Using a Virtual Environment to Examine the Role of Depth in Contextual Cueing of Visual Search

Magda Lapinska

2020

Word count: 7971

A dissertation submitted to Lancaster University in partial fulfilment of the requirements of the degree of BA/BSc (Hons) in Psychology

The work submitted in this report is my own and has not been submitted in substantially the same form towards the award of another degree or other qualifactory work by myself or any other person. I confirm that acknowledgement has been made to assistance given and that all major sources have been appropriately referenced.

☑ By ticking this box I give permission for the Psychology Department to show my 304 report anonymously to students in subsequent years, in order for these students to see how good and excellent projects are structured and to learn from projects in similar areas to their own project topic. Students who see my project will not be told my name but will learn the band in which my 304 mark fell.

The research documented in this report received ethical approval on: 15.07.2018

Name: Magda Lapinska

Signature: Łapińska Magda

Date: 11/03/2020

Abstract

Contextual cueing of visual search refers to the implicit process of learning spatial associations between a target and other objects. It has been found that certain features of a visual scene can effectively guide attention to a target location. However, research on the role of depth perception in the deployment of attention in contextual cueing has yielded mixed results. To explore this issue, the present study used high quality virtual reality settings to provide a good ecological validity, while maintaining the experimental control over stimuli. During two phases of the experiment, participants searched for a capsule-shaped target among cylindrical distractors. Half of the stimuli appeared on the front and the other half on the back hemisphere. Unbeknownst to the participants, half of the presented configurations were repeated throughout the whole first phase in order to establish a contextual cueing effect. For the second phase, in half of those repeated configurations distractors from the front and back hemisphere were swapped, whereas the target remained on the original hemisphere. Analyses revealed that the contextual cueing effect was present only in configurations with far targets and it was not affected by the depth reversal. Far targets also received quicker responses than near ones. Moreover, participants showed an explicit awareness of the far repeated configurations on the recognition test, which was completed at the end of the experiment. The significance and limitations of these findings are discussed in light of previous research.

Keywords: contextual cueing; implicit learning; depth perception; virtual reality

Using a Virtual Environment to Examine the Role of Depth in Contextual Cueing of Visual Search

Imagine your typical morning: you wake up, go to the kitchen and there, you impatiently wait for water for coffee to boil. In the meantime, almost mechanically, you have found your phone to switch off the alarm, searched for your slippers, worked out the way from the bedroom to the kitchen, and looked for a clean cup. By the time you prepare your coffee, you have successfully completed at least four visual searches.

Those efficient searches are possible due to the ability of the human attentional system to prioritise relevant objects and events from the environment through a complex selection process. The mechanisms that govern this process have been examined extensively in the last few decades using a visual search task, in which participants need to find a predetermined object (the target) amongst distractors. The search difficulty may vary by changing the set size, the homogeneity of the distractors, and the similarity of the distractors to the target (Humphreys, Quinlan, & Riddoch, 1989; Palmer & Rock, 1994; Pashler, 1987). Understanding the factors that enhance search efficacy could help us in both theoretical and practical ways by creating more accurate computational models of selection and by improving transparency of public spaces (e.g. airports) and public information (e.g. news, advertisements).

What appears to play a particularly important role in guiding attention during visual search is one's past experience. Although there are many dynamic changes that occur around us, our environment is in large built upon stable spatial regularities between objects and events, in which one object predicts the location of another. The first research to demonstrate how the attentional system takes advantage of this coherency of visual scenes was done by Chun and Jiang (1998) in a paradigm called contextual cueing (CC).

In their experiment, participants searched for a rotated ‘T’-shaped target amongst rotated ‘L’-shaped distractors across a number of spatial configurations. Half of those configurations (‘old’) stayed constant throughout all blocks, while the other half (‘new’) varied. Chun and Jiang (1998) predicted that if people were sensitive to regularities occurring in a visual context, then their attention should be guided by contextual memories of the repeated configurations, leading to quicker reaction times. It was found that indeed, the search performance was enhanced for the repeated configurations compared to the new ones, thus supporting a hypothesis that spatial context can be learnt through a regular exposure to a visual scene. This phenomenon was termed ‘contextual cueing’.

Importantly, participants were not aware of the consistency of some trials, yet their performance was still better for the old configurations. Therefore, Chun and Jiang (1998) hypothesized that contextual cueing could occur implicitly. *Implicit learning* refers to a situation, in which information is acquired and retrieved beyond one’s awareness, as opposed to *explicit learning*, which is a conscious process that requires one’s active engagement (Squire, 1992). To examine whether participants were aware of the learning process, Chun and Jiang used a recognition test in which old and new configurations were presented. Participants were asked to identify which of these configurations had been displayed earlier in the visual search task. Since participants were not able to correctly recognize old configurations, Chun and Jiang advocated that CC involves implicit mechanisms.

To explore this matter further, Chun and Jiang (2003) devised another measure of awareness - a target location guessing task, in which participants were again shown repeated and novel configurations. This time however, configurations consisted only of distractors, as the target was replaced with a distractor identical to others. Participants were asked to indicate which quadrant of the display contained the disguised target, having considered the visual context presented to them. The results from this experiment were in line with previous findings as participants were not able to correctly guess the target location. Moreover, Chun and Jiang demonstrated that CC generated long-lasting effects, since participants who came back after one week still performed better on repeated than random configurations. It was also found that explicit instructions to memorize configurations, did not enhance the learning effect. Since a long-term retention of information and resistance to explicit instructions both constitute features typical for implicit learning, the authors considered it to be conclusive evidence that contextual cueing occurs beyond one's explicit awareness.

Despite many successful replications of these findings (Brockmole, Castelhano, & Henderson, 2006; Geyer, Shi,& Müller, 2010; Ogawa, Watanabe, & Yagi, 2009; Merrill, Conners, Roskos, Klinger, & Klinger, 2013), a closer look at the methodology reveals a number of gaps and shortcomings. As noted by Smyth and Shanks (2008), tests of awareness used in those studies might have been underpowered since they contained only one block of trials. By comparison, visual search tasks included approximately 20-30 blocks. Furthermore, Vadillo, Konstantinidis, and Shanks (2015) have emphasised the problem of Null Hypothesis Significance Testing (NHST) with respect to CC. They noted that a failure to reject the null hypothesis is usually interpreted as absence of awareness. However, drawing conclusions from null results is ambiguous since it merely means that that there is no significant effect consistent with the experimental hypothesis. In other words, NHST is often misunderstood or abused, resulting in flawed interpretations (Levine, Weber, Hullett, Park, & Lindsey, 2008). For all these reasons, implicit learning in CC remains controversial.

It is well acknowledged that the visual system is highly attuned to the processing of depth information. It has been shown that depth in conjunction with colour or motion facilitates target location (Nakayama & Silvermann, 1986). Some authors (Enns & Rensink, 1990a, 1990b; Holliday & Braddick, 1991) have also suggested that three-dimensional properties are captured by a pre-attentive stream in early vision. With a benefit of a preview, depth has also been found to enhance search by inhibiting distractors on an irrelevant depth plane (Dent, Braithwaite, He, & Humphreys, 2012). Collectively, the evidence reviewed here seems to suggest a pertinent role of depth in the attention deployment and information processing.

Despite these findings, most studies examining CC have focused on two-dimensional properties of visual scenes, such as colour or size (Wolfe & Horowitz, 2017). However, given the three-dimensional (3D) nature of the environment, it is likely that spatial associations between objects along the *Z*-axis (which represents depth) are also encoded in the contextual memory and contribute to guiding attention during visual search. To address this research gap, Kawahara (2003) designed a contextual cueing study, in which 3D displays created by stereo shutter goggles induced a perception of two depth planes. Participants were instructed to look for a target (rotated ‘T’) among 11 distractors (rotated ‘L’) and indicate whether it was tilted to the left or to the right. Each depth plane contained an even number of items. The appearance of the target on both depth planes was equalized and the order, in which near and far targets appeared, was unpredictable. Half of the configurations was repeated across the study, while the other half was newly generated on every trial. In the last block of trials, depth planes of distractors in the repeated configurations were reversed. Thus, distractors from the near plane appeared on the far plane, whereas those from the far plane were presented on the near plane. Critically for this study, the target remained on its original plane.

Kawahara (2003) hypothesized that the contextual cueing effect, established for the repeated configurations, would be abolished after the experimental manipulation. To verify it, reaction times from the block before and after the transition were compared. The results supported Kawahara’s expectations, since the mean search time for the switched configurations was significantly longer than for the same configurations prior to the disparity reversal. To determine whether this finding could be associated with a fatigue, which usually appears near the end of such studies, another experiment was carried out, in which a disparity of distractors stayed constant across all blocks. Since the result demonstrated that the CC effect was present in the last block, Kawahara concluded that the findings from the first experiment could be only attributed to the experimental manipulation. Moreover, it was also found that contextual learning of 3D spatial associations is based on implicit mechanisms, as participants were not able to identify repeated configurations in the recognition test. Therefore, this study supports the notion that spatial representations underlying the contextual cueing effect are of a three-dimensional nature and are acquired without one's awareness.

Nevertheless, this study has recently been called into question by Zang, Shi, Muller, and Conci (2017), who argued that maintaining the target on the original depth plane, while swapping a disparity of distractors, could have inadvertently disrupted two-dimensional associations between the target and distractors. To test it, Zang et al. administered a visual search task that was nearly identical to the one used by Kawahara (2003). A crucial change introduced by Zang et al. concerned the experimental manipulation, since it subjected entire configurations to a depth reversal, including the target. This modification led to a different result than the one reported by Kawahara, as the CC remained intact after the swap of depth planes. In the second experiment, Zang et al. used 2D displays, in which the left and right side of configurations were switched. The rest of the design stayed the same. It was found that in the block that followed the left-right swap, the CC effect was disrupted. By comparing results across these two experimental procedures, Zang et al. reached a conclusion that CC relies mainly on 2D spatial representations of the visual scene, with depth information being of secondary relevance.

A key problem with Zang, Shi, Muller, and Conci's study (2017) is that they failed to provide a reliable reason for reversing a disparity of the target. There was no evidence that in Kawahara's original design two-dimensional associations between the target and distractors were affected by moving distractors along the depth dimension. As a matter of fact, it appears that it was Zang et al.'s study that presented a flawed methodology, since their experimental manipulation left 3D spatial relations between the target and distractors the same. The target still shared one depth plane with the same distractors as before. Likewise, other distractors occupied the opposite depth plane to the target, as they did prior to the reversal. As target-distractor 3D associations remained the same, there was no reason for CC to be disrupted.

Up to now, far too little attention has been paid to the relationship between a location of the target on the sagittal plane and contextual learning. Past research on visual asymmetries have demonstrated a consistent preference towards near targets, which was manifested by a faster perceptual processing and slower disengagement from near targets, in comparison to far targets (Downing & Pinker, 1985; Finlayson & Grove, 2015). However, this factor has not been controlled for on a within-subject basis in the studies done by Kawahara (2003) and Zang, Shi, Muller, and Conci (2017).

The aim of the present study is to examine whether depth information is incorporated into spatial representations of the visual context. To this purpose, a contextual cueing paradigm will be adopted, in which half of configurations will be repeated across the first part of the experiment. In the second part, in half of the repeated configurations depth of distractors will be reversed, while the target's initial position will be maintained. As opposed to the previous research (Kawahara, 2003; Zang, Shi, Muller, & Conci, 2017), this study will use a within-subjects design, to avoid drawing comparisons across different experimental procedures and samples. A realistic perception of depth will be provided by a virtual reality environment of a high-quality.

It is expected that in the first phase, participants will learn spatial relationships between the target and distractors. The manipulation of depth in the second phase serves to disrupt these acquired target-distractor associations based on depth information, while maintaining 2D coordinates of all stimuli constant. It is predicted that if mental representations of the visual context involve depth information, then the CC effect should be reduced for switched configurations. In contrast, there should be no changes in the size of CC for non-switched configurations. On the other hand, if depth information is not encoded in the contextual memory, then the CC effect for switched and non-switched configurations should be the same. The impact of a target location on CC will be also explored. However, since this factor has been understudied in relation to CC, a direction of the difference cannot be predicted a priori. At last, a recognition test will be used to assess participants' awareness of the learning process. It is expected that participants will perform in this task at a chance level, thus supporting a view that CC relies on implicit mechanisms.

Methods

Participants

A random sample of 42 students (35 females, 7 males) was recruited via SONA system for a course credit and from among fellow students from Lancaster University and other near universities. It comprised mainly of Psychology students. All of the participants were aged between 18 and 25 (*M* = 20.33, *SD* = 1.46) and had a normal or corrected-to-normal vision based on their self-report. Before the start of the experiment, they were informed of possible side effects of the VR study and asked to stop the experiment in case of any health concerns. The study was approved by the ethics committee of the department of

Psychology at Lancaster University on 15th July 2018. An informed consent was given by all participants. Data gathered from the first two participants was not saved.

Design

The study consisted of three parts: a training, a contextual cueing paradigm, and an awareness test. The training contained 16 trials. All configurations were generated randomly and were different to those displayed in the manipulation phase. The CC paradigm was divided into two phases. In the first ‘learning’ phase, a 2 x 2 x 10 within-subject design was applied, with a factor of Target Location, which referred to a hemisphere on which the target was located (near vs. far), Configuration, which was either repeated across trials or newly generated for every trial (repeated vs. random), and Block (1-10).

The second ‘manipulation’ phase consisted of a 2x3x8 within-subject design, with a factor of Target Location (near vs. far), Configuration (random vs. non-switched vs. switched) and Block (1-8). Three types of configurations were presented: 1) random, which were newly-generated for every trial, 2) non-switched, which were identical to the repeated configurations from the first phase, 3) switched, which resembled the repeated configurations from the first phase with that difference, that the near hemisphere now featured distractors that previously appeared on the far hemisphere, while the far hemisphere now presented distractors that in the first phase were shown on the near hemisphere. While all distractors changed their depth, the target remained on its original hemisphere.

A single block contained 16 trials which in total gave 160 trials in the learning phase and 128 trials in the manipulation phase. Out of these 16 trials, eight configurations were generated randomly, whilst the other eight were repeated. In the second phase, four of these eight repeated configurations were subjected to a depth swap. The order in which all types of configurations were presented was randomized and unpredictable. Configurations were different for all participants. The effect of contextual cueing was measured by a response time (in ms) from the start of a trial until a detection of the response.

For the eight repeated and eight random configurations from the ‘learning’ phase as well as for the eight random configurations from the second phase, the target appeared an equal number of times in each of the four quadrants of the display. Also, an equal number of presentations of the target on the front and on the back sphere was ensured. In the ‘manipulation’ phase for the four switched and four non-switched configurations, a balance in the presentation of the target in each quadrant as well as on each hemisphere was maintained on a between-subjects basis. In other words, all eight repeated configurations from the first phase were subjected to a depth swap in the second phase, however, every participant was exposed to a different pattern of target position across quadrants and depth spheres in the four switched and four non-switched configurations.

The awareness test spanned two blocks of 16 trials each. Participants were presented with eight original repeated configurations from the ‘learning’ phase and with the same number of novel configurations. The performance was measured based on the ratio of hits (correctly recognized old configurations) and false alarms (novel configurations incorrectly classified as old). The use of repeated configurations from the first phase was motivated by an intention to compare results with previous studies that used this design (Kawahara, 2003; Zang, Shi, Muller, & Conci, 2017). Moreover, since the first phase was longer than the second, it was expected that the memory of the repeated configurations prior to the experimental manipulation should be stronger than to the switched configurations.

Materials

The study was conducted on a standard desktop computer. A Unity software was used to program the experiment. An Oculus Rift CV1 headset delivered the experiment in virtual reality, while the Oculus sensor detected movements of the headset. Stimuli were positioned at two different spherical surfaces that were concentric to each other, projecting away from the viewer (who was situated at the centre). The frontal hemisphere, which was closer to the observer, presented ‘near stimuli’, whereas the back hemisphere displayed ‘far stimuli’. The stimuli consisted of 16 red cylindrical distractors, spread evenly across both hemispheres, and one red capsule-shaped target. The background was of a dark gray colour. Practice trials featured a semi-transparent blue disc which followed participants’ head movements. Appendix A provides a 2D illustration of the arrangement and visual features of this display.

The headset was adjustable so eyeglasses could comfortably fit underneath. A spacebar served as a response key in the visual search task. In the awareness test, left mouse button registered responses for configurations regarded as old, while the right mouse button for configurations considered to be novel. At the start of each trial, a semi-transparent fixation sphere was presented in the centre of the display and remained there, until the participant looked directly at it.

Procedure

Participants were seated around 50 cm away from the Oculus sensor, with a keyboard in front of them. The experimenter assisted in setting up the Oculus Rift headset and helped participants with finding the spacebar. Before the experiment started, it was ensured that participants could clearly see in the headset and that their head movements waere detected by the system. When needed, the focus of the headset was adjusted appropriately to participant’s needs.

Participants were instructed to search for a red pill-shaped target amongst red cylinder-shaped distractors in a 3D setting. Firstly, they familiarized themselves with the task in a training session. If they felt ready, they could proceed to the main part of the experiment.

In cases where participants clearly struggled to understand the demands of the task, they were given another chance to undergo the training. All instructions were displayed in front of them in the VR so participants could read them at their own pace, without a need to remove the headset. Before the study began, participants were prompted by the experimenter to ask further questions in case of confusion.

For a response to be detected during visual search, participants had to press the spacebar while looking directly at the target. Head movements were required to examine the configurations thoroughly since some of the stimuli appeared on the periphery of the vision. When a response was approved, the most recent configuration disappeared, and a semi-transparent white fixation sphere emerged in the centre of the display. A new trial was presented only when a participant directed their gaze there. This measure was used to ensure that all participants focused on the same point of the display at the start of every trial. Reaction times were recorded from the beginning of a trial until a response detection.

Participants were not informed that the visual search task was made up of two phases and furthermore, the transition between these phases was not apparent. There were two breaks placed in the learning phase and another two in the manipulation phase. Participants decided themselves whether they wished to rest or to continue. If a participant took the headset off during a break, it was ensured that the focus of the headset was reassessed and, when needed, readjusted before the start of the next trial.

Participants were not informed that half of the configurations would be repeated throughout the experiment. Their explicit awareness was examined through a recognition test (Chun & Jiang, 1998) on newly generated random configurations and configurations which were repeated in the first phase. The reason for this was that it was assumed that the CC effect would be stronger for the original, non-switched repeated configurations than for the switched configurations, due to the primacy effect and a higher number of blocks in the learning phase. Participants were asked to press a left mouse button if a configuration was familiar and a right mouse button, if it was novel. They were instructed to respond as quickly as they could. The whole experiment lasted around 40 minutes.

Results

Trials with timeouts and with more than one extra response were analysed to test whether the amount of difficulties that participants experienced was not substantially different across the factor of target location. On trials with more than one additional response, one high-end extreme outlier with a value of 65.28 was removed. A paired samples *t*-test showed that there was no significant difference between near target trials (M = 12.96, SD = 8.90) and far target trials (M = 15.83, SD = 12.19), $t(35) = 1.72$, $p = .10$. On trials with timeouts, one outlier with a value of 13.19 was identified and removed upon a visual inspection of the data. Results revealed that participants experienced significantly more timeouts on configurations with far targets (M = 2.97, SD = 2.98) than with near targets (M = 1.54, SD = 2.13), $t(35) = 4.67$, $p < .001$. A likely explanation for this is that stimuli appearing on the near hemisphere tended to occlude far targets on some of the trials.

Following this examination, all trials with timeouts and with more than one extra response were removed from the dataset. Three participants were removed from the analysis as outliers since their data fell 1.5 interquartile ranges above the third quartile. The mean reaction time (RT) of the remaining 37 participants was 2591.65 ms. The proportion of trials with one additional response equalled 23.13%. In total, 82.83% of trials contributed to the final analysis.

Learning Phase

To improve statistical power, each two consecutive blocks from the visual search task were collapsed together, forming five epochs in the 'learning' phase. A distribution of data across epochs for all configuration types, presented in Figure 1, demonstrates an overall decrease in mean response times. To test whether this effect was significant, the within-subject factors of Configuration (repeated vs. random), Epoch (1-5), and Target Location (near vs. far) were subjected to a three-way repeated-measures ANOVA.

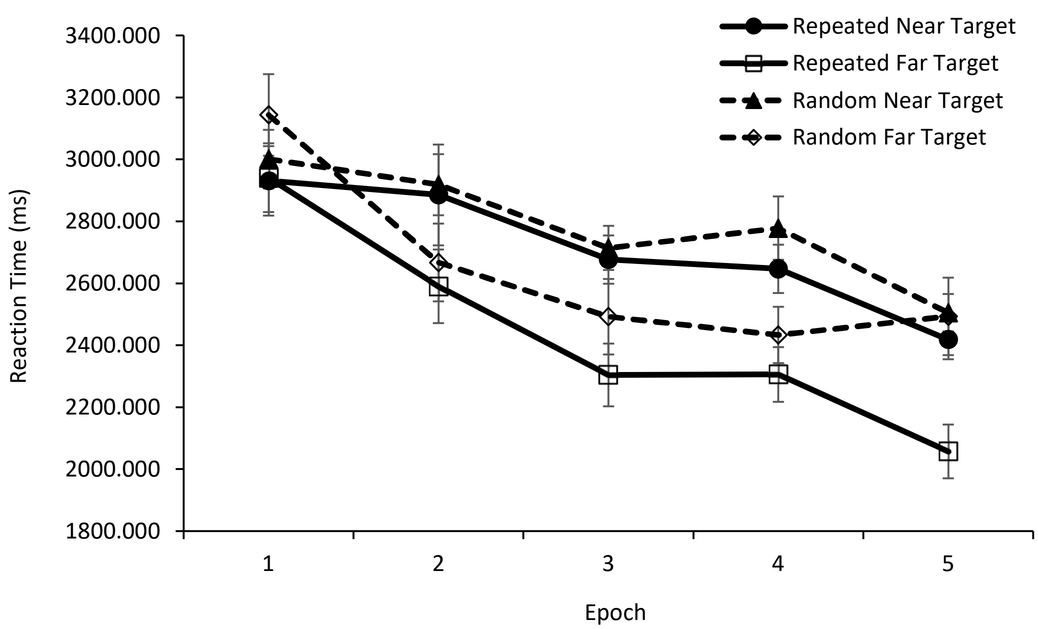

Figure 1. Changes in mean reaction times to near-repeated, far-repeated, near-random, and far-random configurations over the course of five epochs of the 'learning' phase. Error bars represent standard error of the means.

The analysis revealed a significant main effect of epoch, $F(4, 140) = 23.90$, $p < .001$, $\eta_p^2 = .41$, suggesting that an average response time tended to decrease over time. Bonferroni posthoc tests showed that RTs were significantly faster in in the fifth epoch ($M = 2368.64$, $SE = 57.13$) compared to the first ($M = 3004.32$, $SE = 78.80$) and the second epoch ($M = 2765.56$, $SE = 89.87$), both $p < .001$. Moreover, participants responded faster to repeated ($M = 2575.78$, $SE = 59.40$) than random configurations ($M = 2714.81$, $SE = 54.57$), demonstrating a contextual cueing effect, $F(1,35) = 11.60$, $p = .002$, $\eta_p^2 = .25$. Results also indicated that participants performed quicker when the target was positioned on the far hemisphere ($M = 2543.00$, $SE = 70.27$) compared to when it appeared on the near hemisphere ($M = 2747.60$, $SE = 56.58$), yielding a significant main effect of target location on response times, $F(1, 35) = 8.49$, $p = .006$, $\eta_p^2 = .20$.

A follow-up paired-samples *t*-test revealed that a contextual cueing effect occurred in configurations with far targets, $t(36) = -3.08$, $p = .004$, 95% CI [-341.52, -70.28] but not with near targets, $t(36) = -1.27$, $p = .21$, 95% CI [-169.32, 38.81]. As depicted in Figure 2, while there was a considerable improvement in reaction times for far-repeated configurations compared to far-random, there was no difference in participants' performance between near-repeated and near-random configurations.

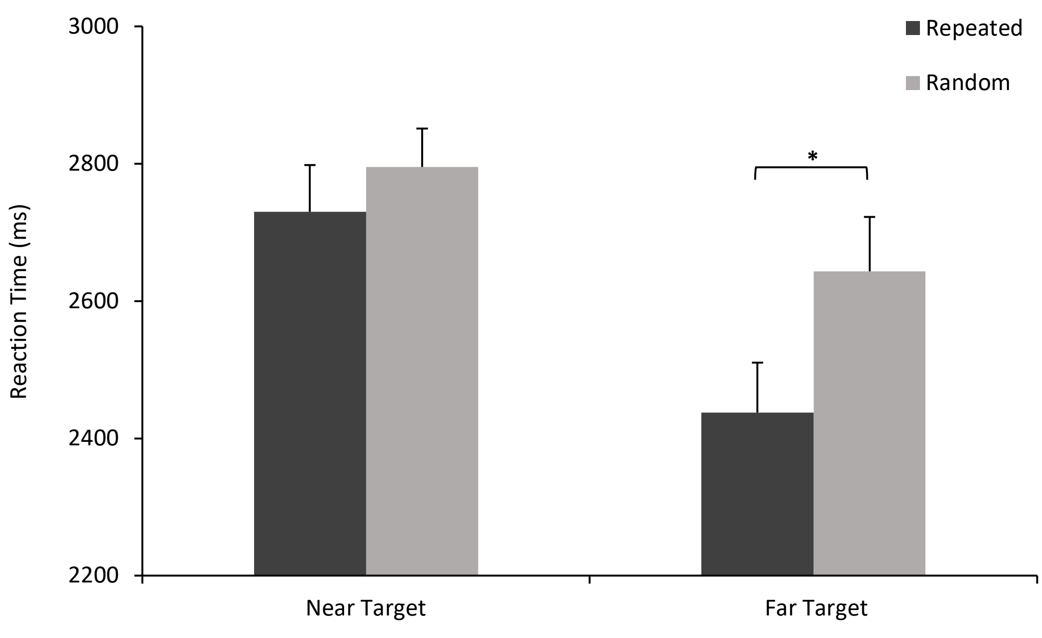

Figure 2. Mean reaction times in the 'learning' phase to repeated and random contexts for configurations with near and far targets. * $p < .05$. Error bars represent standard error of the mean.

Returning to the main three-way ANOVA of the 'learning' phase, for the interaction of epoch by target location, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi2(9) = 18.21$, $p = .03$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .81$). The results showed that there was a significant interaction between epoch and target location, $F(3.59, 125.57) = 3.89$, $p = .007$, $\eta_p^2 = .10$. A test of linear contrast indicated that the average performance for near- and far-target configurations had significant linear, $F(1, 35) = 6.14$, $p = .02$, $\eta_p^2 = .15$, and quadratic components, $F(1, 35) = 10.80$, $p = .002$, $\eta_p^2 = .24$.

Subsequent analyses of the simple main effects showed a significant effect of epoch on near target location, $F(3.33, 119.82) = 9.37$, $p < .001$, $\eta_p^2 = .21$, and on far target location, $F(4, 144) = 20.53$, $p < .001$, $\eta_p^2 = .36$, with RTs for far targets being significantly faster for the second compared to the first epoch ($p < .001$) but similar between consecutive epochs for near targets ($p \geq .07$). Furthermore, a significant difference between far- and near-target locations was found in the epoch two ($F(1, 36) = 4.71$, $p = .04$, $\eta_p^2 = .12$), three ($F(1, 36) = 10.64$, $p = .002$, $\eta_p^2 = .23$), four ($F(1, 36) = 16.76$, $p < .001$, $\eta_p^2 = .32$), and five ($F(1, 36) = 6.02$, $p = .02$, $\eta_p^2 = .14$). The mean RTs in these epochs indicate that participants responded faster for far targets than for near targets. Therefore, factors of epoch and target location both seem to cause the interaction.

The interaction of epoch and context was non-significant, $F(4, 140) = 0.94$, $p = .44$, $\eta_p^2 = .03$, suggesting that the contextual cueing effect remained significant across epochs. The interaction between context and target location, $F(1, 35) = 2.19$, $p = .15$, $\eta_p^2 = .06$. and the three-way interaction, $F(4, 140) = 0.77$, $p = .55$, $\eta_p^2 = .02$, all yielded null results.

Manipulation Phase

As in the first part of the analysis, each two consecutive blocks from the 'manipulation' phase were merged into epochs. Figure 3 illustrates mean RTs across all conditions, showing an overall decreasing trend for configurations with far targets but no consistent pattern for configurations with near targets. A three-way repeated-measures ANOVA with within-subject factors of Configuration (repeated non-switched vs. repeated switched vs. random configurations), Epoch (1-4), and Target Location (near vs. far) revealed a significant main effect of epoch, $F(3, 90) = 6.48$, $p = .001$, $\eta_p^2 = .18$. Subsequent Bonferroni posthoc tests demonstrated that mean RTs were significantly higher in the first epoch (M = 2330.67, SE = 57.17) than in the fourth (M = 2089.09, SE = 51.69), $p = .001$, indicating an occurrence of a procedural learning effect.

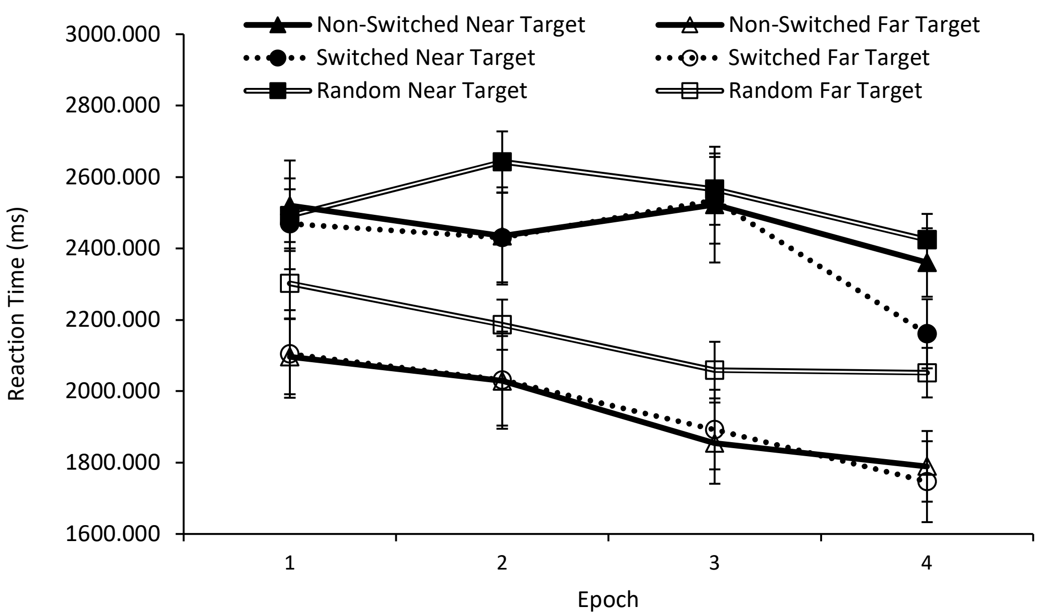

Figure 3. Changes in mean reaction times to switched near, non-switched near, random near, switched far, non-switched far, and random far configurations over the course of four epochs of the 'manipulation' phase. Error bars represent standard error of the means.

For the main effect of context, Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 7.93$, $p = .02$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.81$). A significant effect of context was found, $F(1.69, 50.76) = 3.79$, $p = .04$, $\eta_p^2 = .11$, suggesting that contextual cueing effect occurred in this phase as well. Follow-up Bonferroni tests revealed that participants responded significantly more quickly to switched repeated configurations than to random configurations ($p = .01$). However, the predicted difference in the performance time between non-switched repeated configurations and random configurations was not significant ($p = .08$). Furthermore, no evidence was found for a significant difference between non-switched and switched repeated configurations ($p > .999$).

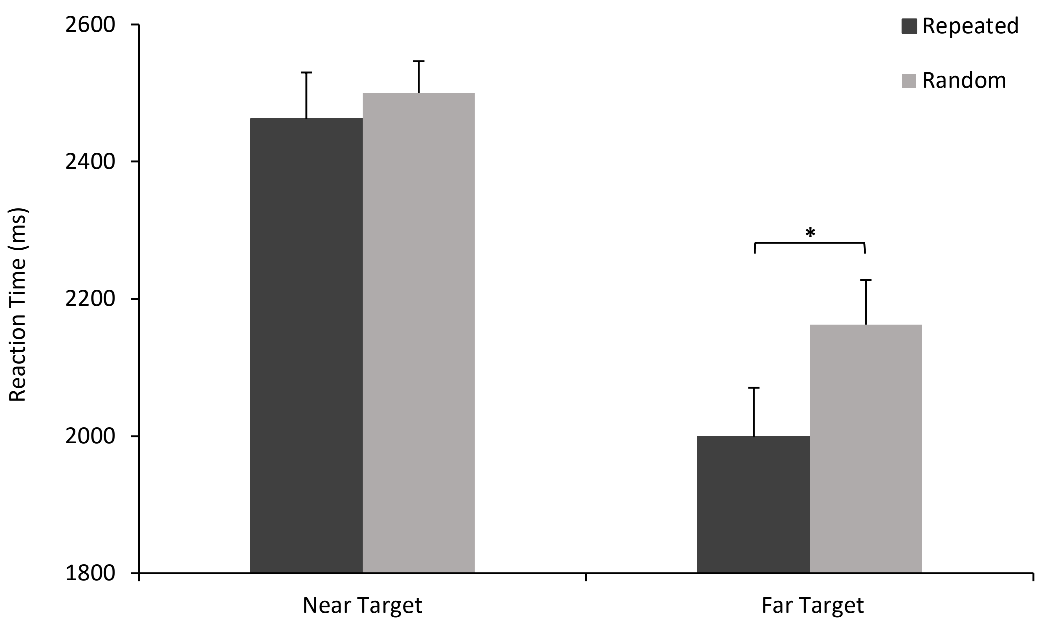

Figure 4. Mean reaction times in the 'manipulation' phase to repeated and random contexts for configurations with near and far targets. * $p < .05$. Error bars represent standard error of the mean.

As perceived from Figure 4, mean reaction times showed a similar pattern as in the first phase, suggesting an effect of contextual cueing for far but not for near targets. A paired-samples *t*-test demonstrated that participants indeed responded faster for far-repeated than far-random configurations, $t(36) = -3.04$, $p = .004$, 95% CI [-273.72, -54.76]. Nonetheless, they did not show any signs of context learning for configurations with near targets, $t(36) = -0.63$, $p = .54$, 95% CI [-161.72, 85.47].

Returning to the main three-way ANOVA, there was a significant main effect of target, $F(1, 30) = 53.72$, $p < .001$, $\eta_p^2 = .64$, with participants detecting targets on the far hemisphere faster ($M = 2012.13$, $SE = 49.68$) than on the near hemisphere ($M = 2463.20$, $SE = 57.00$). All the interactions generated non-significant results, all $F \leq 0.40$, $p \geq .06$, $\eta_p^2 \leq .01$.

Awareness Test

Responses to the awareness test were lost for one person. Therefore, data from 36 participants contributed to the final analysis of the awareness test. Figure 5 illustrates how rates of hits and false alarms were distributed based on a target location. Participants correctly recognized repeated configurations (‘hit rate’; $M = 0.50$, $SD = 0.19$) significantly more often than they incorrectly classified novel configurations as familiar (‘false alarm rate’; $M = 0.38$, $SD = 0.15$), $t(35) = 3.00$, $p = .005$, 95% CI [0.04, 0.21]. This pattern of results indicates that participants were aware of the repeated far target configurations.

Since participants did not demonstrate a contextual cueing effect for configurations with near targets, it was expected that they would not be able to correctly identify near-repeated configurations and hence, there would be no difference between hit and false alarm rates. However, the analysis of near-target configurations showed that participants mistook novel configurations for repeating ($M = 0.50$, $SD = 0.18$) more frequently than they recognised old configurations (*M* = 0.39, *SD* = 0.19), *t*(35) = -2.69, *p* = .01, 95% CI [-0.21, -0.03].

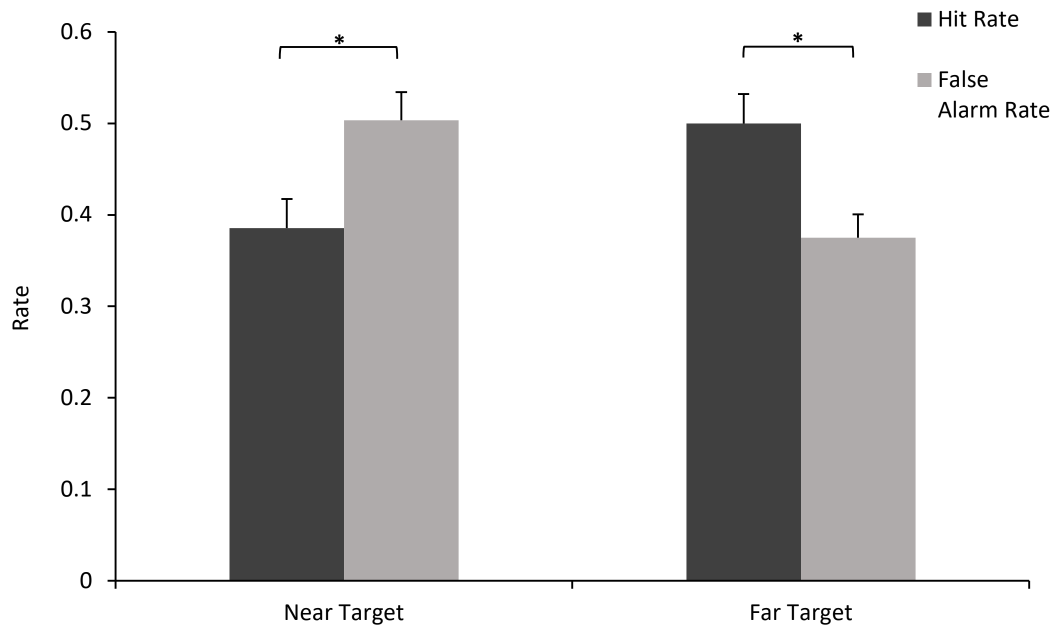

Figure 5. Hit and false alarm rates grouped by target location. The results were obtained from a recognition test in which participants identified displayed configurations as familiar or novel. * *p* < .05 Error bars show standard error of the means.

Discussion

The present study set out with the aim of assessing by means of a contextual cueing paradigm whether depth information is encoded in the contextual memory. Prior literature on this matter reported conflicting results. Kawahara (2003) documented that depth acts as a cue for guiding attention in CC of visual search, whilst Zang, Shi, Muller, and Conci (2017) argued that two-dimensional spatial associations are more relevant than depth information. To address this issue in the current experiment, a complex virtual environment was used in order to evoke a realistic depth perception. Three major findings emerged from this study.

It was hypothesized that the contextual cueing effect, established in the first phase for the repeated configurations, would be maintained for the non-switched configurations in the second phase. In turn, the depth reversal was expected to disrupt three-dimensional spatial associations between the target and distractors, and as a consequence, abolish the CC effect for the switched-repeated configurations. There were no set predictions with respect to a direction of the effect of target location, due to a scarce past research on the relation between target depth and CC. The results showed that there was no difference in RTs for switched and non-switched configurations (Figure 3), which indicates that the experimental manipulation of depth did not alter the CC effect. Notably, additional analyses revealed that CC was present in both phases only in configurations with far targets (Figure 2 and 4). Mean search time for combined far-repeated configurations in the second phase was significantly quicker than for random far configurations. Moreover, it was also found that in both phases, it was easier for participants to find far targets than near ones, as evidenced by shorter reaction times to the former.

Regarding the non-significant effect of the experimental manipulation, the results support the findings of Zang Shi, Muller, and Conci (2017) in so far as they reveal that reversing disparity of distractors is not sufficient to abolish the CC effect (note however, that in the current experiment this observation refers only to configurations with far targets). It is noteworthy that these two studies reached a similar conclusion despite considerable differences in their experimental designs, of which the most crucial was the target position. Whilst Zang et al. swapped entire front and back configurations, including the target, the present study maintained the target's depth invariant and switched only the disparity of distractors. In this way, the three-dimensional information of target-distractor associations became disrupted, while two-dimensional spatial associations remained intact.

Possibly, this study brought different results than Kawahara’s (2003) due to substantial methodological modifications. Unlike the past research (Kawahara, 2003; Zang Shi, Muller, & Conci, 2017), this study used a within-subject design, in which all participants experienced the same type of manipulation. In this way, random noise caused by individual differences was minimized. Moreover, subjecting half of the repeated configurations to the depth reversal allowed for a comparison of performance to switched and non-switched configurations within the same epochs. This measure enhanced reliability of results since fatigue and concentration levels were the same for all types of configurations. Additionally, the Oculus Rift headset provided both a high ecological validity through an immersive experience of a virtual reality and a high experimental control over presented stimuli.

Nonetheless, the absence of effect of the depth manipulation does not necessarily mean that 3D spatial associations are irrelevant for CC. Present results could indicate that significance of depth information could be modulated by experimental conditions. In studies done by Kawahara (2003) and Zang Shi, Muller, and Conci (2017), each search set contained 12 items. By contrast, there were 17 items per trial in this study. It can thus be suggested that depth-defined information was not incorporated in the contextual memory due to a high perceptual demand required to process a large search set. According to load theory (Lavie, 1995), when a perceptual demand of a task is high, attentional capacity is exhausted by selecting relevant information, thus leaving less attentional resources available for processing information about distractors.

Following this line of reasoning, the intact CC effect for switched configurations perhaps indicates that attentional resources, under high perceptual load, were not allocated to process a disparity of distractors. Hence, it could conceivably be hypothesized that 3D target-distractor associations are of a lesser relevance than a two-dimensional spatial information since only the latter was prioritised at an early stage of perceptual processing. This supposition ties well with previous literature which demonstrated that there is an asymmetric interaction between two- and three-dimensional spatial representations, with 2D location information exerting a stronger influence on the depth perception than in reverse (Finlayson & Golomb, 2017b; Finlayson & Golomb, 2016).

Crucially to the proposed explanation, Finlayson, Remington, Retell, and Grove, (2013) found that the influence of depth information can indeed be altered via varying a level of perceptual load of the visual search task. Specifically, when a non-target plane required a conjunction search (i.e. distractors resembled the target on one feature but differed on the other) but the target plane involved a pop-out search (i.e. distractors did not share any features with the target), then participants used a depth segmentation strategy, leaving the non-target plane unattended. However, when both planes contained distractors that exhibited a common feature with the target, no effect of depth segmentation was demonstrated. This finding suggests that under conditions of a high perceptual load, depth cue is regarded as target-irrelevant and filtered out early in the visual processing.

Thus, Finlayson, Remington, Retell, and Grove's study (2013) provides some support for the premise that the high rate of stimuli per trial in this experiment could bring about a heightened perceptual demand, and thereby limited attentional resources necessary for processing information about depth of distractors. As a result, even with a reversed disparity, the CC effect to switched configurations did not differ from those for non-switched configurations, because they all were processed only on the basis of their 2D target-distractor associations. Therefore, this finding indicates that depth does not capture attention in the same automatic fashion as preattentive basic features do. Importantly, this account does not imply that depth information does not contribute to an effective attentional deployment in CC. Instead, it suggests that depth seems to be a second-order property, which is ignored when perceptual load is high, but enhances visual search on tasks of low perceptual load (Finlayson et al., 2013).

With respect to the bias found towards far targets (evidenced by shorter RTs and emergence of the CC effect), it stands in contrast with previous results in the literature, since a majority of them reported a preference for near targets (Chen, Weidner, Vossel, Weiss, & Fink, 2012; Downing & Pinker, 1985). Gawryszewski, Riggio, Rizzolatti, and Umilta (1987) examined the movement of attention in depth by means of a spatial cueing task, in which participants had to detect the presence of a target. The target was displayed in the same 2D location, however, at various distances from the observer. Its position on the sagittal plane was indicated by a centrally-presented cue. Detection time was faster for targets presented near the observer than for those that were farther away. Moreover, redirecting attention from a far to a near depth plane was performed quicker than vice-versa. Plewan and Rinkenauer (2017) found that, in addition to being faster, responses towards close targets were also more powerful, as measured by a force-sensitive key device.

There are several potential explanations of why the current study presented results opposite to the cited above. The first one attributes slower RTs and the lack of CC effect for near targets to head movements, which were required to explore the near hemisphere. To be specific, search on the far hemisphere also involved head movements, however, to a much lesser extent. Therefore, the time taken to make these additional movements could significantly extend search time for near targets. Moreover, since movements of the headset were followed by a sensor placed in front of participants, any significant changes in the head position might have disrupted the connection between these two devices, leading to technical difficulties with a response detection. If this hypothesis holds true, it could be supposed that on near-target trials, participants repeatedly pressed the response button to get their responses approved and also, experienced more timeouts. Nonetheless, the analysis of trials with timeouts and trials with more than one extra response did not demonstrate any bias for near targets. This indicates that the present account is unlikely or at least, insufficient to fully explain the asymmetry found between near and far targets.

Another explanation is based on the uneven distribution of presented stimuli. Since the target constituted an extra item, it could be speculated that a hemisphere containing it attracted participants' attention more than the other hemisphere due to a higher and/or odd number of stimuli. This could be driven by a bottom-up attentional mechanism, which was captivated by a higher/odd number of displayed stimuli, or by a top-down search strategy, in which participants intentionally focused only on a hemisphere with more/odd number of items (Zoest & Donk, 2004). Interestingly however, if either of these attentional processes was engaged, it was apparent only on trials with far targets. Perhaps a higher/odd number of stimuli was more noticeable from a farther perspective, leading to a response bias in favour of configurations with targets on the far hemisphere. Unfortunately, this confound was not considered before the study and therefore, it cannot be ruled out that the effect of target and the absence of CC for near targets occurred solely due to this imbalance.

Another likely explanation derives from attentional research on near and far space. Previc (1990) made a distinction between the space within one's reach (near/peripersonal) and the area beyond it (far/extrapersonal). A large body of research supported this classification by evidencing that peri- and extrapersonal representations are indeed processed separately (Vuilleumier, Valenza, Mayer, Reverdin, & Landis, 1998; Couyoumdjian, Di Nocera, & Ferlazzo, 2003; Park & Reed, 2015). Thus, Plewan and Rinkenauer (2017) suggested that the bias for near targets, found across various studies, could result from a perceptual prioritisation of the space near the body.

Nevertheless, this account cannot be adopted to interpret a rather different pattern of results demonstrated by Maringelli, McCarthy, Steed, Slater, and Umilta (2001). In their study, half of participants were exposed to a virtual body, while the other half was assigned to a body-absent group. The task was to detect a pre-cued target as quickly as possible. The findings revealed that the body-present condition performed faster on trials with near targets, while the body-absent condition - on trials with far targets. The authors hypothesized that these two conditions relied on distinct spatial frames of reference, bringing about different strategies of attentional deployment.

A frame of reference (also known as spatial coding system) refers to an object in relation to which measurements regarding distance, position and size are estimated. Spatial representations can be framed in line with an egocentric (viewer-centred) or allocentric (environment-centred) perspective. The former uses the body's position as the central point in the space, whereas the latter refers to an external anchor point, which is independent of one's own body. Therefore, egocentric representations mainly involve peripersonal space (although they are not limited to it), while allocentric representations predominate in extrapersonal space (Berti, Smania, & Allport, 2001; Weiss et al., 2000). Based on this, Maringelli, McCarthy, Steed, Slater, and Umilta (2001) concluded that a virtual body elicited a viewer-centred spatial coding, whereas the lack of a visual body feedback forced participants to use an environment-centred frame of reference.

Therefore, a reversed pattern of bias in the current experiment could be a result of certain properties of the design, which restricted spatial representations to allocentric coding, known for favouring information from the far space. As demonstrated by Maringelli et al. (2001), the absence of body could heavily contribute to this effect. Furthermore, previous research (Woodin & Allport, 1998) revealed that large-scale layouts may trigger allocentric coding due to their resemblance to the real spatial environment. Since the virtual environment of the current design was not confined (see Appendix A), participants might have perceived it as limitless. Also, the fact that participants were seated and could make movements in the VR only with their heads could have an impact on the frame of reference, as Kimura, Miura and

Yamamoto (2009) reported that static conditions restrain the effect of egocentric coding on a deployment of attention when compared to self-motion. Therefore, in the future, studies should consider all these factors before making a decision about the experimental design of a CC task in VR, to be able to control for a frame of reference.

The last set of findings concerns the results of the recognition test. Contrary to the hypothesis that contextual learning would occur in an implicit manner, participants seemed to be aware of configurations containing targets on the far hemisphere, as demonstrated by a significantly higher rate of recognized configurations than false alarms. This is inconsistent with past research (Chun & Jiang, 1998; Chun & Jiang, 2003; Kawahara, 2003; Zang, Shi, Muller, & Conci, 2017), which showed that the contextual information is encoded incidentally. In fact, it supports an alternative stance represented by Smyth and Shanks (2008), who argued that explicit learning could not be detected in previous studies due to a low power of recognition tests used. Therefore, extending the recognition test in the current study to two blocks could have increased the statistical power and consequently, revealed participants' awareness of the far-repeated configurations. On the other hand, it could be argued that the realism of the virtual environment could have changed the way participants attended to and learnt the configurations. The high ecological validity provided by a complex virtual display perhaps enhanced participants' learning and explicit memories (Marek & Pollmann, 2019). However, this interpretation needs to be treated with caution since these are still early days for research of CC in a VR, and it remains unknown how such an interactive environment could potentially affect attentional as well as perceptual mechanisms.

By contrast, for near-target configurations the rate of false alarms was reliably higher than the rate of hits. Since there was no effect of CC for near targets, no difference between hits and false alarms was expected to be found. When compared with Zang et al.'s study (2017), it can be noticed that their participants were instructed that only half of configurations presented in the recognition test would be old. Since in the current recognition test, the instructions did not include such an information, participants might have believed that a number of familiar configurations was higher. On the other hand, this result might also indicate that a type II error occurred. To resolve this issue, future studies will need to try to replicate this result.

Regarding the limitations of this study, it could be argued that halving the number of far-repeated trials in the second phase reduced the power of the sample. Therefore, it is possible that while there was a difference in RTs between switched and non-switched trials, it could not be detected due to a low statistical power. A major source of concern, as explained earlier, comes from the fact that the target constituted an odd item. It could be a confound since a higher and/or odd number of items might have been more evident on the far hemisphere than on the near. Furthermore, a frame of reference was not controlled for, which means that some features of the design could favour one frame of reference, consequently altering the ways that attention was deployed. It is also noteworthy that a position of the fixation sphere on the sagittal plane was not considered before the study. Since it has been shown that a location of the fixation point can change a frame of reference (Andersen & Kramer, 1993), it could have affected the results of this study as well. Also, the fact that stimuli from the near hemisphere could occlude far targets might have increased RTs to far configurations. Nevertheless, this feature of the design could not be fully controlled for, since it depended on such factors as participants' height or the angle they sat at in relation to the sensor.

The broad implication of the present research is that it appears that depth information about inter-item associations is not encoded in the contextual memory when perceptual load of a task is high. Therefore, it is suggested that perceptual demands modulate processing of 3D visual cues. In future work, investigating whether a depth reversal of distractors could negatively affect CC under low perceptual conditions, might validate this assumption. In addition, this study revealed a strong effect of the 3D target location on search performance. Contradictory to the results of previous research (Downing & Pinker, 1985), participants demonstrated a preference for far targets. Although this effect could be attributed to the uneven number of displayed stimuli, it can also be suggested that certain features of the design framed participants' spatial representations in an allocentric fashion. Since the allocentric coding system processes information from the far space more efficiently, it might have given rise to the bias for far targets. Therefore, it is recommended that future studies use 3D gaze-eye tracking to see how different frames of reference alter distribution of attention across near and far space. Importantly, the findings provide evidence that CC in a 3D visual search task relies on explicit learning mechanisms. It has been argued that this result could be related to a realistic and immersive experience delivered by the VR headset. Future research should certainly further test whether the mechanisms underlying creation of contextual memories differ between a traditional and a VR study.

Word count (excluding figures, references, tables, abstract): 7971

References

Andersen, G. J., & Kramer, A. F. (1993). Limits of focused attention in three-dimensional space. Perception & Psychophysics, 53(6), 658-667.

Berti, A., Smania, N., & Allport, A. (2001). Coding of far and near space in neglect patients. *Neuroimage*, *14*(1), S98-S102.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 699.

Chen, Q., Weidner, R., Vossel, S., Weiss, P. H., & Fink, G. R. (2012). Neural mechanisms of attentional reorienting in three-dimensional space. *Journal of Neuroscience*, *32*(39), 13352-13362.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, *36*(1), 28-71.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 224.

Couyoumdjian, A., Di Nocera, F., & Ferlazzo, F. (2003). Functional representation of 3D space in endogenous attention shifts. *The Quarterly Journal of Experimental Psychology Section A*, *56*(1), 155-183.

Dent, K., Braithwaite, J. J., He, X., & Humphreys, G. W. (2012). Integrating space and time in visual search: How the preview benefit is modulated by stereoscopic depth. *Vision research*, *65*, 45-61.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. P. O. Martin (Ed.), *Attention and Performance XI* (pp. 171–187). Hillsdale: Erlbaum.

Enns, J. T., & Rensink, R. A. (1990a). Influence of scene-based properties on visual search. *Science*, *247*(4943), 721-723.

Enns, J. T., & Rensink, R. A. (1990b). Sensitivity to three-dimensional orientation in visual search. *Psychological Science*, *1*(5), 323-326.

Finlayson, N. J., & Golomb, J. D. (2016). Feature-location binding in 3D: Feature judgments are biased by 2D location but not position-in-depth. Vision research, 127, 49-56.

Finlayson, N. J., & Golomb, J. D. (2017b). 2D location biases depth-from-disparity judgments but not vice versa. *Visual cognition*, *25*(9-10), 841-852.

Finlayson, N. J., & Grove, P. M. (2015). Visual search is influenced by 3D spatial layout. *Attention, Perception, & Psychophysics*, *77*(7), 2322-2330.

Finlayson, N. J., Remington, R. W., Retell, J. D., & Grove, P. M. (2013). Segmentation by depth does not always facilitate visual search. *Journal of Vision*, *13*(8), 11-11.

Gawryszewski, L., Riggio, L., Rizzolatti, G., & Umiltá, C. (1987). Movements of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologia*, *25*(1), 19-29.

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.

Holliday, I. E., & Braddick, O. J. (1991). Pre-attentive detection of a target defined by stereoscopic slant. *Perception*, *20*(3), 355-362.

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single-and combined-feature targets. *Journal of Experimental Psychology: General*, *118*(3), 258.

Kawahara, J. I. (2003). Contextual cueing in 3D layouts defined by binocular disparity. *Visual cognition*, *10*(7), 837-852.

Kimura, T., Miura, T., Doi, S. I., & Yamamoto, Y. (2009). Effects of self-motion on attention in real 3D space. *Acta psychologica*, *131*(3), 194-201.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Lindsey, L. L. M. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*(2), 171-187.

Marek, N., & Pollmann, S. (2019). Contextual Cueing in Virtual (Reality) Environments. In S. Pollmann (Ed.), *Spatial Learning and Attention Guidance* (73-107). Magdeburg: Humana Press.

Maringelli, F., McCarthy, J., Steed, A., Slater, M., & Umilta, C. (2001). Shifting visuo-spatial attention in a virtual three-dimensional space. *Cognitive Brain Research*, *10*(3), 317-322.

Merrill, E. C., Conners, F. A., Roskos, B., Klinger, M. R., & Klinger, L. G. (2013). Contextual cueing effects across the lifespan. *The Journal of genetic psychology*, *174*(4), 387-402.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264-265.

Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition*, *17*(8), 1244-1258.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic bulletin & review*, *1*(1), 29-55.

Park, G. D., & Reed, C. L. (2015). Haptic over visual information in the distribution of visual attention after tool-use in near and far space. *Experimental brain research*, *233*(10), 2977-2988.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception & Psychophysics*, *41*(3), 191-201.

Plewan, T., & Rinkenauer, G. (2017). Simple reaction time and size–distance integration in virtual 3D space. *Psychological research*, *81*(3), 653-663.

Previc, F. H. (1990). Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioral and Brain Sciences*, *13*(3), 519-542.

Smyth, A. C., & Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Memory & Cognition*, *36*(2), 403-415.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience*, *4*(3), 232-243.

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, *23*(1), 87-102.

Vuilleumier, P., Valenza, N., Mayer, E., Reverdin, A., & Landis, T. (1998). Near and far visual space in unilateral neglect. *Annals of neurology*, *43*(3), 406-410.

Weiss, P. H., Marshall, J. C., Wunderlich, G., Tellmann, L., Halligan, P. W., Freund, H. J., ... & Fink, G. R. (2000). Neural consequences of acting in near versus far space: a physiological basis for clinical dissociations. *Brain*, *123*(12), 2531-2541.

Woodin, M. E., & Allport, A. (1998). Independent reference frames in human spatial memory: body-centered and environmentcentered coding in near and far space. *Memory & Cognition*, *26*(6), 1109-1116.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1-8.

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), 17-17.

Zoest, W., & Donk, M. (2004). Bottom-up and top-down control in visual search. *Perception*, *33*(8), 927-937.

Appendix A

Print screen of an exemplary trial from the training session