Title:

The early attentional pancake: minimal selection in depth for rapid attentional cueing

Authors:

Ryan E. O'Donnell[1], Kyrie H. Murawski[1], Ella Herrmann[1], Jesse Wisch[2], Garrett D. Sullivan[1], and Brad Wyble[1]

[1]Pennsylvania State University

[2]Princeton University

Corresponding Author:

Ryan E O'Donnell

Department of Psychology

Pennsylvania State University

State College, PA 16801

(610) 592-7828 /ryanodonnell7@gmail.com

Abstract

There have been conflicting findings on the degree to which rapidly deployed visual attention is selective for depth, and this issue has important implications for attention models. Previous findings have attempted to find depth-based cueing effects on such attention using reaction time measures for stimuli presented in stereo goggles with a display screen. Results stemming from such approaches have been mixed, depending on whether target/distractor discrimination was required. To help clarify the existence of such depth effects, we have developed a paradigm that measures accuracy rather than reaction time in an immersive virtual-reality environment, providing a more appropriate context of depth. Three modified Posner Cueing paradigms were run to test for depth-specific rapid attentional selectivity. Participants fixated a cross while attempting to identify a rapidly masked black letter preceded by a red cue that could be valid in depth, side, or both. In Experiment 1a, a potent cueing effect was found for lateral cueing validity, but a weak effect was found for depth despite an extreme difference in virtual depth (1 vs 300 meters). In Experiment 1b, a near-replication of 1a, the lateral effect replicated while the depth effect did not. Finally, in Experiment 2, to increase the depth cue's effectiveness, the letter matched the cue's color, and the presentation duration was increased; however, again only a minimal depth-based cueing effect – no greater than that of Experiment 1a – was observed. Thus, we conclude that rapidly deployed attention is driven largely by spatiotopic rather than depth-based information.

Keywords: Rapid Attention, Depth, Binocular Disparity, Spatial Attention

Significance Statement

Visual attention is strongly linked to a spatial reference frame, which corresponds to the 2-dimensional retinotopic visual input. Vision often has a strong sense of depth as well, due to a variety of depth cues. Previous results regarding the mind’s ability to rapidly attend to specific depth planes have been mixed. Here, we show that at the earliest stages of rapid attentional deployment, attention minimally incorporates depth information into the 2-dimensional framework, supporting the current implementation of many computational theories of attention and highlighting a key distinction about when attention operates in two dimensions compared to three.

Introduction

Managing the enormous amount of information from the eyes through the optic nerve is a challenge that visual attention helps to manage. There are many aspects of attention, including eye movements as the stereotypically overt form of attention. Complementary to eye movements is the well-studied covert form of visual attention in which changes in neural processing within the visual system allow it to reconfigure the way it processes information (Yeshurun & Carrasco, 1998; Zhang et al., 2011). For example, the deployment of covert attention can emphasize spatial locations (Posner, 1980; Eriksen & Yeh 1985; Yantis & Serences, 2003; Wyble et al., 2020), objects (Egly, Driver & Rafal, 1994; Chen, 2012; Reppa, Schmidt & Leek, 2012), or specific visual features (Chapman & Störmer, 2021; Maunsell & Treue, 2006; Ling, Liu & Carrasco, 2009; Störmer & Alvarez, 2014). One of the most robust forms of spatial covert attention is the reflexive variety that is commonly assessed via the Posner Cueing Paradigm, in which a cue (e.g., an onset stimulus such as a circle or square) elicits a potent and rapid increase in processing efficiency at its own location (Posner, 1980; Posner & Cohen, 1984).

With only a few exceptions, studies exploring covert reflexive visual attention have not manipulated depth as a factor. This is an important limitation of the literature because it has encouraged a view of visual processing that is focused on a flat 2-dimensional representation distributed across cortex. For example, while it is widely acknowledged that binocular disparity cells exist, they are rarely incorporated into models of visual attention (e.g., Tsotsos, 1995; Tsotsos et al., 1995; Itti & Koch 2000, Wyble et al. 2020).

However, it is clearly possible to attend to different depth planes via overt, sustained attention (de Gonzaga Gawryszewski, 1987; He & Nakayama, 1995; Viswanathan & Mingolla, 2002; Marrara & Moore, 2000), and the existence of binocular disparity cells within the visual system leaves open the possibility that rapid forms of covert attention could select for different depth planes as well. Thus, there is a possible mechanism for rapid cueing in depth, and such a finding would challenge the standard notion of 2-dimensional topographic feed-forward projections between layers that characterize many attention models.

Previous studies have explored the question of spatial cueing in depth using binocular disparity as a depth cue. Ghirardelli & Folk (1996) used a standard cueing experiment with a polarizing stereoscopic display. Participants provided a speeded two-alternative forced choice response to a character during two kinds of cueing trials which were randomly intermixed. In half of the trials, the onset cue and target always appeared on the horopter but at the same or different positions on the horizontal meridian (i.e., depth kept constant), while in the other half they always appeared at the same position on the horizontal meridian but on different sides of the horopter (i.e., side kept constant). The paradigm found a standard cueing effect (faster RTs to valid than invalid trials with neutral RTs roughly in between) on trials where depth was constant but horizontal position was changed. However, there was essentially no effect of cueing, either for valid or invalid trials, when the target and cue were at the same horizontal position but different binocular depths.

Atchley, Kramer, Anderson & Theeuwes (1997) followed this up with a modified binocular disparity task that was made more difficult and had increased reaction times by forcing subjects to identify a target presented among distractors. In this version, there was a significant RT benefit for a valid depth cue at an SOA of 150 ms. When they eliminated the distractor, the RTs were much faster on average (more like those in Ghirardelli & Folk, 1996) and the depth validity effect disappeared. The authors explained this as an effect of perceptual load, such that the need to distinguish between targets and distractors revealed the depth validity effect. This is not consistent with spatiotopic attention because standard visual cueing studies going back to the original Posner paradigm have revealed RT effects with only a single target item, even in the simplest of stimulus-response paradigms.

Since the depth cueing effect using RT seems to hinge on task difficulty in a way that is not consistent with conventional spatial cueing effects, we decided to explore depth-based cueing effects using accuracy as a dependent measure instead of RT. When properly calibrated, measuring cueing effects through accuracy for strongly masked letter targets can reveal very potent reflexive cueing effects (Wyble, Bowman & Potter 2009). Moreover, like Atchley et al. (1997) we used a display that contained a distractor so that the target had to be discriminated. Additionally, we added a staircase to ensure that the stimuli were presented at a speed that was highly challenging but perceivable for the participants. Finally, the other advantage we provided for a depth-based finding was to use a fully immersive environment such that any binocular disparity cues would be embedded in an environment that was consistent with depth planes. This environmental factor is important because perception is strongly contextual, and the fact that binocular displays are often presented on a monitor that is known to be flat could have reduced the effectiveness of disparity.

In this study, participants completed a modified Posner Cueing Paradigm, in which a covert cue could appear at the exact location of the target letter, at the same point on the horizontal meridian but at a different depth plane, at the same depth plane but on the opposite side of the horizontal meridian, or (in Experiment 2 only) at both the opposite side and different depth plane of the cue. Importantly, the experiments were designed so that participants could only rely on visual disparity of the targets to discriminate the depth planes. If rapid attention can be allocated to a specific depth plane, as observed by Atchley et al. (1997), then we should see reliably lower accuracy when depth is invalidly cued (but side is validly cued; termed *depth invalid*) compared to when both depth and side are validly cued (termed *perfect valid*). However, if attention is unable to be allocated to a specific point in depth at rapid timescales, we should observe no reliable accuracy difference between the perfect and depth invalid cues. In addition, we moved the location of the fixation cross across experiments to assess the influence of sustained fixation on a particular depth plane in target identification accuracy.

Experiment 1a

In our first experiment, we created a virtual environment with a large difference in virtual disparity designed to be detectable to human observers without monocular depth cues (Allison, Gillam & Vecellio, 2009; Amigo, 1963; Dees, 1966; Howard & Rogers, 2002; Kaufman et al., 2006; McKee, Levi & Bowne, 1990; Morgan, 2003; Ogle, 1958; Palmer, 1999; Wheatstone, 1838): the front plane was 1 virtual meter from the observer while the back plane was 300 m away. Additionally, the sizes of the stimuli in the front and back depth planes were matched for apparent size by visual superposition to avoid differences in depth perception according to the apparent size phenomenon (Hochberg & Hochberg, 1952).

Methods

Participants: Previous studies assessing for cueing validity effects via accuracy have used sample sizes ranging from 12 to 14 participants and achieved greater than 80% power to detect 14% or greater differences between valid and spatially invalid cueing conditions with similar paradigms (Wyble, Bowman & Potter, 2009). However, because we were interested in comparing accuracy-based cueing effects between valid and depth invalid cueing, we elected to recruit a larger sample size. Thus, 22 Pennsylvania State University undergraduate students participated in the experiment in exchange for course credit. This study was approved by the Pennsylvania State University IRB, and all participants provided consent before participating.

Apparatus: Participants sat in a chair centered in a 10 x 20 m tracking area and wore an HTC Vive Virtual Reality headset (100° H ×110° V field of view, 1080 x 1200 pixels-per-eye resolution). The headset receives signals from two "lighthouses" secured to the ceiling that emit IR pulses using diodes. The HTC Vive uses these pulses to determine the position and orientation of the headset at all times. The virtual environment was generated on a graphics PC (Alienware, Windows 10, NVIDIA 1070 graphics card) using the Unity software to render the images. Custom C# code controlled the timing and placement of objects within the Unity software.

Stimuli and Procedure: In the virtual environment, (Figure 1), participants were placed in a four-sided ceiling-less room 330 m (virtual meters in Unity coordinates) from the left wall, 342.5 m from the right wall, 347 m from the back wall, and 435 m from the front wall[1]. All walls were 248.5 m tall. The front wall was gray, the side and back walls were dark gray with a blue grid pattern, and the ground was white with a gray grid pattern. A black fixation cross was positioned 230 m in front of the participant and 37 m off the ground. Shadows were included in the environment but were not visible due to the height of the observer over the floor. The participants were positioned on one side of the ceiling-less room 45 m from the virtual ground.

To elicit strong percepts of stereopsis, stimuli were presented at two depth planes: a front plane 1 m from the observer and a back plane 300 m from the observer. Four invisible black wire

---

[1] All size dimensions reported reflect virtual values in the Unity environment that were then represented as such through the Vive headset.

cubes were positioned in the room[2]. The front two cubes were sized 0.1 m X 0.1 m X 0.1 m and smaller than the back two cubes, sized 30.5 m x 30.5 m x 30.5 m. All of the boxes were positioned inside the room at a height of 28 m off the ground angled towards the camera by 11 degrees so as to align the boxes with the participant's view as depicted in Figure 1a. Participants could not navigate through the virtual space; they were only capable of rotating their head within the virtual environment. The head's virtual position was fixed.

Participants were instructed to identify the letter presented to them. Each target letter was presented alongside one distractor digit (Figure 1b). The target was randomly selected from the set A, B, C, D, K, N, V, or Z. The distractor was randomly selected from the set 4, 6, and 9. Targets and distractors were scaled to be smaller or larger than their original size depending on their position in the virtual environment in order to appear to be the exact same size when projected onto the VR display. The target and distractor were also rotated 11 degrees to be orthogonal to the viewer.

At the beginning of each trial, one of the invisible wire-frame boxes flashed red for 40 ms, serving as the cue. An interval of 40 ms occurred between the cue offset and the target and distractor onset. One target and one distractor, which were 2-Dimensional flat images of a letter and number, appeared at the position of one of the four box frames. The target and distractor identity and positions were randomized, with the stipulation that each stimulus appear on opposite sides and in different depths. A staircase algorithm controlled the duration of the target display, increasing the duration by 2ms if the participant was incorrect and shortening the duration by 1.5ms if correct across all trials to aim for an accuracy of about 80% across all

---

[2] The wire frames were made invisible because they were matched for apparent size. Had the boxes been made visible, the front plane would have occluded the back plane, and any back plane cues would be undetectable.

conditions (Leek, 2001). The starting value for the staircase was 30 ms. After the target and distractor disappeared, a set of four masks consisting of a superimposed “#” and “@” symbol (also scaled to appear of equal size across depth planes on the VR display) appeared in the center of each of the wire cubes and remained on-screen until the participant’s verbal response as to what target they saw was recorded. After reporting the target’s identity, participants then reported the depth plane of the target as front or back. Verbal responses were recorded by the researcher in the room. The researcher also encouraged the participant to guess if they did not remember which letter they saw, and informed participants when their guess was not in the letter set before encouraging them to guess again. After responding, participants received feedback of what letter appeared and the next trial would begin.

There were three types of trials (Figure 1b): 1) *perfect valid* trials in which the cue and target appeared at the same depth and on the same side (i.e., the target letter appeared within the wireframe box that flickered red), 2) *depth invalid* trials in which the cue and target appeared on the same side but different depths, and 3) *side invalid* trials in which the cue and target appeared at the same depth but on different sides. Each trial type was randomly selected so as to occur approximately ⅓ of the time. Participants completed 5 practice trials with the target on-screen for 500 ms, and then completed 120 trials with the parameters described above.

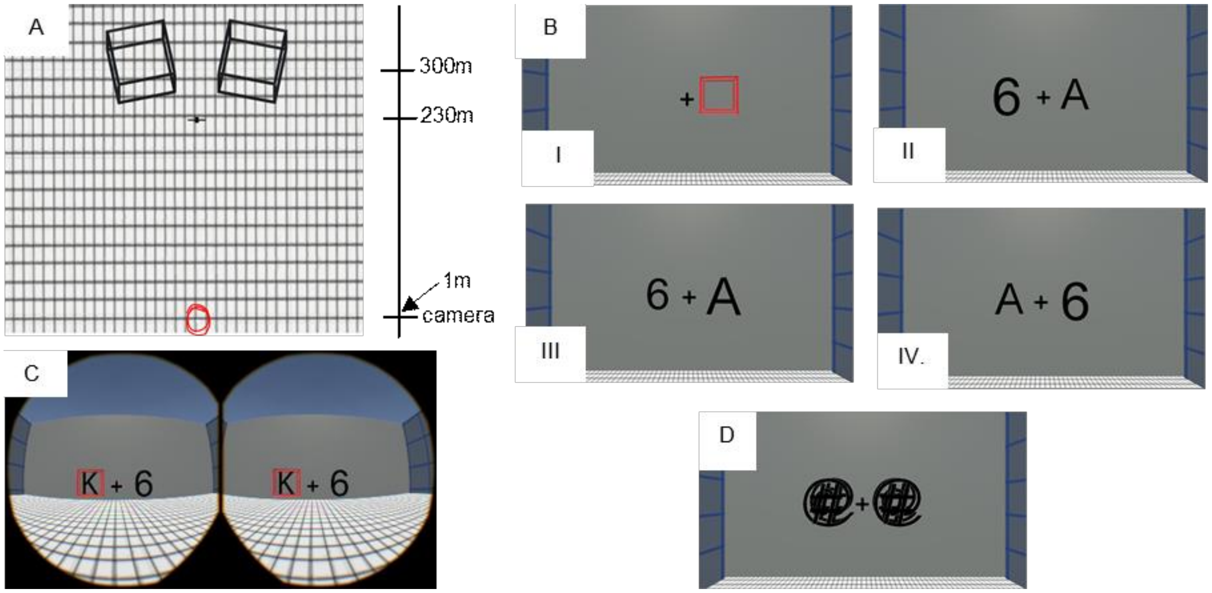

Figure 1: Illustration of the paradigm in Experiment 1a. A) Bird's eye view of virtual setting from the Unity editor, highlighting the distance in depth between the front and back depth planes. Note that the fixation cross appears in front of the back depth plane and the large disparity in stimuli sizes between the front and back depth plane. The red circle is not part of the display but indicates the viewer position and front depth plane placeholders, which are too small to see in this image. B) Sample monocular view of trial displays. This is how participants would see the paradigm. I. Example of cue appearing on back right position. II Example perfect valid trial. III. Example depth invalid trial. IV. Example side invalid trial. C) Binocular display of experiment. The participant would see a merged view, as seen in B. Note, the cue and target are superimposed for emphasis D) Monocular display of masks, which appear shortly after target onset.

Analysis: A repeated measures ANOVA assessed for an overall difference between groups. Following this test, separate paired samples t-tests compared accuracy between the perfect valid condition and each of the two invalid conditions. We also assessed if the location of the target in depth had an impact on performance by comparing accuracy at identifying front targets versus back targets throughout the whole experiment and independently in each cueing condition via paired t-tests. For each paired t-test, a Bayes Factor was also computed via JASP (Wagenmakers et al., 2018) to provide further evidence in favor of the alternative hypothesis (BF_{10}), Cauchy scale = 0.707. $BF_{10} < .33$ is seen as evidence *in favor* of the null hypothesis, whereas $BF_{10} > 3$ is considered substantial evidence in favor of the alternative hypothesis. Values that fall between 0.33 and 3 are considered anecdotal or inconclusive evidence (Kass & Raftery, 1995; Stefan et al., 2019).

To quantify the difference in cueing magnitude between side invalid and depth invalid cueing, we generated 95% bootstrapped confidence intervals (bCI) around the effect sizes of the depth invalid versus perfect valid and side invalid versus perfect valid paired t-tests (Banjanovic & Osborne, 2016). The magnitude of the effect sizes was deemed significantly different if the observed effect sizes did not overlap in the opposing bCI. Each bCI was generated with 10,000 iterations of t-tests run from bootstrapped resampling of the observed data. Additionally, a 95% bootstrapped confidence interval was calculated around the accuracy identifying the target's depth plane (10,000 iterations of sampling with replacement), with depth accuracy deemed as significantly above chance if the bCI did not contain chance performance (50%).

Results

Target identification accuracy within each cueing condition is depicted in Figure 2. There was an overall significant effect of condition, $F(2,20) = 17.87$, $p < 0.001$, $\eta_p^2 = 0.46$. Pairwise comparisons revealed that participants exhibited greater accuracy at identifying the letter when cued perfectly ($M = 63\%$, $SD = 8\%$) compared to both the side invalid cue ($M = 44\%$, $SD = 11\%$), $t(21) = 5.28$, $p < 0.001$, $\eta_p^2 = 0.57$, $BF_{10} = 781.1$, and, surprisingly, the depth invalid cue ($M = 58\%$, $SD = 8\%$), $t(21) = 2.18$, $\mathrm{p} = 0.041$, $\eta_p^2 = 0.18$, though with inconclusive evidence according to the Bayes Factor, $BF_{10} = 1.57$. Additionally, bootstrapped confidence intervals (bCIs) of the effect sizes revealed that the effect size of depth invalid cueing was significantly smaller than side invalid cueing, depth invalid $\eta_p^2 = 0.18$, 95% bCI [.007, .472], side invalid $\eta_p^2 = 0.57$, 95% bCI [.369, .762].

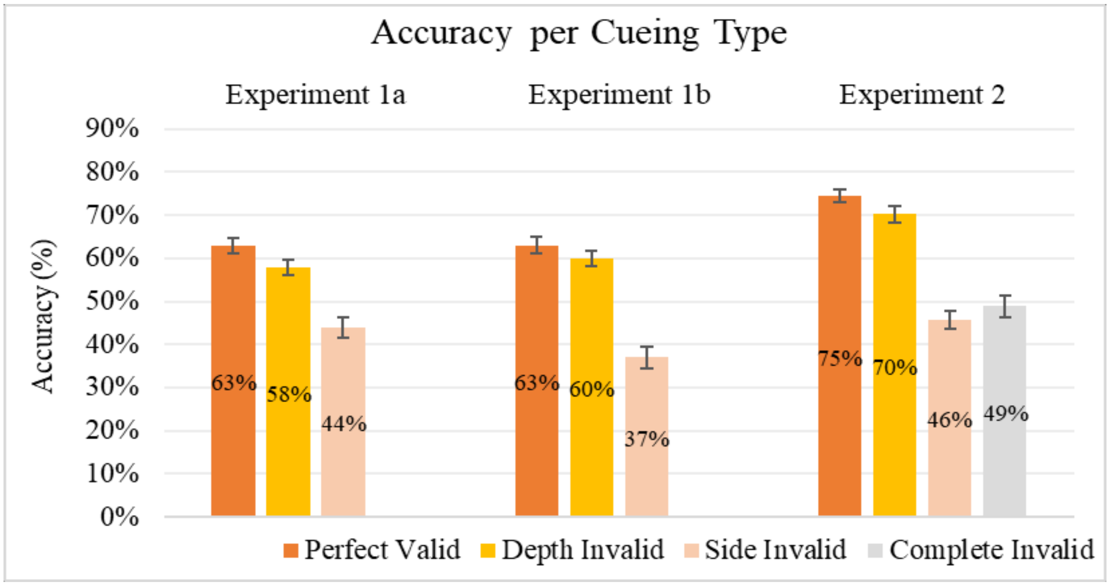

Figure 2 The average percent accuracy of each trial type – perfect valid, depth invalid, side invalid, or complete invalid (Experiment 2 only) – for Experiments 1a, 1b, and 2. Error bars represent ±1 SE from the standard mean.

Overall, participants were more accurate at identifying targets when they were presented in the back depth plane ($M = 58\%$, $SD = 6\%$) compared to the front ($M = 51\%$, $SD = 6\%$), $t(21) = -2.87$, $p = 0.009$, $\eta_p^2 = 0.28$, $BF_{10} = 5.38$ (Figure 3). This difference was of minor note in the perfect valid condition ($M_{Front} = 58\%$ vs $M_{back} = 67\%$), $t(21) = -2.42$, $p = 0.025$, $\eta_p^2 = 0.22$, $BF_{10} = 2.36$. Though numerically present, this difference was not significant in the invalid depth ($M_{Front} = 56\%$ vs $M_{back} = 61\%$), $t(21) = -1.03$, $p = 0.31$, $\eta_p^2 = 0.048$, $BF_{10} = 0.36$, or invalid side conditions, ($M_{Front} = 42\%$ vs $M_{back} = 47\%$), $t(21) = -1.73$, $p = 0.10$, $\eta_p^2 = 0.12$, $BF_{10} = 0.79$.

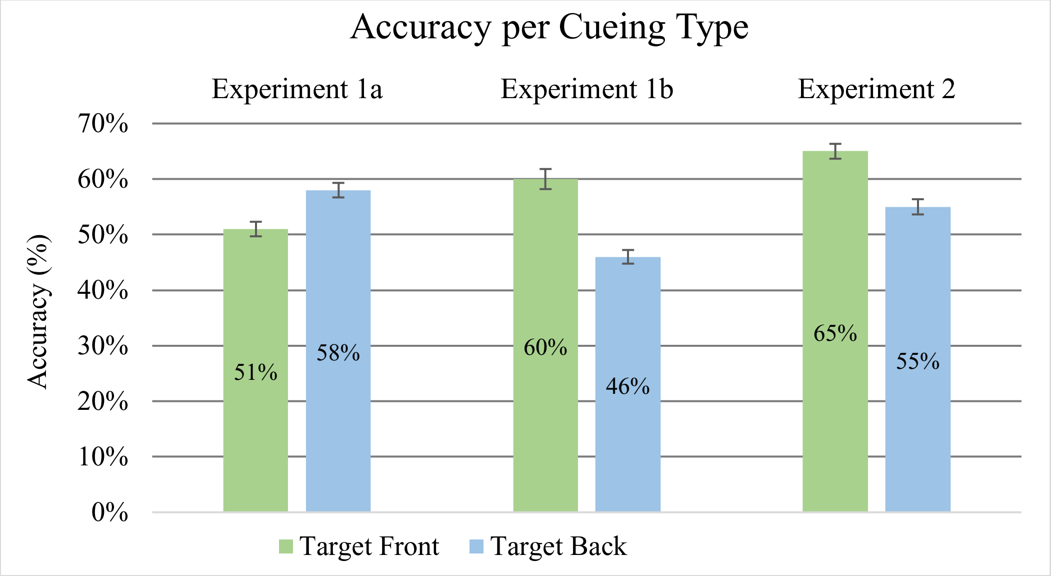

Figure 3 Average percent accuracy for identifying targets presented on the front or back depth plane for Experiments 1a, 1b, and 2. In Experiment 1a, the plane of fixation cross was closer to the back, whereas in Experiments 1b and 2 it was closer to the front. Error bars represent ±1 SE from the mean.

When determining if participants were capable of explicitly reporting the target's depth plane, we observed that participants were on average 52% accurate at making such a judgement. This accuracy was not significantly greater than chance performance, as a 95% bCI generated from the observed data included chance accuracy within its bounds, 95% bCI [49.4%, 55%]. This inability was consistent throughout the experiment, 1st half trials $M = 52\%$, 95% bCI [48.9%, 55.8%], 2nd half trials $M = 52\%$, 95% bCI [48.6%, 55.4%].

Discussion

As expected, we observed a substantial side invalid effect, showing that when the target and the cue were in different spatial locations, the ability to report the target was dramatically reduced compared to when they were in the same location. Our results also revealed conflicting evidence regarding an observer's ability to rapidly incorporate depth information into attentional orienting. First, a small, inconclusive effect of depth validity was observed, yet participants also demonstrated an inability to identify the depth plane of the target at rapid timescales via explicit report. However, this inability to detect depth information was not because of a limitation in the paradigm or design of the virtual room. When asked to do so, two authors and one naïve participant were able to explicitly identify the depth plane of the cue at the same rapid timescales of the of the experiment (100%, 85%, and 80% respective accuracy out of 20 trials). Each of these samples was individually significantly different from chance using a binomial test, all *ps* < .002. The same three individuals were able to identify the depth plane of the target at rapid timescales when the four masks were removed (100%, 90%, and 90% respective accuracy out of 20 trials).

Additionally, target identification accuracy was sensitive to the depth plane of the target, independent of the cueing manipulation. Specifically, targets presented on the back plane, which was closer to the fixation cross, were more accurately identified. Thus, it is unclear how strong of an influence sustained fixation to the depth plane of the fixation cross has on the observed depth-based cueing observation, and if this effect is masking the sensitivity of rapid attentional deployment to depth information. To further assess this possibility, we replicated Experiment 1a, while changing the depth plane of the fixation cross to independently assess the reliability of a depth-based cueing effect and an endogenous attentional shift to the depth plane of the fixation cross.

Experiment 1b

In Experiment 1a, the fixation cross was positioned at the back of the virtual room, which would have encouraged participants to set a focal depth in the far distance. This created a discrepancy in accuracy for back compared to front targets. In Experiment 1b, the fixation cross was moved to a virtual depth plane of 0.8 m from the observer (Figure 4a) to set the focal depth at a near distance and assess whether the discrepancy in front versus back targets reverses. 20 new undergraduates from Pennsylvania State University participated for course credit after providing informed consent. Experiment 1b was identical to Experiment 1a in all other aspects.

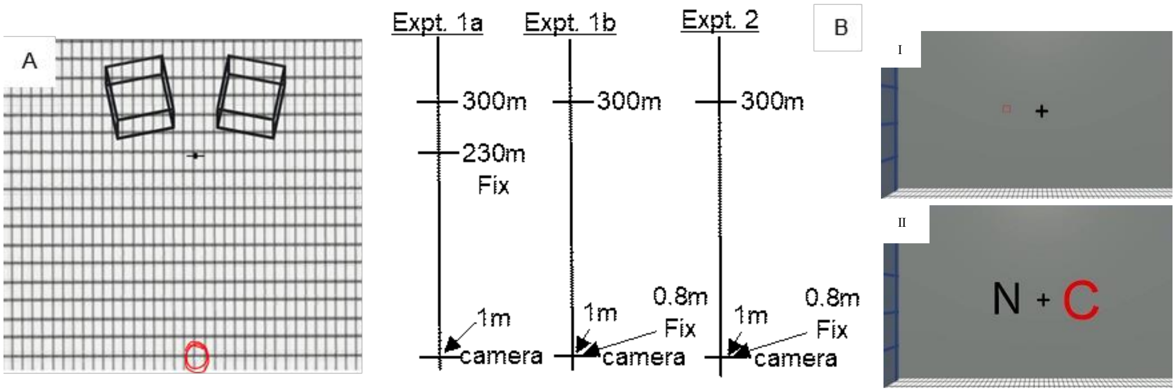

Figure 4: Layout of depth planes in Experiments 1a, 1b, and 2. A) Changes in depth plane of the three experiments, alongside a Bird's eye view for visualization. Note the viewer is represented by the location of the camera, and that only fixation cross (labeled "Fix" in the legend) moves between experiments, as can be seen in the legend. B) Monocular display of the paradigm in Experiment 2. I. Example of cue. II Example of target and distractor. The target now matched the cue color, and the distractor was changed to a black letter.

Results

Results of Experiment 1b are visualized in Figure 2. There was an overall significant effect of condition, $F(2,18) = 42.94$, $p < 0.001$, $\eta_p^2 = 0.69$. Pairwise comparisons revealed that participants exhibited greater accuracy at identifying the letter in the perfect valid cue condition ($M = 63\%$, $SD = 9\%$) compared to the side invalid cue ($M = 37\%$, $SD = 11\%$), $t(19) = 7.45$, $p < 0.001$, $\eta_p^2 = 0.75$, $BF_{10} = 3.53 \times 10^4$. However, unlike Experiment 1a there was no longer a significant difference in accuracy when the cue was perfectly valid compared to when the cue was invalid in depth ($M = 60\%$, $SD = 8\%$), $t(19) = 1.73$, $p = 0.101$, $\eta_p^2 = 0.14$, though with only anecdotal evidence in favor of the null hypothesis, $BF_{10} = 0.82$. Additionally, bCIs of the effect sizes revealed that the effect size of depth invalid cueing was significantly smaller than side invalid cueing, depth invalid $\eta_p^2 = 0.14$, 95% bCI [.002, .428], side invalid $\eta_p^2 = 0.75$, 95% bCI [.588, .882].

Moving the fixation cross to the front depth plane reversed the accuracy boost seen from identifying targets presented in the back depth plane (Figure 3). In this experiment, participants exhibited a greater ability to identify targets presented in the front plane ($M_{Front} = 60\%$, $SD = 8\%$) versus the back plane ($M_{back} = 46\%$, $SD = 6\%$), $t(19) = 6.00$, $p < 0.001$, $\eta_p^2 = 0.65$, $BF_{10} = 2411.17$. This difference was significant in all three cueing conditions, but inconclusive in the perfectly valid condition: perfect valid ($M_{Front} = 66\%$ vs $M_{back} = 60\%$), $t(19) = 2.24$, $p = 0.038$, $\eta_p^2 = 0.21$, $BF_{10} = 1.77$, depth invalid (M_{Front} 70% vs M_{back} 52%), $t(19) = 3.79$, $p = 0.001$, $\eta_p^2 = .43$, $BF_{10} = 30.32$, side invalid (M_{Front} 44% vs M_{back} 29%), $t(19) = 5.06$, $p < 0.001$, $\eta_p^2 = 0.57$, $BF_{10} = 388.15$.

Unlike Experiment 1a, participants were able (albeit slightly) to identify the depth plane of the target, $M = 53\%$, 95% bCI = [50.6%, 54.8%]. However, this ability seemed to grow over time, as participants were not able to identify the correct depth plane in the 1st half of the experiment, $M = 52\%$, 95% bCI [49.1%, 55.3%], but could do so in the 2nd half, $M = 53\%$, 95% bCI [50.9%, 55.9%].

Discussion

Taken together, the results of Experiment 1a and 1b suggest that when binocular disparity serves as the only informative depth signal to the observer in an immersive headset, participants cannot reliably attend to different depth planes at a rapid timescale. Though the location of the fixation cross can influence where focal attention is initially allocated in depth, and attention can rapidly deploy to a lateral position in a 2-dimensional plane, the onset of the cue cannot rapidly shift attention to different depth planes in a way that strongly and reliably influences processing at a given depth plane. Moreover, in both Experiments 1a and 1b, the magnitude of the side invalid effect size was significantly greater than that of the depth invalid effect size. Collectively, these results suggest that when covert reflexive attention is cued to a new location, attentional allocation is largely compressed into a 2-dimensional format.

One alternative account to our observations is that the paradigm is not designed to maximize rapid attentional deployment in depth. In particular, in Experiments 1a and 1b, though the cue was unpredictive, as in that it did not inform participants as to which condition was being presented, the nature of the experimental design made it so that the correct side was cued 2/3 of the time, and the correct depth 2/3 of the time. Thus, the cue was pseudo-predictive for the two spatial scales. This pseudo-predictability may have encouraged participants to over rely on the lateral nature of the cue, potentially encouraging them to ignore depth information. Additionally, the cue may have been an imperfect design, as it only matched the target by its abrupt onset similarity. In our next experiment, we attempted a replication that controlled for these possible alternatives for why depth information was a minimal influence on attentional deployment in depth.

Experiment 2

Experiment 2 was similar to Experiments 1a and 1b with the exception that the parameters of the experiment were designed to strengthen the influence of depth validity. Specifically, participants now saw a red and a black letter on each trial and would report the red letter while ignoring the black letter (Figure 4b). The cue was also red, so that there was a match between the color of the cue and the task relevant feature of the target. Additionally, a fully invalid condition was introduced to make the cue spatially uninformative, ensuring that our previously observed results were not due to a biased utilization of lateral spatial cues. Finally, the display duration of the target was increased, to remove the possibility of feature conjunctions (the red cue masked the red letter when presented in the same location).

Methods

Participants: 22 Pennsylvania State University students participated in Experiment 2 after providing informed consent.

Stimuli & Apparatus. This experiment replicated the design of Experiment 1b except in the following ways. A red cue was used as before, but the search display now consisted of a black letter and a red letter (selected from the same set). Red letters were masked by a red mask and black letters by a black mask.

Procedures. Participants were instructed to report the red letter and ignore the black letter and that they would see a red cue that may inform them of the target's location. The match between the color of the cue and the target letter caused feature conjunctions at very short lags. This required us to shrink the size of the cue (length of one side of the cube reduced by factor of 7) and start the duration of the staircase for letter duration at 100ms instead of 30ms, but this second change also provided participants more time to process the target's depth information in the display. The staircase adjusted this value upwards by 8ms on an incorrect trial and downwards by 6ms on a correct trial. Additionally, a fourth, *complete invalid* condition, was introduced in this experiment, where the cue appeared on the incorrect side and depth plane of the target. This change was introduced to make the cue's location completely unpredictive.

Finally, with slower display durations, practice trials were no longer necessary and now used as the first 5 trials of the experiment. Thus, there were 125 trials in total in this experiment.

Results

Results of Experiment 2 are visualized in Figure 2. There was an overall significant effect of condition, $F(2,20) = 42.75$, $p < 0.001$, $\eta_p^2 = 0.67$. Pairwise comparisons revealed that participants exhibited greater accuracy at identifying the letter in the perfect valid cue condition ($M = 75\%$, $SD = 7\%$) compared to the side invalid cue ($M = 46\%$, $SD = 10\%$), $t(21) = 9.36$, $p < 0.001$, $\eta_p^2 = 0.81$, $BF_{10} = 2.03\text{x}10^6$. Like Experiment 1a, there was a significant difference in accuracy when the cue was perfect valid compared to when the cue was invalid in depth ($M = 70\%$, $SD = 9\%$), $t(21) = 2.11$, $p = 0.047$, $\eta_p^2 = 0.175$, but again this finding was inconclusive, $BF_{10} = 1.41$. Additionally, bCIs of the effect sizes revealed that the effect size of depth invalid cueing was significantly smaller than side invalid cueing, depth invalid $\eta_p^2 = 0.175$, 95% bCI [.003,.519], side invalid $\eta_p^2 = 0.81$, 95% bCI [.713, .897].

Keeping the fixation cross at the front depth plane maintained the accuracy benefit for front compared to back targets as seen in Experiment 1b. (Figure 3). Participants exhibited a greater ability to identify targets presented in the front plane ($M_{Front} = 65\%$, $SD = 6\%$) versus the back plane ($M_{back} = 55\%$, $SD = 6\%$), $t(21) = 3.99$, $p < 0.001$, $\eta_p^2 = 0.43$, $BF_{10} = 51.45$. This difference was present in all conditions, but inconclusive in all but the depth invalid condition: perfect valid ($M_{Front} = 79\%$ vs $M_{back} = 71\%$), $t(21) = 2.33$, $p = 0.03$, $\eta_p^2 = 0.21$, $BF_{10} = 2.01$, depth invalid (M_{Front} 78% vs $M_{back} = 62\%$), $t(21) = 4.09$, $p < 0.001$, $\eta_p^2 = .44$, $BF_{10} = 63.58$, side invalid ($M_{Front} = 51\%$ vs $M_{back} = 41\%$), $t(21) = 2.02$, $p = 0.056$, $\eta_p^2 = 0.16$, $BF_{10} = 1.22$, complete invalid ($M_{Front} = 55\%$ vs $M_{back} = 44\%$), $t(21) = 2.22$, $p = 0.037$, $\eta_p^2 = 0.19$, $BF_{10} = 1.69$.

With the increased search duration, participants were (as expected) able to identify the depth plane of the target, *M* = 56%, 95% bCI = [51.1%, 62.2%]. The ability to recognize depth, was consistent over time: 1st half of trials, *M* = 58%, 95% bCI [50.5%, 61.8%], 2nd half, *M* = 59%, 95% bCI [50.7%, 63.0%].

Discussion

It is possible that the combination of a target-cue color mismatch, a pseudo-laterally predictive cue, and a rapid target presentation duration used in Experiments 1a and 1b minimized the effectiveness of cueing depth, thereby failing to reveal a cueing effect. However, Experiment 2 demonstrated that even with both a task relevant, laterally unpredictive cue and a greater than chance ability to identify the depth plane of the target, there was still a stronger effect of laterality-based cueing compared to a weak – deemed inconclusive – depth-based cueing effect on target report. That this minor depth benefit is no greater than the effect previously observed in Experiment 1a – when no measures were taken to maximize depth-based attentional deployment – suggests that rapid depth-based attentional cueing by disparity is near inconsequential.

Moreover, this experiment again demonstrated a discrepancy of target report based on the depth plane it was presented. Like Experiment 1b, targets presented on the same plane as the fixation cross were better reported, suggesting that the binocular effects of the paradigm are able to drive differences in fixation depth with consequent effects on perception at similar virtual depth planes. Thus, we can observe depth-based perceptual effects in VR, but they are largely non impactful with respect to rapid attentional orienting.

General Discussion

The results of three experiments which utilize extreme manipulations of binocular disparity revealed small and unreliable evidence of attentional modulation across depth planes for highly salient cues presented about 100 ms prior to the target's onset. While these cues were able to reduce accuracy in reporting the target when cueing the opposite side of the visual display, there was very little difference in accuracy when the cue was presented in a different depth plane relative to the target.

One possibility is that participants could not rely upon depth-based cueing because of their inability to perceive depth without monocular cues. However, a core component of our argument is that depth cannot be successfully utilized at rapid timescales. As previously confirmed in Experiment 1a, two authors were able to report the depth of both the cue and target when presented at the rapid timescales of the experiment with nearly perfect accuracy when the target was unmasked. Therefore, depth is discernable in our paradigm, but potentially not able to guide rapid attention. Another possibility for a minimally significant finding is that some participants were better able to perceive (and thus subsequently incorporate) depth information than others. As a test of this possibility, we correlated each participant's depth report accuracy (or their ability to perceive the target’s depth) with their depth validity effect (perfect valid accuracy minus depth invalid accuracy). If participants who are better able to recognize depth manipulations are subsequently affected by cueing manipulations, then we should see a significant positive correlation between depth report accuracy and depth cueing validity. In all three experiments, this correlation was not significant: Experiment 1a, $r = -.13$, $p = 0.56$; Experiment 1b, $r = .34$, $p = .14$; Experiment 2, $r = .22$, $p = .32$, which supports the claim that depth-based cueing has a minimal-at-best influence on rapid attentional deployment. Thus, despite the ability to rapidly orient to a different two-dimensional location, additional time is required to shift attention to a different depth plane.

Our results are consistent with findings that attention can deploy to different depths at slower time courses, given that the depth plane of fixation had a clear effect on target accuracy across experiments (see also de Gonzaga Gawryszewski, 1987; He & Nakayama, 1995; Marrara & Moore, 2000). However, our results add support for further distinguishing rapid versus sustained forms of attention, as the rapid component of attention seems unable to use a depth cue effectively. This could also explain the discrepancy between paradigms with fast RTs that find no evidence of attention in depth (Iavecchia & Folk, 1994; Ghirardelli & Folk, 1996; Marrara & Moore, 2000) and paradigms with slower RTs or longer display durations that find some evidence (Atchley et al., 2007; Bauer et al., 2012; Marrara & Moore; 2000), if the latter investigations provided more of an opportunity for depth-based mechanisms to affect response time. Though our findings are suggestive of such an account, future work should more directly address this question regarding the time course of disparity based attentional deployment in depth.

Ultimately, our findings contribute to our understanding of the mechanisms that guide rapid attention in a spatial reference frame. The results suggest that in the initial, rapid stages of attentional deployment, 3-dimensional space is largely pancaked into a 2-dimensional reference frame, by attending to all (or at least all task relevant) information within that space regardless of disparity. Our mixed results additionally suggest further nuance from the findings presented in the introduction. Though we utilized distractors in the task like Atchley et al., (1997) and a very difficult target discrimination, we could not find a reliable depth-based cueing effect most likely because such an effect is small and inconsistent at rapid timescales. However, like Ghirardelli &

Folk (1996), despite minimal to no evidence for depth-based attentional orienting, we show clear evidence of attentional orienting to lateral cues. This discrepancy between the importance of lateral versus depth information in rapid attentional orienting supports theories of spatial attention in which reflexive decisions to attend to a location are largely mediated by a 2-dimensional representation that lacks depth information (e.g., Tsotsos, 1995; Tsotsos et al., 1995; Itti & Koch 2000, Wyble et al. 2020). In other words, at least in the early – perhaps primarily feedforward – stages of attentional deployment, 3-dimensional space is largely (and imperfectly) pancaked into 2 dimensions, consistent with current models of spatial attention. Following an initial deployment of 2-dimensional attention, subsequent mechanisms could elaborate these initial representations into depth-aware representations. These findings illustrate the importance of understanding how visual representations unfold in time, allowing depth information to be computed after an initial feed-forward pass of processing.

Acknowledgements

This work was funded by NSF Grant 1734220. The authors would like to thank Justin Lee for their assistance in coding Experiment 2 of this manuscript, and Joyce Tam and Shekoofeh Hedayati for their assistance in implementing Bayes Factor analyses. The authors would also like to thank all members of the Wyble Lab for their time in piloting these experiments and collecting data.

Open Practices Statement

All experimental scripts, analysis files and data can be found at this link: https://osf.io/qgu2c/. No experiments in this manuscript were pre-registered.

References

Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*(1), 10. doi: 10.1167/9.1.10

Amigo, G. (1963). Variation of stereoscopic acuity with observation distance. *Journal of the Optical Society of America*, *53*, 630–635.

Atchley, P., Kramer, A. F., Andersen, G. J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychonomic Bulletin & Review*, *4*(4), 524-529.

Banjanovic, E. S., & Osborne, J. W. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical Assessment, Research, and Evaluation*, *21*(1), 5.

Bauer, D., Plinge, A., Ehrenstein, W. H., Rinkenauer, G., & Grosjean, M. (2012). Spatial orienting of attention in stereo depth. *Psychological Research*, *76*(6), 730-735.

Chapman, A. F., & Störmer, V. S. (2021). Feature-based attention is not confined by object boundaries: spatially global enhancement of irrelevant features. *Psychonomic Bulletin & Review*, 1-9.

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, *74*(5), 784-802.

de Gonzaga 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.

Dees, J. W. (1966). Accuracy of absolute visual distance and size estimation in space as a function of stereopsis and motion parallax. *Journal of Experimental Psychology*, *72*, 466–476.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161-177.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583-597.

Ghirardelli, T. G., & Folk, C. L. (1996). Spatial cuing in a stereoscopic display: Evidence for a "depth-blind" attentional spotlight. *Psychonomic Bulletin & Review*, *3*(1), 81-86.

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proceedings of the National Academy of Sciences*, *92*(24), 11155-11159.

Hochberg, C. B., & Hochberg, J. E. (1952). Familiar size and the perception of depth. *The Journal of Psychology*, *34*(1), 107-114.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth, Vol. 2: Depth perception*. University of Toronto Press.

Iavecchia, H. P., & Folk, C. L. (1994). Shifting visual attention in stereographic displays: A time course analysis. *Human Factors*, *36*(4), 606-618.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489-1506.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Kaufman, L. Kaufman, J. H. Noble, R. Edlund, S. Bai, S. King, T. (2006). Perceptual distance and the constancy of size and stereoscopic depth. *Spatial Vision*, *19*, 439–457.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279-1292.

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194-1204.

Marrara, M. T., & Moore, C. M. (2000). Role of perceptual organization while attending in depth. *Perception & Psychophysics*, *62*(4), 786-799.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317-322.

McKee, S. P. Levi, D. M. Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779.

Morgan, M. (2003). *The space between our ears: How the brain represents visual space*. London: Weidenfeld & Nicolson.

Ogle, K. N. (1958). Note on stereoscopic acuity and observation distance. *Journal of the Optical Society of America*, *48*, 794–798.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge: MIT Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531-556.

Reppa, I., Schmidt, W. C., & Leek, E. C. (2012). Successes and failures in producing attentional object-based cueing effects. *Attention, Perception, & Psychophysics*, *74*(1), 43-69.

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042-1058.

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology*, *24*(17), 1985-1988.

Tsotsos, J. K. (1995). Toward a computational model of visual attention. *Early vision and beyond*, 207-218.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1-2), 507-545.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception*, *31*(12), 1415-1437.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76.

Wheatstone, C. (1838). Contributions to the physiology of vision—Part the first On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society*, *128*, 371–394.

Wyble, B., Callahan-Flintoft, C., Chen, H., Marinov, T., Sarkar, A., & Bowman, H. (2020). Understanding visual attention with RAGNAROC: A reflexive attention gradient through neural AttRactOr competition. *Psychological Review*, *127*(6), 1163-1198. doi: 10.1037/rev0000245.

Wyble, B., Bowman, H., & Potter, M. C. (2009). Categorically defined targets trigger spatiotemporal visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 324.

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology*, *13*(2), 187-193.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72-75.

Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences*, *108*(21), 8850-8855.