Examining the role of depth information in contextual cuing using a virtual reality visual search task

Tom Beesley[1], Ying Yun Tou[2], and & Jack Walsh[2]

[1] Lancaster University
UK
[2] UNSW Sydney
Australia

Author Note

Tom Beesley, Department of Psychology, Lancaster University. Ying Yun Tou, School of Psychology, UNSW Sydney. Jack Walsh, School of Psychology, UNSW Sydney.

The data, analysis scripts, experimental materials, and the manuscript source files, are available at github.com/tombeesley/CCVR.

Correspondence concerning this article should be addressed to Tom Beesley, Department of Psychology, Lancaster University, UK, LA1 4YD. E-mail: t.beesley@lancaster.ac.uk

Abstract

Two experiments examined the role that depth plays in the formation of associations during contextual cuing of visual search. Current associative models make predictions about the spatial constraints placed on learning within two-dimensional procedures, but there exists very little evidence of how these predictions translate to three-dimensional space. A virtual reality procedure was used to project the stimuli in three-dimensions. Experiment 1 established a contextual cuing effect using this procedure, while Experiment 2 examined whether the relative distance between repeated distractors and the target, or the position of the distractors relative to the observer modulated contextual cuing. It was found that the contextual cuing effect was consistent across these different conditions. As a result, there was no evidence to suggest that depth information forms a significant part of the representations that form during contextual cuing. These data are therefore broadly consistent with the mechanisms of current associative models of contextual cuing.

Public significance statement: This study provides a test of the way in which visual scenes are processed and learnt about across three-dimensional space. Current models of this behaviour only consider learning in two-dimensional space. The current work therefore provides important insights into how depth information contributes to the memory representations of familiar scenes.

Keywords: Visual search, contextual cuing, depth, virtual reality

Word count: 7971

Examining the role of depth information in contextual cuing using a virtual reality visual search task

Visual search through the environment is guided by stored representations of our past experience. Within the laboratory setting, the power of this cognitive process is perhaps best demonstrated by studies of "contextual cuing" (Chun & Jiang, 1998, hereafter "CC"). In a typical task, participants have to locate a target amongst a set of distractors as quickly as possible while maintaining high accuracy. Unbeknownst to participants, some of these visual search displays are repeated across trials of the experiment, such that some trials contain "repeated configurations," and others "random configurations." A CC effect is shown by faster reaction times to repeated configurations compared to random configurations, an effect which emerges quickly in the task, and is robust and observed in the majority of participants, with an effect size typically above $d_z = 1$ (Vadillo et al., 2016). Some have claimed that the CC effect is the result of an implicit learning system, which does not yield consciously accessible memories of repeated configurations (e.g., Colagiuri & Livesey, 2016), though this continues to be the subject of debate (e.g., Kroell et al., 2019; Vadillo et al., 2016).

The CC effect fits a clear associative framework for memory representation. Participants learn to associate the repeating configurations of distractors with the location of the target, such that the perception of the configuration of distractors (or at least a sample of the configuration) triggers a learnt behavioural response, likely to be primarily an earlier ceasing of random search behaviour and a shift of attention towards the target location (Beesley et al., 2018). Brady and Chun (2007)'s associative model provided an important first step in formalising the representational theory of CC. The model is based on a simple two-layer connectionist network, with input units reflecting the spatial representation of present stimuli (candidate targets) in the visual field, and output units reflecting the expectation for a target in each of these positions. Simple error-correction learning mechanisms drive the formation of associations between the input units for present (distractor) stimuli and the output unit representing the target position. These associations form in the case of repeated configurations but not random configurations, and so the model easily provides a simulation of the CC effect.

The model also sought to explain empirical data showing that CC occurs most readily between those distractors appearing proximally to the target. Olson and Chun (2002) segmented their display into two halves and local contexts were defined where the repeated distractors shared the same half of the screen as the target. Long-range contexts on the other hand, had repeated distractors found on the opposite side of the screen to the target. Observers were trained on repetitions of either contexts and a facilitation in target localisation was observed when local information were repeated. Long-range contexts had been shown to also facilitate search only when there is no extraneous information, such as non-predictive distractors, segregating it from the target. This result demonstrated two important findings – first, the implicit learning of visual context is sensitive enough to parse noise from signal within a given context and second, perceptual constraints limited to spatially proximal objects influence the information that is processed and encoded into our spatial maps. The Brady and Chun (2007) model incorporates *spatial constraints* on the learning of associations, such that changes to the associations for those distractors that are close to the target occur more readily compared to those that are further away from the target (see Figure 1).

Extensions of the Brady and Chun (2007) model have looked to extend and refine this associative framework for representing CC. Beesley et al. (2015) demonstrated the importance of local associative connections between input units, such that memory enocding of the configuration occurs irrespective of the associations that exist between the distractors and the target location. Similarly Beesley et al. (2016) developed a model that works with "configural representations" at the input layer and associates the entire configuration with the target location.

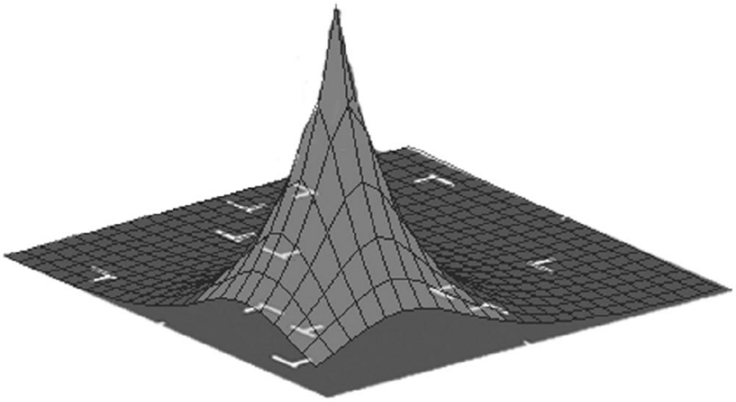

Figure 1

Illustration of the spatial constraints on associative weight change in the Brady and Chun (2007) model

Importantly, all of these models have worked with two-dimensional representations of the search space, encoding the spatial information on an x/y coordinate space. This is because the vast majority of visual search research, and nearly all CC research, has presented stimuli across a two-dimensional coordinate space (i.e, a screen). Yet a great many (perhaps the vast majority) of real-world visual search tasks are performed on stimuli that exist in a three-dimensional coordinate space. From both an empirical and theoretical (model) perspective, it is important to determine the role that depth plays in the formation of representations of visual scenes and how that affects the CC of visual search.

It is well acknowledged that the visual system is highly attuned to the processing of depth information, and that the position of stimuli across depth planes can have important effects on visual search efficiency (e.g., McSorley & Findlay, 2001; Nakayama & Silverman, 1986). We focus here on the small group of studies that have explored the role of depth information in CC of visual search. Kawahara (2003) (Experiment 2A and 2B) used shutter glasses to render stereoscopic projections of stimuli across two depth planes, *near* and *far* from the viewer, with targets appearing equally across the two planes. Critically they were able to show a CC effect with this procedure, and when the distractors across the two planes were swapped in a final stage, the CC effect was reduced, though not completely abolished. The reduction of the effect demonstrates that the CC that had been established was reliant on the encoding of depth information, which was disrupted by the manipulation in the final stage.

Zang et al. (2017) have suggested that Kawahara's (2003) method of maintaining the position of the target during the switch of distractor depths may have disrupted the local associations between the target and distractors in terms of their two-dimensional representations (x/y). To test this, Zang et al. (Experiment 1) subjected entire configurations to a depth reversal, including the depth of the target. Interestingly, they found that, in contrast to the findings of Kawahara (2003), the CC effect remained intact after the swap of depth planes. When the configurations of distractors were disrupted by switching the left and right sides (Experiment 2), the CC effect was significantly attenuated. Thus the results of Zang et al.'s (2017) Experiment 1 call into question how critical depth information is in the representations driving CC, at least to the extent that associations either do not appear to be forming across different depth planes, or are resistant to significant generalization decrements.

The current set of experiments further explore the role that depth plays in CC of visual search. The experiments use a virtual reality (VR) device to project the stimuli in stereoscopic 3D. This allows stimuli to be positioned across a wide field of view, with head movements required to process the visual display in full during the search process. Experiments 1a and 1b provide a simple demonstration of the CC effect in this novel procedure with targets placed relatively near to, or far from, the observer. In Experiment 2 we manipulate the validity of distractors at these different depths to explore the structure of the associative representations that form during visual search across three-dimensions. To our knowledge, the only experiment that has used a VR device to examine CC is the recent paper by Marek and Pollmann (2020). This experiment used a design that was quite similar to that used in Experiment 1 (at the time of running the experiment we were unaware of this work): participants experienced two sets of repeating configurations, with one set paired with targets near to the observer, and one set paired with targets that were further from the observer. They observed CC effects for configurations paired with near and those paired with far targets (we return to a discussion of these data later). Similarly, the primary aim of Experiment 1 was to establish a CC effect within our VR procedure for configurations with targets at different depths, before exploring the role that depth plays in the representations contributing to the CC effect in Experiment 2.

Transparency and Openness

The data, analysis scripts, experimental materials, and the manuscript source files, are available at github.com/tombeesley/CCVR. We report our determination of sample sizes, the data exclusion criteria, and all of the manipulations. The analyses reported in this manuscript are computationally reproducible from the manuscript source files (using R v4.1.1), which are available at the github repository. The study design and analyses were not pre-registered.

Experiment 1

The current experiment aimed to demonstrate CC in the VR procedure, while also providing a simple manipulation of the target depth. Stimuli in the experiment were positioned at one of two distances from the observer (hereafter ‘near’ and ‘far’ depths). On one half of all trials, repeated configurations were presented, in which all distractors across the two depths were positioned in set locations. On the other half of all trials, all distractors were arranged randomly. Orthogonal to this factor, targets were positioned at either near or far locations from the observer.

Vadillo et al. (2016) estimated an effect size of $d_z = 1$ in typical CC experiments. Given the novelty of the procedure, we took a cautionary approach by estimating a diminished effect size of $d_z = 0.50$. To achieve power of 0.90, with $\alpha = 0.05$, required a minimum sample of 44, which was achieved in Experiment 1 (Experiment 1a and 1b combined).

While running Experiment 1, we observed that participants struggled to complete certain trials, and the experiment could not progress until the target had been accurately detected. To alleviate this issue, we added a timeout to the trial procedure, terminating the trial after 10 seconds if no valid response had been made. We also made other minor changes to the procedure. We present the original procedure as Experiment 1a and the modified procedure as Experiment 1b. We assess the impact this modification made to the ease of target detection, and we include the "sub-experiment" factor in our analysis.

Experiment 1a

Method

Participants.

Twenty-five undergraduate psychology students (mean age = 19.44, SD = 2.02; 19 identified as male and 6 as female) from UNSW Sydney participated in the study in exchange for course credit (one participant failed to complete the task and their data were not analysed). Participants were required to have normal or corrected visual acuity. Participants wearing glasses (but not those wearing contact lenses) were excluded from this experiment due to constraints with the amount of space available in the VR headset. Six participants were excluded from the final analysis (see below for details).

Materials.

An Oculus Rift CV1 headset (Oculus VR), connected to a PC with a NVIDIA GeForce GTX 970 graphics card, delivered the experiment in VR. The Oculus Rift CV1 is a head-mounted display that utilises a stereoscopic OLED display with 2160 x 1200 resolution (1080 x 1200 per eye) at a refresh rate of 90 Hz. Within the headset, images on the computer undergo transformations through two convex hybrid Frensel lenses, warping images and extending the environment to a wide 110° field of view. The headset has a dial that adjusts the separation of the lenses to accommodate for the varying interpupillary distances across participants. Two external infrared tracking sensors were used to track the headset and position the user in 3D space. The experiment was programmed in the Unity engine, which created all the stimulus properties, controlled timing and recorded responses. All dimensions of the program were measured in "Unity units." The spacebar was used for indicating the detection of the target.

The VR environment depicted an empty room with grey walls where the viewer was situated in the middle of the room. The program allows objects to be placed on any of nine concave surfaces, which radiated from the observer at the origin point (see Figure 2). Two of these surfaces were selected for the two depths (near and far) used in the task. Objects on each surface were approximately equidistant from the observer. Each surface represented an increment of 2 units from the observer.

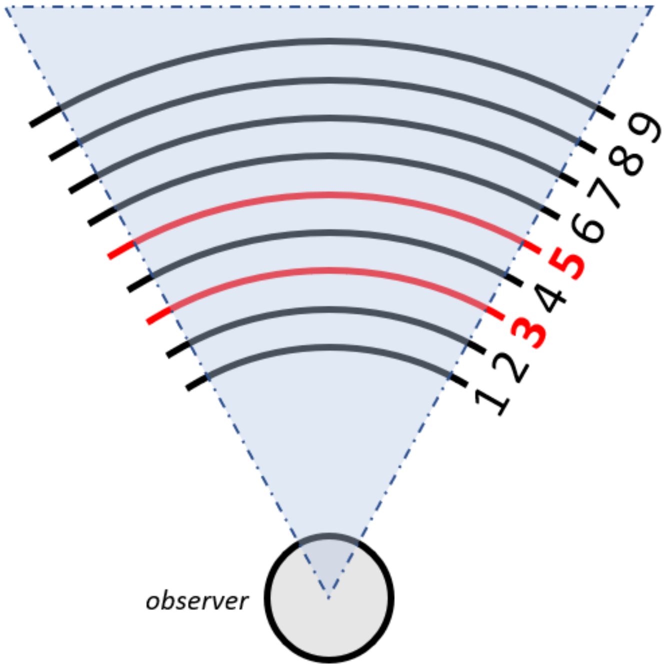

Figure 2

Schematic of the surface arrangement from a top-down perspective. The the choices for 'near' (3) and 'far' (5) surfaces are shown in red.

Object placement was restricted such that objects did not appear too far to the left or right of the viewer and too far above and below the viewer. As it was possible for stimuli to appear close to the observer, head movements were required to examine the configurations and to find the target. Surfaces 3 and 5 were chosen to provide an environment where stimuli were quite clearly positioned at different depths, while balancing the need for excessive head movements.

Each trial contained 16 cylinder-shaped distractor objects and one capsule-shaped (rounded) target. All stimuli were red and had a scale of 0.3 x 0.24 x 0.3 Unity units (x,y,z) regardless of depth. There were no constraints on the rotation on the x, y and z-axis for each target and distractor, but rotations were fixed for repeating configurations.

A semi-transparent blue disk was used in the practice phase, scaled at 0.18 x 0.18 x 0.001 Unity units (x,y,z), was presented in the middle of the participant's visual field (see Figure 3). A white fixation disc was used throughout the task for refocusing, which was scaled at 1 x 1 x 0.001 Unity units(x,y,z).

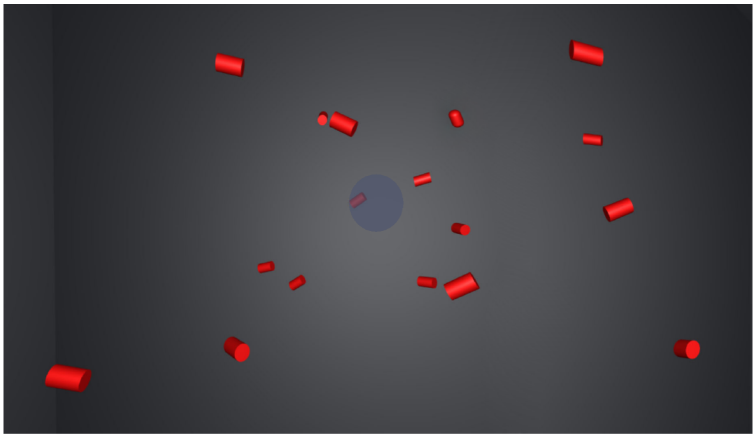

Figure 3

Example stimulus arrangement. The central blue disc represents the centre of the participants view and was only present during the practice trials. The rounded target shape is located in the upper right quadrant.

Design.

A within-subjects design was used, with independent variables of *configuration type* (repeated vs. random), *target-depth* (near vs. far), and epoch (6 epochs, each comprising 5 blocks). The dependent variable of primary interest in all experiments was reaction time (head movements were recorded and those data are available at the data repository, but are not reported).

Each configuration comprised 16 cylinder-shaped distractors, which were equally distributed across the two surfaces, resulting in eight distractors per surface (Figure 3). Within each surface, the eight distractors were distributed equally across the four quadrants of the x/y plane. The target was placed on one or other surface depending on the configuration set (see below), and within each set the target was equally frequent in each quadrant of the x/y plane. The same target locations were used in the repeating and random configuration types that had targets at the same depth. Two sets of repeated configurations (four configurations in each set) were presented across the blocks of the experiment: *Near-T-repeated* had targets on the near surface and repeating distractor configurations on both surfaces, while *Far-T-repeated* had targets on the far surface and repeating distractor configurations on both surfaces. Two sets of “random configurations” were used in which distractors on the two surfaces were randomly arranged: *Near-T-random* had targets on the near surface and *Far-T-random* had targets on the far surface.

Procedure

Participants were given 10 practice trials to familiarize themselves with the task. This provided participants with a guide as to how they should direct their head position towards the centre of the target object, prior to responding with the spacebar. Participants were told to search for a “pill-shaped target” and to press the spacebar once the central blue circle region was over the target object. This provided a guide to how the program would register accurate target detection. The blue circle was used for training only and not presented in the main phase of the experiment. After a successful detection of the target, the white fixation circle appeared. The program detected the gaze on this fixation circle before initiating the next trial.

The main task consisted of 480 trials, divided into 30 blocks (16 trials per block). Each block contained four trials from each of four different configuration sets. Within each set, and within each block, targets were equally distributed across the four quadrants of the x/y plane. The order of trials within each block was randomised, with the constraint that the target could not occupy the same location across consecutive trials. Participants were given breaks every 160 trials (at the end of every 10 blocks) and breaks lasted until participants wished to resume.

Reaction times were measured in milliseconds (ms) from the beginning of a trial to an accepted target detection response for each trial. Any rejected target detection responses - those that were made when the headset was not oriented accurately on the target - were recorded and used to determine the difficulty of target detection on any given trial.

Experiment 1b

Method

Participants.

Twenty participants (mean age = 23.65, SD = 5.39; 5 participants identified as male and 15 as female) took part in the study in exchange for AUS$15. All other aspects of participant recruitment, exclusions and requirements were identical to Experiment 1a. Two participants were excluded from the final analysis (see below for details).

Materials, Design, and Procedure.

Experiment 1b employed some minor alterations to the materials and procedure. Sixteen practice trials were used instead of 10. Breaks were given every 64 trials (every four blocks) compared to every 160 trials in Experiment 1A. In an attempt to combat issues with object occlusion and poor target detection rates, a timeout was given (the word "TIMEOUT" appeared) after 10 seconds of the trial. Experiment 1b also used an updated algorithm controlling the registration of the participants view over targets in an attempt to improve valid target detection.

Results

Since Experiment 1B used a timeout of 10 seconds, we imposed the same cutoff for our analysis of reaction times in Experiment 1A. Trials which led to timeouts and trials that had more than two response attempts were removed (i.e., up to 2 spacebar responses were allowed). Following this, RTs greater or less than 2.5 SDs from the participant mean RT were removed. Mean RTs for each participant were computed and the mean across the sample was 2310 ms (SD = 388). No participants were identified as outliers in terms of their mean RT.

Two participants had an unusually high proportion of trials removed (greater than 2.5 SDs of the mean proportion of trials removed), and we excluded these two participants from further analysis (both from Experiment 1a). We next examined the impact of these trial exclusions on the proportion of data contributing to the analysis across trials with targets at the two different depths. We identified 6 participants (4 from Experiment 1a and 2 from Experiment 1b) that had a high discrepancy in the proportion of excluded trials across the factor of target depth (greater than 15% difference in the proportion of trials excluded) and we excluded these participants from further analysis. The final proportion of trials contributing to the analysis in Experiment 1a was 93.9% and 88.2% for near and far targets, respectively, and in Experiment 1b was 87.9% and 88.4% for near and far targets, respectively.

Within-subject error bars were computed by a process of normalising the RT data for the sample (Cousineau, 2005). Figure 4 shows the RT data across the 6 epochs, plotted separately for the data from Experiment 1a and Experiment 1b. RTs decreased with practice on the task, and were shorter for repeated configurations than for random configurations. There was also a clear target location effect, with responses being much slower when the target was located on the near surface compared to the far surface.

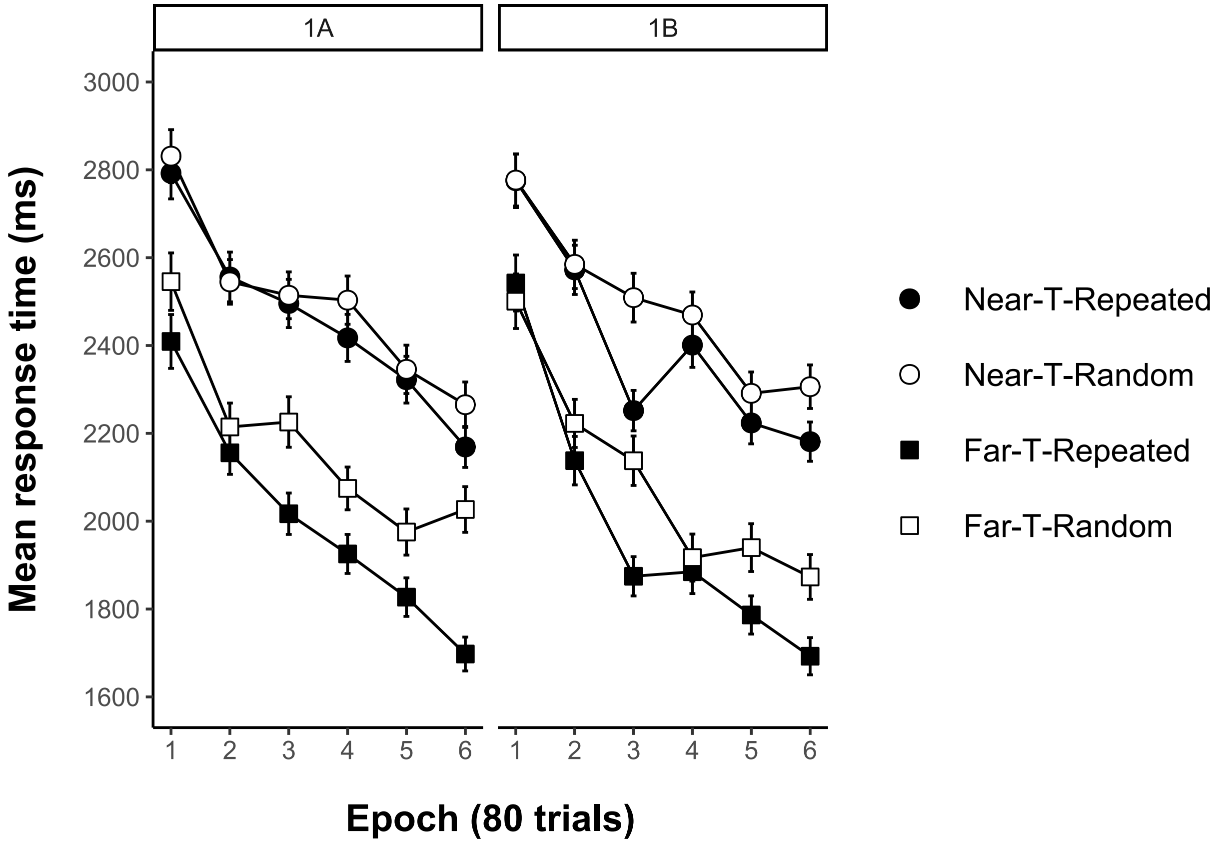

Figure 4

RT data for Experiment 1A and 1B. Circles reflect trials with near targets (T) and squares those with far targets. White symbols are used for trials with distractors that are entirely randomised, while black symbols are those trials with repeating distractors.

These data were subjected to a four-way repeated measures ANOVA, using the *afex::aov_car()* function, which included a between-subject factor of *sub-experiment* (1a vs. 1b), and within-subject factors of *configuration type* (repeated vs. random), *target-depth* (near vs. far) and *epoch* (1 to 6). Where Mauchly's test of sphericity revealed violations, Greenhouse-Geisser corrections were made. There was no main effect of *sub-experiment*,

$F(1,\ 35) = 0.07$, $p = .800$, $\eta_p^2 < .01$ and the only significant interaction effect involving *sub-experiment* was the *sub-experiment* by *configuration type* by *epoch* interaction, $F(3.78,\ 132.16) = 2.76$, $p = .033$, $\eta_p^2 = .07$, suggesting that the contextual cuing effect may have emerged at slightly different rates in the two procedures.[1]

There was a main effect of *configuration type*, $F(1,\ 35) = 15.58$, $p < .001$, $\eta_p^2 = .31$, with response times for repeated configurations (2220.6 ms, SD = 379) faster than those for random configurations (2323.7 ms, SD = 380). There was a main effect of *target-depth*, $F(1,\ 35) = 80.12$, $p < .001$, $\eta_p^2 = .70$, with response times for far targets (2076.6 ms, SD = 384.5) faster than those for near targets (2467.8 ms, SD = 402.8). There was also a main effect of *epoch*, $F(3.50,\ 122.46) = 69.44$, $p < .001$, $\eta_p^2 = .66$ with response times decreasing across the experiment. The *configuration type* by *target-depth* interaction was not significant, $F(1,\ 35) = 3.23$, $p = .081$, $\eta_p^2 = .08$. The *configuration type* by *epoch* interaction was significant, $F(3.78,\ 132.16) = 4.73$, $p = .002$, $\eta_p^2 = .12$, as was the *target-depth by epoch* interaction, $F(3.98,\ 139.39) = 3.84$, $p = .006$, $\eta_p^2 = .10$, suggesting that the RT decrease was more pronounced for configurations with far targets. There were no other significant interaction effects.

While there was a clear main effect of *configuration type*, the present data present some uncertainty as to whether this CC effect is modulated by the depth of the target (the *configuration type* by *target-depth* interaction). A Bayesian ANOVA was conducted, using the *BayesFactor::anovaBF()* function (priors were set at the default "medium" width). This found that there was no confirmatory evidence for the absence of an interaction effect, BF=0.40 (a value of 0.33 would reflect moderate support for the null, Dienes, 2014). Thus the experiment was somewhat underpowered for detecting differences in the size of the contextual cuing effect across these conditions.

---

[1] The sample size for Experiment 1 was determined on the basis of detecting within-subjects differences in RT. We acknowledge that the between-subjects comparisons across the factor of sub-experiment may be underpowered.

The overall CC effect (the main effect of *configuration type* from the ANOVA) resulted in a Cohen's $d_z = 0.66$, with 78% of participants showing a positive CC effect. The CC effect was present in both configurations with near targets, $t(36) = 2.08$, $p = .045$, $d_z = 0.34$, and those with far targets, $t(36) = 3.76$, $p < .001$, $d_z = 0.62$.

Discussion

Experiment 1 established a VR procedure for observing a CC effect, which resulted in a moderately large effect size of $d_z = 0.66$, but one that is slightly weaker than the CC effect seen in typical standard 2D implementations of the task (Vadillo et al., 2016). This effect was present for both configurations paired with near and far targets. The effect was numerically larger in the case of configurations paired with far targets, but the difference between the CC effect for near and far targets was not significant. However, we note that the experiment was potentially underpowered to detect what could have been small differences in the size of the CC effect between these conditions.

Marek and Pollmann (2020) have recently demonstrated a CC effect in a VR procedure using a similar design. Like in our task, participants experienced targets appearing on either a near surface or a far surface. In their task, and unlike in our results, participants took longer to respond to targets on the far surface compared to the near surface. This is likely to be because, in Marek and Pollmann's (2020) task, the targets on the far surface were placed outside of the immediate field of view, and hence head movements were needed to locate the targets on the far surface. In contrast, in Experiment 1 it was the targets on the near surface that could be placed outside of the initial field of view, whilst the targets on the far surface had a far greater chance of being contained within the initial field of view. The data from Experiment 1 and those of Marek and Pollmann (2020) therefore demonstrate, somewhat unsurprisingly, that the presence of the target within the initial field of view is critical to the speed of target detection.

Marek and Pollmann (2020) also found that the CC effect was equivalent in magnitude for configurations paired with near and far targets. In our analysis we did not see any difference in the size of the CC effect between these conditions, though Bayesian analysis failed to find supportive evidence of an absence of this effect. Taken together, the current data and those of Marek and Pollmann (2020) suggest that CC develops for both configurations with targets at different depths, and that the VR procedure is suitable for exploring the associations that develop during CC.

Experiment 2

Experiment 1 demonstrated significant CC for configurations paired with near and far targets. In Experiment 2 we sought to examine the role that depth plays in the formation of associations *within* a repeating configuration. The data from Zang et al. (2017) suggested that when the distractors at different depths were switched (near distractors placed further from the viewer; far distractors placed nearer), then the CC effect was retained. This generalisation of performance across these different depths suggests a flat, 2D representation of the context that possibly does not encode depth information. If this is true, we would expect to see equivalent CC effects for distractors placed at different depths within a configuration. To test this, Experiment 2 selectively paired distractor configurations at one of the two depths, with targets that were placed at one of the two depths. As such, for each repeating configuration of distractors, half of the distractors were repeated across presentations and the other half were randomly arranged (see Table 1). For some participants, the distractors on the proximal surface to the target repeated, while for other participants the distractors on the distal surface to the target repeated. By also manipulating the target depth (relative to the observer) across configurations, this produced orthogonal factors of target depth, and repeated distractor depth.

This design allowed us to test the strength of associations forming between distractors and targets (i.e., to test whether CC is greater when repeating distractors are positioned close to the target or far away from the target). We decided to manipulate the

Table 1

Types of repeated configurations in Experiment 2

Condition	Target position	Repeated distractors	Random distractors
Proximal-NT	Near	Near	Far
Proximal-FT	Far	Far	Near
Distal-NT	Near	Far	Near
Distal-FT	Far	Near	Far

relationship between the repeating distractors and the target on a between-subjects basis in order to reduce the total number of configurations participants were exposed to (given the clear limitations on learning in standard CC designs, cf. Smyth & Shanks, 2008). Thus participants were exposed to eight repeated configurations: four with the target placed on the near surface, and four with the target on the far surface (within each set, the target was placed once within each quadrant).

In order to maximise the learning of these distractor-target associations (in what were quite stochastic configurations), we exposed participants to repeated configurations in "training" blocks, without the presentation of any random configuration trials. In order to then test for CC, we presented the entirely random configurations alone in "test" blocks. If CC had occurred during training, then we would expect to see increases in RT on the test blocks. By ensuring random configurations shared the same target positions, we were able to examine the CC effect for different types of configurations. For example, performance on the repeated configurations with near targets can be directly compared to the performance on the random configurations with those same near targets.

We also included an "awareness test" at the end of the experiment. Participants were presented with repeated configurations from the main experiment and newly generated configurations. On each presentation, the target was replaced with a distractor and participants were asked to identify (guess) which distractor had replaced the target.

Method

Participants.

Sixty-eight participants (mean age = 19.32, SD = 2.33; 20 participants identified as male and 48 as female) from UNSW Sydney took part in the study in exchange for AUS$15. Participants were randomly allocated to one of the two target-distractor contingency conditions, “proximal” and “distal.” 34 participants were allocated to each condition. One participant from each condition was excluded from further analysis on the basis of their mean RT (see below). Five participants were excluded from the final analysis (see below for details).

All other aspects of participant recruitment, exclusions and requirements were identical to Experiment 1a and 1b.

Materials, Design, and Procedure.

The materials used were the same as those used in Experiment 1b, with the exception that surfaces 3 and 6 were used to accentuate the differences in depth (see Figure 2)

A mixed-model design was used, with a between-subject factor of *target-distractor contingency* (proximal vs. distal), and within-subject factors of *configuration type* (repeated vs. random), *target-depth* (near vs. far) and *epoch* (31 epochs, each comprising 2 blocks).

Configurations were created in the same manner as for Experiment 1a and 1b, except that the sets of repeated configurations contained repeating distractors on only one surface. Each presentation of one of these repeating configurations contained a novel random configuration of distractors on the “non-repeating surface.”

The procedure was the same as that used in Experiment 1b, with the exception that random configurations were only presented in epochs 9, 10, 19, 20, 29, and 30, and repeated configurations were not presented in these 6 epochs.

At the end of the task participants were given a “generation awareness test” , to test the extent to which participants could explicitly recall the relationship between the distractor configurations and the target location. Two blocks of 16 trials were given, which each included the 8 repeating configurations from the main task, and 8 new repeating configurations that had been created in an identical manner (i.e., these new configurations were repeated in the second block). On each trial, the target object was replaced by a distractor object. Participants had to centre their view over the distractor they thought was most likely to be the target and press the spacebar. The program computed the discrepancy between the true target and the centre of the visual field, computing this error in two-dimensional space (x and y). This error measure (in unity units) acted as the dependent variable for the awareness task. A smaller error score on repeated compared to random configurations would indicate a level of awareness of where the target was positioned in repeating configurations.

Results

Trials were excluded on the same basis as in Experiment 1a and 1b. Mean RTs for each participant were computed and the mean across the sample was 2190 ms (SD = 345). As detailed above, two participants were excluded as outliers, as their mean RT was more than 2.5 SDs higher than the mean of the sample. Three participants had a difference in their exclusion rate of greater than 15% between near and far targets (the same criterion as used in Experiment 1) and these participants were removed from further analysis. The final proportion of retained trials was 94.1% and 93.5% for near and far targets, respectively.

Figure 5 shows the RT data across the 31 epochs. RTs decrease with practice on the task, and are shorter for repeated configurations than for random configurations. Like in Experiment 1a and 1b, there is a clear target location effect, with responses being much slower when the target was located on the near surface compared to when it was on the far surface.

Learning effects under each condition were calculated by comparing RTs for those epochs containing random configurations (9, 10, 19, 20, 29, and 30) with the adjacent epochs containing repeated configurations (8, 11, 18, 21, 28, 31). The data from these three "test periods" are plotted in Figure 6. Overall there was a strong CC effect across conditions and test phases. Averaging across the test data, 81% of participants showed a positive (numerical) CC effect.

The data were subjected to a four-way mixed-model ANOVA with a between-subject factor of *target-distractor contingency* (proximal vs. distal), and within-subject factors of *configuration type* (repeated vs. random), *target-depth* (near vs. far) and *test* (test 1, 2, or 3). Where Mauchly's test of sphericity revealed violations, Greenhouse-Geisser corrections were made. There was a main effect of *configuration type*, $F(1, 61) = 32.51$, $p < .001$, $\eta_p^2 = .35$, with response times for repeated configurations (2068.5 ms, SD = 310.9) faster than those for random configurations (2185.6 ms, SD = 303.1). There was also a main effect of *target-depth*, $F(1, 61) = 194.42$, $p < .001$, $\eta_p^2 = .76$, with response times for far targets (1864.6 ms, SD = 331.2) faster than those for near targets (2389.4 ms, SD = 330.8). The main effect of *test* was also significant, $F(1.99, 121.58) = 64.49$, $p < .001$, $\eta_p^2 = .51$ with response times decreasing across the three tests (Test 1: 2299.4 ms, SD = 340.3; Test 2: 2111 ms, SD = 323.6; Test 3: 1970.7 ms, SD = 307.1). The main effect of *target-distractor contingency* was not significant, $F(1, 61) = 0.45$, $p = .503$, $\eta_p^2 < .01$.

There was a significant interaction between *configuration type* and *test*, $F(1.86, 113.48) = 6.01$, $p = .004$, $\eta_p^2 = .09$, indicating that the difference in response times between repeated and random configurations increased across the three tests. There was no significant *configuration type* by *target-depth* interaction, $F(1, 61) = 1.41$, $p = .239$, $\eta_p^2 = .02$, however there was a significant three-way interaction between *configuration type*, *target-depth*, and *test*, $F(1.91, 116.69) = 8.88$, $p < .001$, $\eta_p^2 = .13$. Post-hoc t-tests

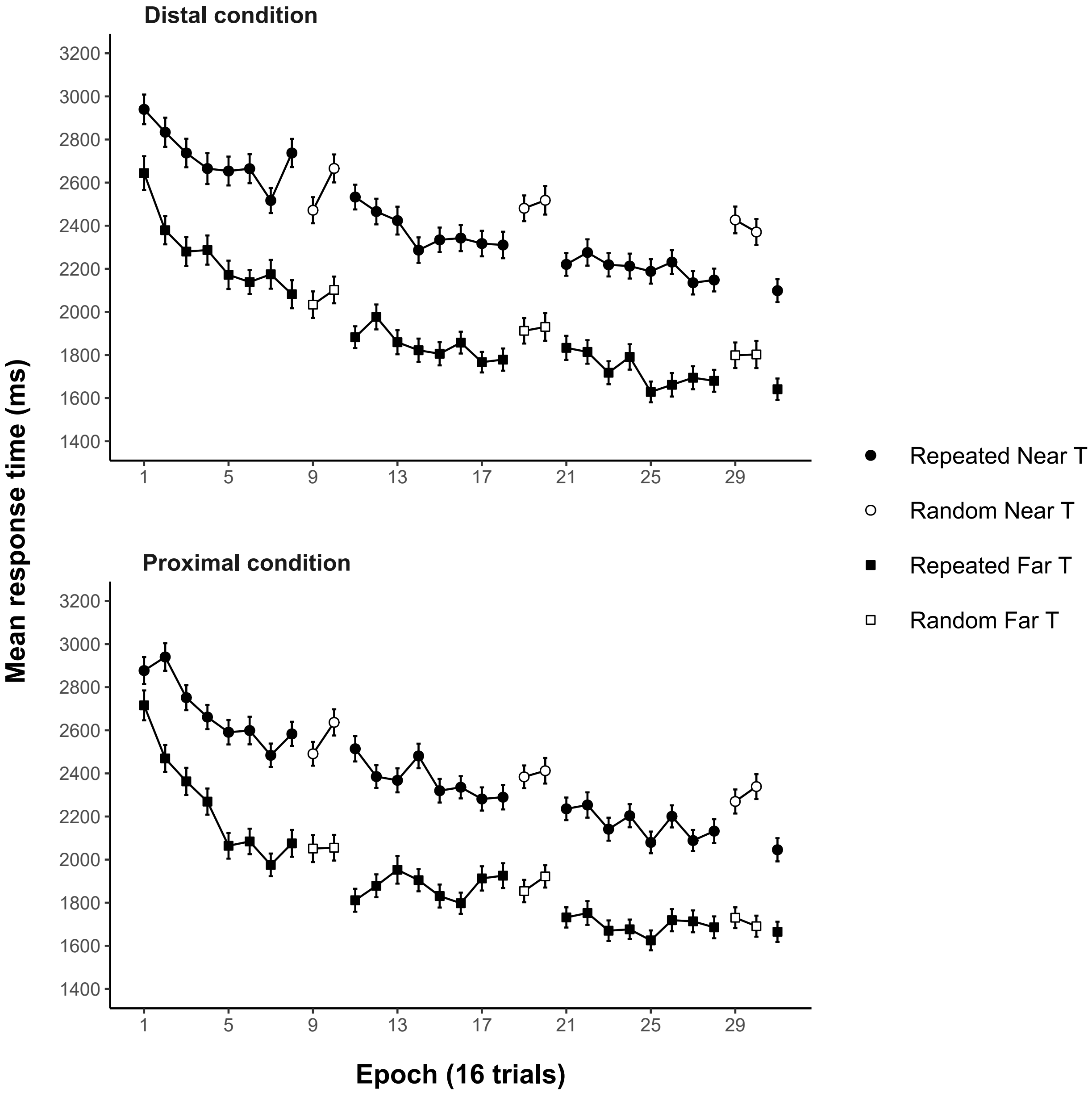

Figure 5

RT data for Experiment 2. Participants in the "Distal" condition received repeating configurations in which those distractors that were close to the target were randomised on each trial, while those distractors far from the target were predictive of the target location. In the "Proximal" condition this relationship was reversed: those far from the target were randomised, while those close to the target were predictive of the target location. Circles reflect trials with near targets (T) and squares those with far targets. White symbols are used for trials with distractors that are entirely randomised, while black symbols are those trials with repeating distractors.

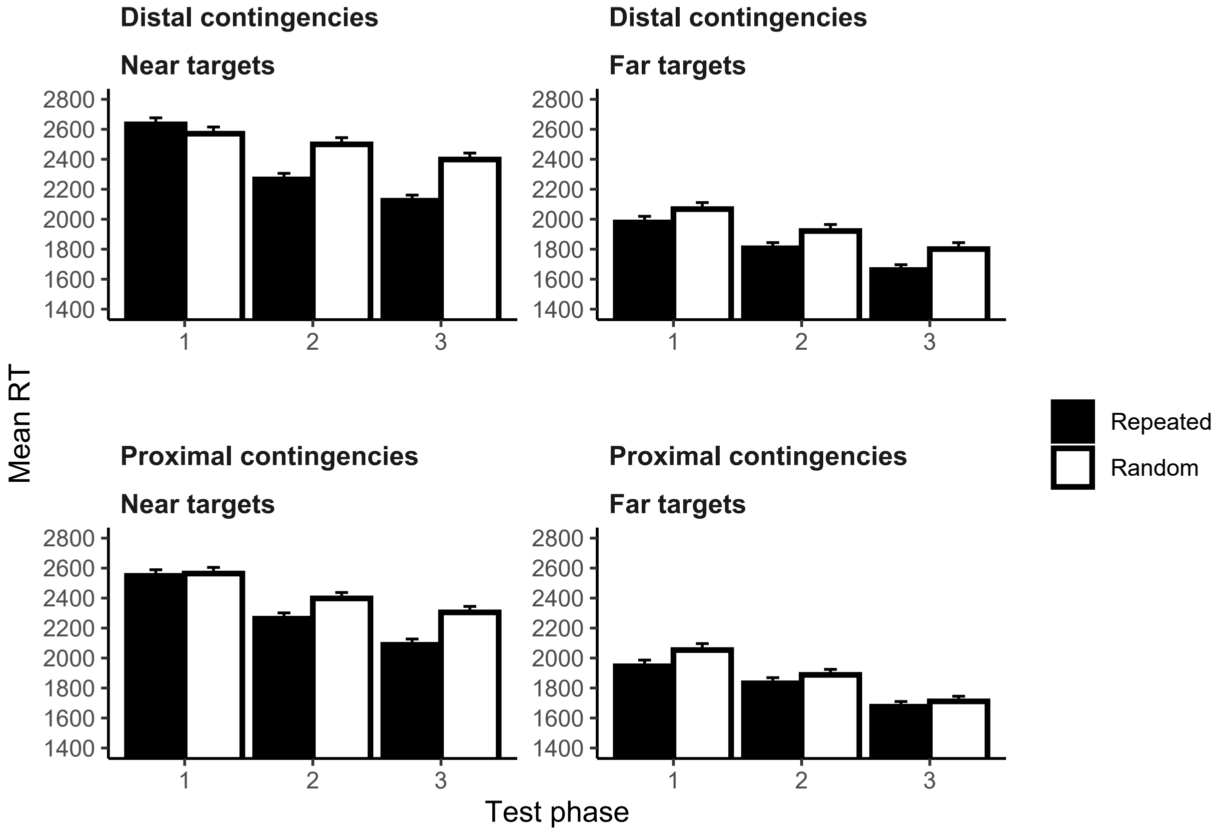

Figure 6

RT data from the "test blocks" of Experiment 2. Each test compares the reaction times in blocks containing random configurations to reaction times in adjacent blocks that contained repeated configurations.

compared differences in CC (RT to random configurations minus RT to repeated configurations) for configurations with near and far targets at each of the three tests. In Test 1, the CC effect was greater for configurations with far targets compared to configurations with near targets, $t(62)$ = -2.12, p = .038, d_z = -0.27. There was no difference in the CC effect in Test 2, $t(62)$ = 1.79, p = .078, d_z = 0.23, but by Test 3 the difference in the size of the CC effect had reversed, with a greater CC effect observed for configurations paired with near targets compared to those paired with far targets, $t(62)$ = 3.00, p = .004, d_z = 0.38. It should be noted that this interaction effect was not supported in the results of the Bayesian analysis, reported below. No other interaction effects were significant (Fs < 1.9).

The ANOVA results suggest that the *target-distractor contingency* factor did not play a significant role in determining reaction times. To confirm this, a Bayesian ANOVA (priors were set at the default “medium” width) was conducted with factors of *configuration type*, *target-depth*, and *target-distractor contingency*. The model which produced the largest Bayes Factor was that containing only factors of *configuration type* and *target-depth* with no interaction effects, BF_{10} = 2.8x10ˆ82. The comparison of this model with that containing an interaction between these two factors revealed support for the null, BF = 0.2. Considering each of the models containing only an individual factor, there was considerable support for the model containing just *configuration type*, BF_{10} = 93.2, and for the model containing just *target-depth*, BF_{10} = 2.1x10ˆ78. In contrast, there was moderate support for the null when considering the model containing just the factor of *target-distractor contingency*, BF_{10} = 0.2.

We next explored the possible interaction effects involving *target-distractor contingency* (i.e., a comparison of the model with the interaction to the model without the interaction). There was support for the absence of an interaction effect between *target-distractor contingency* and *configuration type*, BF = 0.2, and also between *target-distractor contingency* and *target-depth*, BF = 0.1. There was also support for the absence of a three-way interaction effect, BF = 0.2.

An additional Bayesian ANOVA (priors were set at the default “medium” width) was run with factors of *configuration type*, *target-depth*, and *test* to explore the three-way interaction that was significant in the ANOVA reported earlier. This revealed a Bayes Factor for the interaction that did not amount to supporting evidence, BF = 3.1.

Analysing the egocentricity of learning

So far our analysis has focused on the *allocentric* nature of the contingencies: whether the position of the repeating distractors in respect of the target-depth affects the

CC that develops. However, our manipulation of the distractor contingencies, with only one surface containing repeated distractors (and the other random distractors) means that we can also assess the CC effect with respect to the position of the observer. To do this, we recoded the factors of distractor-target contingency and target depth as a single factor of egocentricity: egocentric-near (proximal contingencies with near targets; distal contingencies with far targets) or egocentric-far (proximal contingencies with far targets; distal contingencies with near targets). These data are presented in Figure 7.

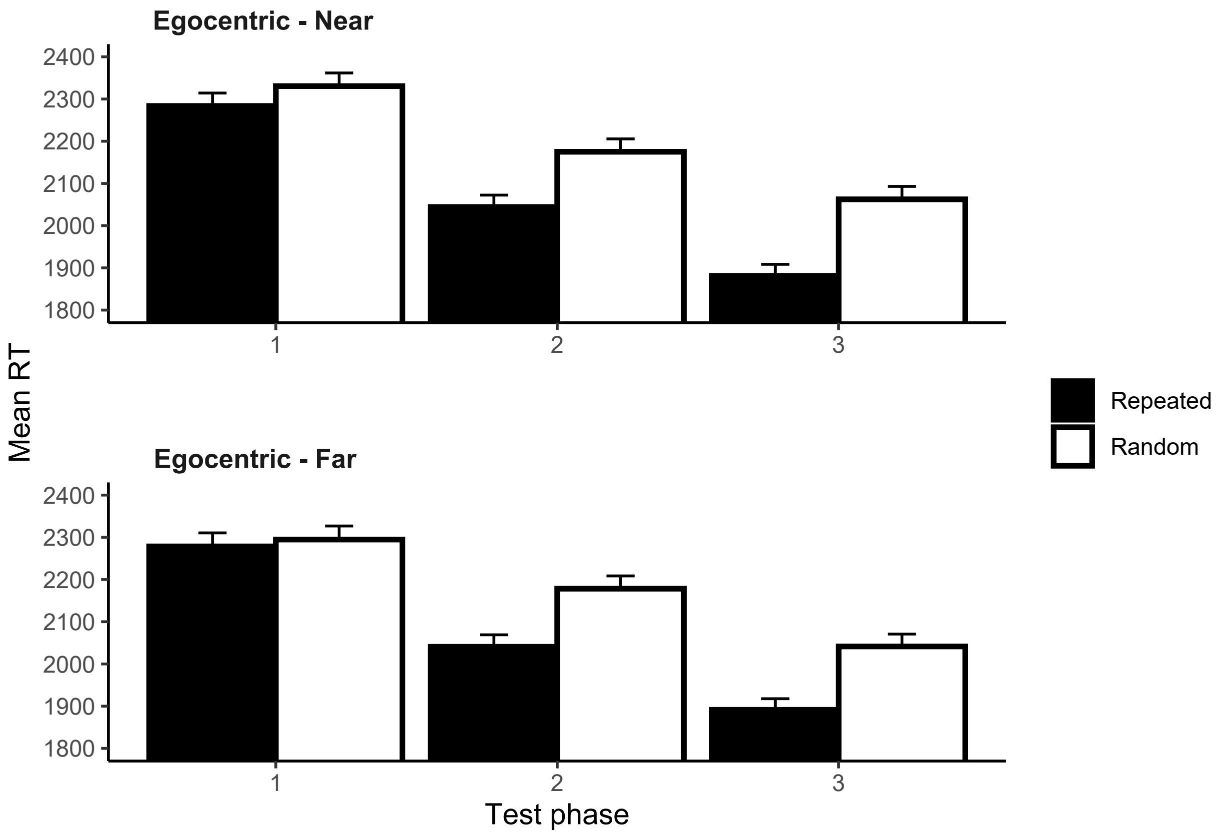

Figure 7

RTs across the tests in Experiment 2, as a function of egocentricity of the repeated configurations (near to the observer or far from the observer).

These data were subjected to a three-way repeated-measures ANOVA with factors of *egocentricity*, *configuration type* and *test*, which found no main effect of egocentricity, and egocentricity did not interact with the other factors (Fs < 1). To confirm the CC effect in both conditions, separate ANOVAs were run on the near and far egocentric conditions. This revealed main effects of configuration type and test in both conditions (Fs > 15). Numerically, the size of the CC effect increased across tests in both conditions, but the interaction between configuration type and test was not significant in either the far condition, $F(1.92, 119.00) = 2.94$, $p = .059$, $\eta_p^2 = .05$, or the near condition, $F(1.87, 115.88) = 3.01$, $p = .057$, $\eta_p^2 = .05$.

Awareness test

For the awareness test, participants error scores (Euclidean distance on the x/y plane, between the correct target position and central head position in “Unity units”) on configurations with far targets was 4.50 (SD = 2.73) for repeated configurations, and 4.55 (SD = 2.75) for random configurations, while for near targets it was 3.11 (SD = 1.81) for repeated configurations, and 3.35 (SD = 1.78) for random configurations. These data were subjected to a two-way ANOVA with factors of *configuration* and *target-depth*, which revealed a main effect of target-depth $F(1, 62) = 150.20$, $p < .001$, $\eta_p^2 = .71$,[2] but no main effect of configuration, $F(1, 62) = 1.78$, $p = .187$, $\eta_p^2 = .03$ and no interaction, $F(1, 62) = 0.65$, $p = .423$, $\eta_p^2 = .01$. A Bayesian ANOVA revealed a Bayes Factor of 0.22 for the model containing only the configuration factor, suggesting reasonable support for the null model. We examined whether the size of the CC effect (RT to random configurations minus RT to repeated configurations, averaged across the three tests) was correlated with the awareness performance (error scores on random configurations minus error scores on repeated configurations). There was no evidence to suggest these scores were positively correlated, $r(61) = .03$, $p = .811$ (though see Vadillo et al., 2021 for a discussion of the pitfalls of such approaches to assessing awareness).

---

[2] The main effect of target-depth is due to the relative distances between the objects on the near surface being shorter than those on the far surface. Therefore any given decision made will be more likely to be closer to the actual target position for near surface targets

Discussion

Experiment 2 sought to examine the associative learning occurring between distractors and targets across different depths during visual search. For all participants, targets appeared on either the near or far surface equally often, and half of the distractors were repeated (and so could be learnt in relation to the target), while the other half of the distractors on each trial were randomly arranged. For half of the participants, the repeating distractors were those that were proximal to the target (same surface), and for the other half of participants the repeating distractors were those that were distal to the target (different surface). We found that the proximity of the repeating distractors to the target did not affect the size of the CC effect. As in Experiment 1, the target-depth appears to play a more important role in determining visual search and CC. As was the case in Experiment 1, targets that appear nearer to the observer were harder to detect (slower RTs). Unlike in Experiment 1, there was some evidence to suggest that the CC effect for distractors paired with near targets was larger than that for those paired with far targets (in Test 3), however, this pattern was not supported by the Bayesian analysis.

We also examined any potential effect of the ego-centricity of repeating distractors on CC, by coding the configurations as either containing repeating distractors that appeared near or far from the observer. This confirmed that ego-centricity did not affect CC and that both levels of ego-centricity resulted in CC.

An awareness test, in which participants had to indicate the location of an absent target, revealed that participants were as accurate for new configurations as they were for those that they had experienced in the main task. The procedure involved the presentation of each configuration twice (32 trials in total), which goes some way towards alleviating issues of test sensitivity, though we acknowledge that a test involving more trials would be beneficial here (c.f. Smyth & Shanks, 2008). As such, while our awareness test failed to find any evidence of conscious awareness of the repeating configurations, we acknowledge the issues with drawing such conclusions (see Vadillo et al., 2016, 2021).

General Discussion

Two experiments sought to establish a CC effect within a three-dimensional VR procedure. Experiment 1 established a moderate effect of CC (Cohen's $d_z = 0.66$) and we observed a CC effect for both configurations with targets near to the observer, and those with targets more distant from the observer. Experiment 2 sought to examine the learning of proximal and distal contingencies within the repeating configurations, by selectively pairing the target with repeated configurations on the same surface, or the alternative surface. While we found a robust CC effect overall (0.72), there was no effect of the position of the repeated distractors on the size of the CC effect. These data overwhelmingly suggest that where contingencies exist within the repeating configurations, learning will occur readily for those target-distractor contingencies.

The placement of the target at different depths within three-dimensional space had a noticeable effect on the efficiency of visual search: having a target closer to the observer made search more difficult. Interestingly, there was some evidence in our data that the target placement also modulates the extent to which a CC effect develops. Experiment 1 observed a numerically larger CC effect for configurations paired with far targets over those paired with near targets (though this was not a statistically significant difference). In contrast, in Experiment 2 we found a similar advantage was observed only in the very first part of the experiment (Test 1), whereas later in the experiment this advantage had reversed (Test 3), with stronger CC effects observed for those configurations with targets on the near surface. This latter interaction effect was not supported by the Bayesian analysis and likewise Marek and Pollmann (2020) found no difference in the CC effect between configurations paired with near and far targets, with substantial support for the null hypothesis; the effect of the target position on the development of contextual cuing will require further investigation before firm conclusions can be drawn.

The finding that depth relationships did not modulate the size of the CC effect in Experiment 2 is consistent with the results of Zang et al. (2017), where it was observed that after successful CC had been established, switching the distractors (and target) between front and back planes resulted in a maintenance of CC. Taken together, these data suggest that CC relies predominantly (and possibly exclusively) on the formation of associations within a two-dimensional representation of space. Our analysis of the ego-centricity of the contingencies illustrates this feature: distractors occurring near to the participant were learnt as readily as distractors far from the participant.

In this sense, the associative framework put forward by Brady and Chun (2007; see also Beesley et al., 2016) to account for data from two-dimensional CC tasks would not require a major revision to allow for the encoding of depth information. As it stands, the three-dimensional configuration of distractors could be recoded as a two-dimensional configuration, ignoring the distance in depth between distractors and between distractors and the target. This lack of an effect of depth contrasts with the effect of spatial proximity on the x/y plane in the standard 2D CC procedures. It has been shown that those distractors that are located closest to the target position are most dominant in driving the development of CC. When the local context surrounding the target is removed, then CC is abolished, whereas the presence of the wider, global context is not critical for the observation of CC (Olson & Chun, 2002).

Two caveats should be made to this conclusion. Firstly, Kawahara (2003) showed that reversing the depth of two surfaces within previously learnt configurations can produce a disruption in CC, but in the case of their experiment (in comparison to Zang et al., 2017), the target-depth was maintained across these reversals, and it is therefore likely that the two-dimensional relationships between distractors and targets were also disrupted. As such, this attenuation of the CC effect would be expected according to models encoding only two-dimensional representations of the configurations. The second caveat is that we observed a strong effect of target location on the efficiency of visual search, with longer search times on targets closer to the observer compared to targets appearing further away. Prima facie, this result would seem to necessitate the encoding of depth within the model to account for such effects on visual search.

Our tasks demonstrate the great potential VR procedures have for expanding the range of manipulations that can be achieved in CC (see also Marek & Pollmann, 2020), and therefore advancing our understanding of the cognitive and perceptual processes responsible for learning in visual search. Future work will provide further tests of whether the ubiquitous CC effects across depth that we have observed withstand manipulations of distractor positions, providing further data to understand the discrepancy in findings between the current data, those of Zang et al. (2017) and those of Kawahara (2003). This in turn will provide key data to revise our models of CC. Another potentially fruitful line of experimental work would involve an exploration of the role of the observer position on the development and expression of CC. In Experiment 2 we briefly considered the data in terms of the relationship to the observer (the egocentricity of the repeating distractors), but the VR technology allows much greater control over manipulations of observer position. Of note is recent work by Zheng and Pollmann (2019), that explored the rotation of configurations in a standard CC procedure. They observed that the CC effect was preserved in a final test phase in which a 45 degree rotation was made of the configuration. This suggests a certain degree of allocentric encoding of the configuration, suggesting that the expression of CC could persist after changes in viewpoint in the VR scene. Future work will explore these possible manipulations.

References

Beesley, T., Hanafi, G., Vadillo, M. A., Shanks, David. R., & Livesey, E. J. (2018). Overt attention in contextual cuing of visual search is driven by the attentional set, but not by the predictiveness of distractors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(5), 707–721. https://doi.org/10.1037/xlm0000467

Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 348–362. https://doi.org/10.1037/xlm0000033

Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2016). Configural learning in contextual cuing of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1173–1185. https://doi.org/10.1037/xhp0000185

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*(4), 798–815. https://doi.org/10.1037/0096-1523.33.4.798

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Colagiuri, B., & Livesey, E. J. (2016). Contextual cuing as a form of nonconscious learning: Theoretical and empirical analysis in large and very large samples. *Psychonomic Bulletin & Review*, *23*(6), 1996–2009. https://doi.org/10.3758/s13423-016-1063-0

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in*

Psychology, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00781

Kawahara, J. (2003). Contextual cueing in 3D layouts defined by binocular disparity. *Visual Cognition*, *10*(7), 837–852. https://doi.org/10.1080/13506280344000103

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

Marek, N., & Pollmann, S. (2020). Contextual-Cueing beyond the Initial Field of View-A Virtual Reality Experiment. *Brain Sciences*, *10*(7). https://doi.org/10.3390/brainsci10070446

McSorley, E., & Findlay, J. M. (2001). Visual search in depth. *Vision Research*, *41*(25-26), 3487–3496. https://doi.org/10.1016/s0042-6989(01)00197-3

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264–265. https://doi.org/10.1038/320264a0

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302. https://doi.org/10.1080/13506280042000162

Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403–415. https://doi.org/10.3758/MC.36.2.403

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, *23*(1), 87–102. https://doi.org/10.3758/s13423-015-0892-6

Vadillo, M. A., Malejka, S., Lee, D. Y. H., Dienes, Z., & Shanks, D. R. (2021). Raising awareness about measurement error in research on unconscious mental processes. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01923-y

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. https://doi.org/10.1167/17.5.17

Zheng, L., & Pollmann, S. (2019). The contribution of spatial position and rotated global configuration to contextual cueing. *Attention, Perception, & Psychophysics*, *81*(8), 2590–2596. https://doi.org/10.3758/s13414-019-01871-9