.27$, $p = 0.61$, $\eta_p^2 = 0.02$; epoch, $F(7, 105) = 0.58$, $p = 0.77$, $\eta_p^2 = 0.04$; interaction, $F(7, 105) = 1.03$, $p = 0.42$, $\eta_p^2 = 0.06$.

Training session

A two-way repeated-measures ANOVA, with context (old vs. new) and epoch (1–4) as factors, was computed for the mean RTs. The results revealed significant main effects of epoch and context: epoch, $F(1.96, 29.37) = 8.38$, $p < 0.001$, $\eta_p^2 = 0.36$; context, $F(1, 15) = 17.68$, $p < 0.001$, $\eta_p^2 = 0.54$; RTs were 258 ms faster in Epoch 4 as compared to Epoch 1, and response latencies were overall shorter, by 92 ms, for old as compared to new displays. The interaction was marginally significant, $F(3, 45) = 2.32$, $p = 0.09$, $\eta_p^2 = 0.13$, mainly due to the cueing effect being nonsignificant in Epoch 1, $t(15) = 0.19$, $p = 0.85$, but significant from Epoch 2 onward: Epoch 2, $t(15) = 3.73$, $p = 0.002$; Epoch 3, $t(15) = 2.16$, $p = 0.047$; and Epoch 4, $t(15) = 3.90$, $p = 0.001$. To further investigate contextual cueing independently of changes in the overall RT latencies (see main effect of epoch), we calculated normalized contextual cueing scores for each epoch (see Figure 3C) with RT in new displays serving as the baseline (see e.g., Howard, Jr., Dennis, Howard, Yankovich, & Vaidya, 2004): $[RT(\text{new}) - RT(\text{old})] / RT(\text{new})$. A series of t tests comparing these normalized scores against zero revealed a significant cueing effect from Epoch 2 onward: Epoch 2, $t(15) = 4.19$, $p = 0.001$; Epoch 3, $t(15) = 2.41$, $p = 0.029$; and Epoch 4, $t(15) = 4.37$, $p = 0.001$. To summarize, there was a robust contextual-cueing effect in the training session of the experiment; that is, participants learned the repeated 3D spatial contexts, facilitating their search.

Test session

A two-way repeated-measures ANOVA with context and epoch (5–8) as factors revealed a significant contextual-cueing effect and a procedural learning effect as indicated by significant main effects of context,

$F(1, 15) = 16.42$, $p < 0.001$, $\eta_p^2 = 0.52$, and epoch, $F(1.70, 25.55) = 6.54$, $p < 0.001$, $\eta_p^2 = 0.30$. Old contexts were responded to faster, by 129 ms, than new contexts, and RTs decreased by 117 ms across epochs. The latter effect shows that overall procedural learning of the task continued after the change in depth. The Context \times Epoch interaction was not significant, $F(3, 45) = 1.54$, $p = 0.22$, $\eta_p^2 = 0.09$, which suggests the significant contextual-cueing effect maintained across epochs. This was supported by further paired sample t tests for individual epochs (see Figure 3A, all $ps < 0.012$), and by additional comparisons of the normalized contextual cueing scores in each epoch against zero (see Figure 3C, all $ps \leq 0.0011$).

Finally, the mean RTs in Epochs 4 and 5 were subjected to a two-way repeated-measures ANOVA with the factors epoch and context, to investigate potential changes in contextual cueing before and after the swap of depth planes. The results again revealed a significant main effect of context, $F(1, 15) = 16.64$, $p < 0.001$, $\eta_p^2 = 0.53$, but no main effect of epoch and no Context \times Epoch interaction: $F(1, 15) = 0.69$, $p = 0.42$, $\eta_p^2 = 0.04$, and $F(1, 15) = 2.06$, $p = 0.17$, $\eta_p^2 = 0.12$, respectively. This indicates that the contextual memory acquired during training is preserved across the swap of the front and back layers; that is, the depth change does not significantly affect the learned associations underlying contextual cueing.

Recognition session

In the recognition test, participants' mean hit rate was $54.69\% \pm 16.38\%$ and mean false alarm rate was $54.69\% \pm 10.96\%$, providing no evidence of explicit context memory in the current experiment.

Discussion

Experiment 1 revealed that repeated 3D spatial context with items positioned across different depth layers is learned effectively, demonstrating that depth-based segmentation of the search display does not, to any significant extent, impede contextual learning (for a comparable finding with display layouts that were segmented by means of color, see Conci & Von Mühlennen, 2011). Thus, in general, this finding shows that 3D contextual cueing reveals a comparable effect (and a comparable effect size) as standard 2D search procedures. Importantly, despite of a clear, depth-defined segmentation of the display into separate layers, the substantial contextual facilitation was maintained after the swapping of contextual configuration between the front and back layers (cueing effects of 92 and 129 ms before and after the swapping). Note that this finding contrasts with Kawahara (2003), who

reported contextual cueing to be reduced when the disparity of distractors but not that of the target was reversed.¹ This difference in the pattern of cueing effects suggests that the target–distractor relations within a certain depth plane are crucial for contextual cueing in 3D visual search. There are two possible explanations for this robust transfer effect: On the one hand, contextual cueing may be relatively “flexible” to depth changes; that is, depth information may not at all, or only to a limited extent, be integrated into a given contextual memory representation—perhaps with depth providing additional but redundant information to contextual cueing. In this view, contextual memory representations would essentially encode 2D inter-item relations. Therefore, contextual facilitation acquired during training is transferred after the swap of the front and back layers. Alternatively, instead of being flexible particularly (or just) to depth changes, contextual cueing might exhibit flexibility to a number of possible spatial changes, such as a swapping between the two halves of the display. If this were the case, reliable contextual facilitation should be maintained after a swap of the context between the left and right halves of the display. This alternative was examined in Experiment 2.

Experiment 2

The experimental design and procedure of Experiment 2 were essentially the same as in Experiment 1, except for the following: instead of swapping the front and back display layers (as in Experiment 1), the spatial context was swapped between left and right display halves (see Figure 1 for an example). In other words, the display items presented originally (during training) in the left half of the display changed (during test) to the right half, and vice versa for the items originally in the right half. Thus, the spatial context swapped along the *X*-axis in Experiment 2, as compared to the swapping along the *Z*-axis in Experiment 1 (see Figure 1). In addition, instead of presenting all search items on a circular grid, in Experiment 2, we used a rectangular arrangement of 64 possible item locations (see Figure 2B). Within this grid, targets could appear at all item locations except for the four center locations around the initial fixation point. The reason to change to a rectangular layout was to maintain the global structure before and after the swapping of the left/right halves of the configuration, which would not have been possible with a circular arrangement: the latter arrangement would have turned into a display with two half-circles facing each other subsequent to the left/right swap (i.e., a ● layout would change into a ◐ layout), thus presenting rather dissimilar displays in the learning and

test sessions. The edge length of the rectangle area used in Experiment 2 was the same as the diameter of the largest circle used in Experiment 1 (16.37 cm), and the variation in (*X*-axis) swapping distance was about 8.2 cm, which was somewhat shorter than the (*Z*-axis) swapping distance in Experiment 1 (14.7 cm). If the acquired contextual associations can be transferred from the originally presented (old) to the swapped displays, then contextual cueing would appear to be rather flexible in compensating for absolute positional variations overall. In contrast, if no transfer were observed, then contextual cueing would appear to be flexible in particular to variations in depth (Experiment 1), but not to left/right swaps. In Experiment 2, we tested 16 participants (10 women, six men; mean age: 26.56 ± 4.35 years old) with normal or corrected-to-normal visual acuity.

Results

As in Experiment 1, the first two trials in each block and trials with erroneous responses or RTs that were faster than 200 ms or slower than 2.5 standard deviations from the individual mean were excluded from further analysis. The overall mean error rates were low (1.09%) and a repeated-measures ANOVA showed no significant differences across epoch (1–8) and context (all p s > 0.5).

Training session

A two-way repeated-measures ANOVA, with the factors context and epoch revealed both main effects to be significant: epoch, $F(1.34, 20.38) = 19.10$, $p < 0.001$, $\eta_p^2 = 0.56$; context, $F(1, 15) = 6.40$, $p = 0.01$, $\eta_p^2 = 0.30$. On average, RTs were 339 ms faster in Epoch 4 than in Epoch 1, and 120 ms faster for the old compared to the new displays. The Context \times Epoch interaction was also significant, $F(3, 45) = 4.89$, $p = 0.007$, $\eta_p^2 = 0.23$, due to a nonsignificant difference between old and new contexts in Epoch 1, $t(15) = 0.18$, $p = 0.86$, followed by a marginal effect in Epoch 2, $t(15) = 1.90$, $p = 0.08$, and reliable differences in Epochs 3, $t(15) = 3.99$, $p = 0.001$, and 4, $t(15) = 2.92$, $p = 0.011$. Subsequent analyses of normalized contextual cueing scores also revealed a significant contextual facilitation in Epochs 3, $t(15) = 3.87$, $p = 0.001$, and 4, $t(15) = 2.53$, $p = 0.023$, but only a marginal effect in Epoch 2, $t(15) = 1.89$, $p = 0.078$, and no effect in Epoch 1, $t(15) = -0.27$, $p = 0.79$ (see Figure 3D). Taken together, these results suggest that contextual cueing developed as the experiment progressed and became stable from Epoch 3 onward. Overall, the normalized contextual cueing scores were roughly comparable in size during the training session in both experiments (5.32 and 5.35 in Experiments 1 and 2,

respectively), thus showing that variations in the display arrangement (circular vs. rectangular) did not have any major influence on contextual learning.

Test session

During the test session, a significant procedural learning effect was found, as evidenced by the main effect of epoch, $F(3, 45) = 11.07$, $p < 0.001$, $\eta_p^2 = 0.43$, which revealed a significant RT reduction of 164 ms from Epoch 5 to Epoch 8. The main effect of context was only marginally significant, $F(1, 15) = 4.19$, $p = 0.059$, $\eta_p^2 = 0.22$, depicting a mean contextual-cueing effect of 105 ms. The interaction was also marginally significant, $F(3, 45) = 2.25$, $p = 0.095$, $\eta_p^2 = 0.13$, with old contexts being responded to faster only toward the end of the test session in Epochs 7, $t(15) = 3.52$, $p = 0.003$, and 8, $t(15) = 2.36$, $p = 0.032$. Simple one-sample t tests performed on the normalized contextual-cueing scores also revealed (marginally) significant above-zero contextual facilitation in Epochs 7, $t(15) = 3.72$, $p = 0.002$, and 8, $t(15) = 2.08$, $p = 0.055$, but not in Epochs 5, $t(15) = 0.95$, $p = 0.36$, and 6, $t(15) = 0.51$, $p = 0.62$. This pattern is indicative of a transient reduction (following the swap) and a subsequent recovery of contextual cueing.

Finally, to test the effect of the swapping on contextual cueing across training and test sessions, a two-way repeated-measures ANOVA of mean RTs was performed with epoch (4, 5) and context as factors. This analysis revealed no main effect of epoch, $F(1, 15) = 0.005$, $p = 0.94$, $\eta_p^2 < 0.01$, but a significant main effect of context, $F(1, 15) = 5.16$, $p = 0.038$, $\eta_p^2 = 0.26$, and a Context \times Epoch interaction, $F(1, 15) = 5.55$, $p = 0.033$, $\eta_p^2 = 0.27$. This effect pattern reflects the fact that contextual cueing decreased significantly from the last training epoch, 193 ms in Epoch 4, $t(15) = 2.92$, $p = 0.011$, to the first test epoch, 72 ms in Epoch 5, $t(15) = 1.18$, $p = 0.26$. In this view, contextual cueing as acquired during the training session did not (effectively) transfer to the test session when the spatial context swapped along the X -axis. However, relearning of the swapped spatial context was nevertheless possible, with the time required for relearning being comparable to the initial contextual learning in the training session: effective relearning required some two epochs of repeated exposure to swapped contextual layouts (i.e., the same number of epochs as for the original learning).

Recognition session

Participants' mean hit rate ($47.92\% \pm 20.53\%$) was numerically higher than their mean false-alarm rate ($40.62\% \pm 22.33\%$), but this difference was nonsignificant, $t(15) = 1.48$, $p = 0.17$, JZS Bayes Factor = 0.64; that is, there was no reliable evidence of explicit memory of the spatial context.

Discussion

Experiment 2 revealed both a procedural learning effect and a contextual-cueing effect during the training session, further confirming that repeated spatial context in 3D vision can be effectively learned to guide search. However, the contextual facilitation vanished completely when the left and right halves of the display were swapped. By contrast, comparable swaps in depth between the front and back planes in Experiment 1 had no effect on the magnitude of contextual cueing. Interestingly, the repeated context in the test phase was relearned after two epochs of repetition, suggesting that novel contextual learning was successful after the change.

General discussion

The present study investigated the role of depth information in 3D layouts in a contextually guided visual search task. During the training session of both experiments, participants searched for a T-shaped target among L-shaped distractors in 3D layouts, with half of the search items presented in the front and the other half in the back layer of the 3D display. During the following test sessions, the front and back depth planes were swapped in Experiment 1 or the left and right display halves were exchanged in Experiment 2, yielding comparable contextual changes from training to test along the Z - and X -axis, respectively. The results showed significant contextual cueing facilitation during the training session in both experiments, suggesting that spatial contexts distributed in 3D space could be learned effectively to facilitate search. Additional recognition tests at the end of the experiments suggested that learning of the 3D layouts was implicit; that is, observers could not explicitly recognize the (original) repeated displays.

The results of the test sessions revealed that the facilitation deriving from the learned contextual associations could only be transferred across front–back swaps, but not across left–right swaps. Note that the only change in Experiment 1 between the original display and the swapped display was the depth ordering, while the item associations on the 2D retinal image as well as on each depth plane were preserved. By contrast, the left–right change of the two halves of the displays in Experiment 2, while preserving the depth of individual items, caused rather large variation in the inter-item associations within each depth plane. Accordingly, contextual cueing appears to be rather flexible to, and can thus be transferred across, depth changes as long as the inter-item associations remain (largely) unchanged. This might be taken to mean that

the acquired contextual memory representation of a (repeated) display layout encodes inter-item associations without particularly registering binocular disparity information; that is, mental representations of 3D spatial context may be based essentially on a 2D projection of the scene. Thus, the spatial context memory may derive primarily from item associations within X/Y coordinates, while relying to a lesser extent on associations along the Z -axis.

Importantly, representations of the 3D spatial context in terms of a 2D projection do not necessarily mean that depth plays no role in contextually guided visual search. Rather, depth might still act as an effective cue to group the visual search items into different (depth-defined) clusters, with contextual associations being established primarily within (but not across) segmented groups. For instance, various studies (with 2D search layouts) have shown that subset formation on the basis of color similarity (Conci & Von Mühlenen, 2011; Geyer, Shi, & Müller, 2010), region segmentation (Conci, Müller, & Von Mühlenen, 2013; Conci & Von Mühlenen, 2009; Zang et al., 2016), or with temporally segmented clusters (Hodsoll & Humphreys, 2005) constrains the build-up of contextual associations, which tend to evolve primarily within, but not across groups of items. A comparable mechanism might also operate for depth-defined displays. In this view, grouping in depth leads to a segmentation of the search layout into distinct depth-defined subsets, which in turn constrains the learning of contextual associations. For instance, Kawahara (2003) found contextual cueing to be dependent on the invariant arrangement of the search items in the depth layer that contained the target, whereas cueing is unaffected by variation of the item relations in the nonattended layer (i.e., the layer that did not contain the target). Moreover, a change of the depth ordering of items within a segmented layer reduces contextual cueing (Kawahara, 2003), whereas an overall change in depth of the entire group of items leaves the magnitude of cueing unaffected (as in Experiment 1). It thus appears that the segmentation of the display provides a subset within which the local contextual relations of the target to the items in its vicinity can then be learned.

In addition to a potential role of depth-based segmentation in constraining contextual cueing, our study found that the transfer of contextual learning in 3D crucially depends on the maintenance of item associations of the 2D retinal image. Previous studies already emphasized the role of item associations in memory-guided search tasks. For instance, Beesley et al. (2015) found that invariant distractor–distractor associations support the manifestation of the contextual-cueing effect. Conversely, Endo and Takeda (2005) reported that the contextual facilitation was reduced when swapping the search items between the upper and

lower halves of a display. Moreover, Manginelli and Pollmann (2009) and Zellin, Conci, Von Mühlenen, and Müller (2013) found that a change of the target location leads to sustained costs in contextual cueing. However, these studies were all limited to 2D visual search, and our study, for the first time, presents evidence that also supports the essential role of inter-item associations in a 3D context-guided search task.

The importance of item associations in contextual cueing may also explain differences in the results between the current experiments and previous studies (Chua & Chun, 2003; Kawahara, 2003), with the latter suggesting that depth information is incorporated in contextual memory. In Chua and Chun's (2003) study, 2D search displays with apparent depth information were presented, whereby depth was manipulated “apparently” via the size of the objects. When changing the viewing angle from training to transfer, a reduced cueing effect was found with an increasing difference in rotation angle. However, this rotation of the display in pseudo-depth also systematically changed the inter-item associations on the 2D retinal image, which, on the basis of our results, are actually crucial for exploiting a learned layout.

An actual influence of binocular disparity in contextual cueing was previously investigated by Kawahara (2003), who reported that contextual cueing decreased with a reversal in the disparity of the distractors while keeping the target in the same depth plane. In his study, however, binocular disparity variations potentially weakened or removed the learned local item association within a given depth layer that contained the target (because the spatial configuration of the target's neighboring items changed to a different depth plane), again suggesting that the reduction of contextual cueing resulted from changes to the learned inter-item associations within particular depth planes and not because of pure binocular disparity-defined depth variations. Together with the current results, these findings may be taken to suggest that distractors are preferentially learned if presented in the same depth plane as the target, so that a change in depth of the entire learned configuration does not affect contextual cueing. Our results agree with the model of Brady and Chun (2007): they suggested that only few distractors in the vicinity—here, in the depth plane—of the target suffice to establish reliable contextual cueing; but when these local target–distractor contextual relations are disturbed, contextual cueing is considerably weakened.

In sum, our findings show that contextual learning depends to a large extent on intact and invariant inter-item associations, while at the same time, depth-related coding of disparity does not seem to be incorporated in contextual memory. Context-based learning therefore appears to extract and encode, in some sense, a rather

vague or crude representation of the reoccurring statistics in the environment.

Keywords: 3D visual search, contextual learning, contextual retrieval

Acknowledgments

We would like to thank Ms. Chen Cui for her help with data collection and help with drawing the 3D figures. This research was supported by the German Research Foundation (DFG, Grants CO 1002/1-1 and GE 1889/4-1 to MC and SH 166/3-1 to ZS); the National Natural Science Foundation of China (Grant 31600876 to XZ); and the China Postdoctoral Science Foundation (Grant 2016M600663 to XZ). The funding bodies played no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Commercial relationships: none.

Corresponding author: Xuelian Zang.

Email: zangxuelian@gmail.com.

Address: College of Psychology and Sociology, Shenzhen University, China.

Footnote

¹ It should be noted that, in the current study, depth was defined by not only disparity, but also the presentation of semitransparent layers in the display (see Figure 1), strengthening the segmentation in depth (e.g., as compared to Kawahara, 2003). But irrespective of the strong depth cues, swapping of the depth layers did not influence the magnitude of contextual cueing.

References

- Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cueing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 348–362.
- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cueing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 798–815.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Chua, K. P., & Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. *Perception & Psychophysics*, *65*(1), 72–80.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365, doi:10.1111/1467-9280.00168.
- Conci, M., & Müller, H. J. (2012). Contextual learning of multiple target locations in visual search. *Visual Cognition*, *20*(7), 746–770.
- Conci, M., Müller, H. J., & Von Mühlénen, A. (2013). Object-based implicit learning in visual search: Perceptual segmentation constrains contextual cueing. *Journal of Vision*, *13*(3):15, 1–17, doi:10.1167/13.3.15. [PubMed] [Article]
- Conci, M., Sun, L., & Müller, H. J. (2011). Contextual remapping in visual search after predictable target-location changes. *Psychological Research*, *75*(4), 279–289.
- Conci, M., & Von Mühlénen, A. (2009). Region segmentation and contextual cueing in visual search. *Attention, Perception, & Psychophysics*, *71*, 1514–1524.
- Conci, M., & Von Mühlénen, A. (2011). Limitations of perceptual segmentation on contextual cueing in visual search. *Visual Cognition*, *19*(2), 203–233.
- Endo, N., & Takeda, Y. (2005). Use of spatial context is restricted by relative position in implicit learning. *Psychonomic Bulletin & Review*, *12*(5), 880–885.
- Geyer, T., Shi, Z., & Müller, H. J. (2010). Contextual cueing in multiconjunction visual search is dependent on color- and configuration-based intertrial contingencies. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 515–532. doi:10.1037/a0017448.
- Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends In Cognitive Sciences*, *19*(9), 524–533.
- Hodsoll, J. P., & Humphreys, G. W. (2005). Preview search and contextual cueing. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1346–1358.
- Howard, J. H., Jr., Dennis, N. A., Howard, D. V., Yankovich, H., & Vaidya, C. J. (2004). Implicit spatial contextual learning in healthy aging. *Neuropsychology*, *18*, 124–134.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in

- spatial contextual cuing—Configuration or individual locations? *Perception and Psychophysics*, 66(3), 454–463.
- Kawahara, J.-I. (2003). Contextual cueing in 3D layouts defined by binocular disparity. *Visual Cognition*, 10(7), 837–852.
- Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search? *Psychological Research*, 73, 212–221.
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, 9(3), 273–302.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Schankin, A., & Schubö, A. (2009a). Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. *Psychophysiology*, 46(3), 668–679.
- Schankin, A., & Schubö, A. (2009b). The time course of attentional guidance in contextual cueing. In L. Paletta & J. K. Tsotsos (Eds.), *Lecture Notes in Computer Science: Vol. 5395. Attention In Cognitive Systems* (pp. 69–84). Berlin, Germany: Springer.
- Shi, Z., Zang, X., Jia, L., Geyer, T., & Müller, H. J. (2013). Transfer of contextual cueing in full-icon display remapping. *Journal of Vision*, 13(3):2, 1–10, doi:10.1167/13.3.2. [PubMed] [Article].
- Song, J. H., & Jiang, Y. (2005). Connecting the past with the present: How do humans match an incoming visual display with visual memory? *Journal Of Vision*, 5(4):4, 322–330, doi:10.1167/5.4.4. [PubMed] [Article].
- Zang, X., Geyer, T., Assumpção, L., Jia, L., Müller, H. J., & Shi, Z. (2016). From foreground to background: How task-neutral context influences contextual cueing of visual search. *Frontiers in Psychology*, 7, 852.
- Zang, X., Jia, L., Müller, H. J., & Shi, Z. (2015). Invariant spatial context is learned but not retrieved in gaze-contingent limited-viewing search. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(3), 807–819.
- Zellin, M., Conci, M., Von Mühlénen, A., & Müller, H. J. (2013). Here today, gone tomorrow—Adaptation to change in memory-guided visual search. *Plos ONE*, 8(3), e59466.
- Zellin, M., Von Mühlénen, A., Müller, H. J., & Conci, M. (2014). Long-term adaptation to change in implicit contextual learning. *Psychonomic Bulletin & Review*, 21(4), 1073–1079.