The Effects of Task Difficulty on Visual Search Strategy in Virtual 3D Displays

Pomplun, M.; Garaas, T.W.; Carrasco, M.

Abstract

Analyzing the factors that determine our choice of visual search strategy could shed light on visual behavior in everyday situations. Previous results suggest that increasing task difficulty leads to more systematic search paths. Here we analyze observers' eye movements in an 'easy' conjunction search task and a 'difficult' shape search task to study visual search strategies in stereoscopic search displays with virtual depth induced by binocular disparity. Standard eye-movement variables – such as fixation duration and initial saccade latency – as well as new measures proposed here – such as saccadic step size, relative saccadic selectivity, and x-y target distance – revealed systematic effects on search dynamics in the horizontal-vertical plane throughout the search process. We found that in the 'easy' task, observers start with the processing of display items in the display center immediately after stimulus onset and subsequently move their gaze outwards, guided by extrafoveally perceived stimulus color. In contrast, the 'difficult' task induced an initial gaze shift to the upper-left display corner, followed by a systematic left-right and top-down search process. The only consistent depth effect was a trend of initial saccades in the easy task with smallest displays to the items closest to the observer. The results demonstrate the utility of eye-movement analysis for understanding search strategies and provide a first step toward studying search strategies in actual 3D scenarios.

Journal of Vision

**The Effects of Task Difficulty on Visual Search Strategy
in Virtual 3D Displays**

Marc Pomplun*[1], Tyler W. Garaas[2], and Marisa Carrasco[3]

[1] Department of Computer Science, University of Massachusetts at Boston,
100 Morrissey Boulevard, Boston, MA 02125-3393, USA

[2] Mitsubishi Electric Research Laboratories,
201 Broadway, Cambridge, MA 02139

[3]Department of Psychology and Center for Neural Science, New York University,
6 Washington Place, New York, NY 10003

*Corresponding author
Phone: 617-287-6443
Fax: 617-287-6433
e-mail: marc@cs.umb.edu

Abstract

Analyzing the factors that determine our choice of visual search strategy could shed light on visual behavior in everyday situations. Previous results suggest that increasing task difficulty leads to more systematic search paths. Here we analyze observers' eye movements in an "easy" conjunction search task and a "difficult" shape search task to study visual search strategies in stereoscopic search displays with virtual depth induced by binocular disparity. Standard eye-movement variables – such as fixation duration and initial saccade latency – as well as new measures proposed here – such as saccadic step size, relative saccadic selectivity, and x-y target distance – revealed systematic effects on search dynamics in the horizontal-vertical plane throughout the search process. We found that in the "easy" task, observers start with the processing of display items in the display center immediately after stimulus onset and subsequently move their gaze outwards, guided by extrafoveally perceived stimulus color. In contrast, the "difficult" task induced an initial gaze shift to the upper-left display corner, followed by a systematic left-right and top-down search process. The only consistent depth effect was a trend of initial saccades in the easy task with smallest displays to the items closest to the observer. The results demonstrate the utility of eye-movement analysis for understanding search strategies and provide a first step toward studying search strategies in actual 3D scenarios.

Introduction

The research paradigm of visual search is derived from the common real-world task of looking for a visually distinguished object in one's surroundings (for recent reviews see Eckstein, 2011, and Nakayama & Martini, 2011). In a typical laboratory visual search task, observers are asked to report a feature of a designated target item among distractor items. In most studies, response times and error rates have been analyzed as functions of the number of items in the display (set size). When measuring search performance, it is important to account for the observer's ability to increase search speed at the expense of response accuracy and vice versa. The response-signal speed-accuracy tradeoff (SAT) procedure (Reed, 1973) has been used to jointly assess speed and accuracy in visual search (e.g., Carrasco & McElree, 2001; McElree & Carrasco, 1999). Even though such an analysis is not always possible, it is important to analyze both speed and accuracy when assessing visual search performance.

The main mechanism enabling observers to successfully perform visual search is visual attention. Attention can be deployed to a given location overtly, accompanied by eye movements, or covertly, in the absence of such eye movements (for a recent review see Carrasco, 2011). Although the dynamics of attention and eye movements are not identical, they are closely coupled in visual search tasks in which observers are free to move their eyes (Findlay, 2004). For this reason, several studies have examined eye movements in visual search tasks, providing fine-grained measures that supplement global performance indicators such as RT and error rate. Eye movements can also be used to examine the factors that determine our choice of visual search strategy, which may greatly improve our understanding of visual behavior in everyday situations. Characterizing eye movements has revealed some influences of set size and task difficulty on visual search strategies (e.g., Pomplun, Reingold, & Shen, 2001; Zelinsky, 1996). In this study, we investigated observers' search strategies in a 3D virtual display and developed novel measures to analyze observers' eye movements with displays of different sizes and task difficulties (see **Figure 1a**).

Several theories of visual search have been proposed. An early theory of visual search, Feature Integration Theory (Treisman & Gelade, 1980; Treisman, Sykes, & Gelade, 1977), suggests that preattentive feature maps represent individual stimulus dimensions such as color or shape. If a search target is defined by a feature in a single stimulus dimension, it can be detected based on the information in a single feature map and can be detected very efficiently – it "pops out." However, when the target is defined by a conjunction of features, attention is necessary to locally integrate the information of the multiple feature maps, leading to serial search patterns and lower search efficiency. These assumptions explained the set-size effect – reduced search performance with larger displays – that is commonly found in conjunction search but to a much lesser extent in feature search. Subsequent studies yielded inconsistent results with the assumptions of this theory, leading to modifications of its original version (Treisman, 1991a, 1991b, 1993; Treisman & Gormican, 1988) and to new approaches such as the Guided Search theory (e.g., Cave & Wolfe, 1990; Wolfe 1994, 1996; Wolfe, Cave, & Franzel, 1989). According to this theory, preattentive processes create an activation map indicating likely target locations in the display. Those items that are least similar to their neighboring items ("bottom-up" influences) or are most similar to the target ("top-down" influences) receive more activation. This activation map guides serial shifts of attention in the subsequent visual search process.

An alternative explanation for the finding of reduced search performance with greater set size is based on Signal Detection Theory (SDT; Green & Swets, 1966). SDT models account for noise in vision and decision making tasks by assuming that greater set size increases the noise in the observer's visual input. This noise, in turn, may raise the probability of a target being confused with

a distractor (e.g., Cameron, Tai, Eckstein & Carrasco, 2004; Eckstein, 1998; Geisler & Chou, 1995; Palmer, Verghese & Pavel, 2000).

An important finding with regard to visual search strategies is the eccentricity effect. With short display presentations and in the absence of eye movements, performance in visual search tasks deteriorates – reaction time is lengthened and accuracy decreased – as the target is presented at farther peripheral locations, and this effect becomes more pronounced with greater set size (Carrasco & Chang, 1995; Carrasco & Frieder, 1997; Carrasco et al., 1995, 1998; Geisler & Chou, 1995). This reduction in performance is attributed to the poorer spatial resolution at the periphery because more eccentric receptive fields integrate over a larger area and therefore include more distractors. Consistent with this spatial resolution explanation, when stimulus size is enlarged according to the cortical magnification factor performance is constant across eccentricity (Carrasco & Frieder, 1997; Carrasco et al., 1998). These findings suggest that spatial resolution is a limiting factor in visual search.

Given that this eccentricity effect is well documented, many experimenters place stimuli at isoeccentric locations to mitigate perceptual differences (e.g., Cameron et al., 2004; Giordano, McElree & Carrasco, 2009; Eckstein, 1998; Moher et al., 2011; Palmer et al., 2000; Talgar, Pelli & Carrasco, 2004). However, differences in performance at isoeccentric locations can be quite pronounced, even at parafovea. The shape of the visual performance field, with eccentricity held constant, is characterized by a Horizontal–Vertical Anisotropy (HVA), in which performance is better along the horizontal than the vertical meridian, and a vertical meridian asymmetry (VMA), in which performance is better in the lower than the upper region of the vertical meridian.

These performance fields emerge in contrast sensitivity and spatial resolution tasks (e.g., Abrams, Nizam & Carrasco, 2012; Cameron, Tai, & Carrasco, 2002; Carrasco, Penpeci-Talgar, & Cameron, 2001; Mackeben, 1999; Montaser-Kouhsari & Carrasco, 2009; Silva et al., 2010; Skrandies, 1985; Talgar & Carrasco, 2002), as well as in visual search tasks (Carrasco, Giordano, & McElree, 2004; Kristjánsson & Sigurdardottir, 2008; Najemnik & Geisler, 2008, 2009). Both the HVA and the VMA become more pronounced as target eccentricity, target spatial frequency, and the set size increase (Cameron, Tai, & Carrasco, 2002; Carrasco, Talgar, & Cameron, 2001). Moreover, information accrual, i.e., the rate at which discriminability rises from chance to two-thirds of its asymptotic level, also manifests these asymmetries; accrual is faster along the horizontal than the vertical meridian, and it is faster along the lower than the upper vertical meridian (Carrasco, Giordano, & McElree, 2004).

To study the selection of saccade targets, some visual search experiments have used a variety of eye-movement measures including fixation duration, saccade amplitude, number of fixations per trial, initial saccadic latency, and the distribution of saccadic endpoints. These measures are sensitive to manipulations considered to influence cognitive processes underlying visual search performance (e.g., Bertera & Rayner, 2000; Findlay & Gilchrist, 1998; Jacobs, 1987; Motter & Belky, 1998; Rayner & Fisher, 1987; Williams & Reingold, 2001; Williams et al., 1997, Zelinsky & Sheinberg, 1997). For instance, consistent with the predictions of the Guided Search Theory, such studies have documented that the spatial distribution of saccadic endpoints is biased towards distractors sharing a particular feature such as color or shape with the target item (e.g. Findlay, 1997; Hooge & Erkelens, 1999; Luria & Strauss, 1975; Motter & Belky, 1998; Pomplun, 2006; Shen, Reingold, & Pomplun, 2000; Tavassoli et al., 2007; 2009; Williams & Reingold, 2001; Williams, 1967, but see Zelinsky, 1996).

Task difficulty seems to be a crucial factor affecting the observers' strategies in visual search. It determines the capacity of visual processing; that is, the amount of task-relevant information that can be processed during a fixation (also termed "visual span" or "useful field of view"; see Bertera

& Rayner, 2000). Easy tasks often allow the processing of multiple display items within a single fixation, and saccades are typically directed towards the centers of item clusters rather than individual items, a finding termed the "global effect" or "center-of-gravity" effect (Findlay, 1982; 1997). This observation served as the basis of the Area Activation Model (Pomplun et al., 2003) that predicts the statistical distribution of saccadic endpoints by assuming a maximization of the amount of relevant information to be processed during the subsequent fixation. Difficult tasks, on the other hand, seem to emphasize the influence of strategic factors on the large-scale structure of scanpaths during search (e.g., Pomplun, Reingold, & Shen, 2001; Zelinsky, 1996). For example, some observers tend to scan a search display in their usual reading direction. Such strategies facilitate efficient scanning by reducing the frequency of individual items being overlooked or being fixated more than once, which is particularly beneficial for difficult search tasks due to the greater performance gain.

Given the importance of visual search dynamics for the performance of everyday tasks, the purpose of the present eye-movement study was to systematically explore search strategy as a function of task difficulty and set size. To vary task difficulty, two different search tasks were used: First, an easy conjunction search task using colored bars shown at different tilt angles (**Figures 1b and 1d**), and second, a difficult shape search task using arbitrarily oriented shapes, with the target being a mirror image of the distractors (**Figures 1c and 1d**).

According to earlier results on performance fields described above, and the finding that when a signal is embedded in noise, saccades are more frequent to the locations of lower visual sensitivity (Najemnik & Geisler, 2008; 2009, but see Morvan & Maloney, 2012), we could hypothesize that eye movements may be directed to the upper vertical meridian most, followed by the lower vertical meridian and then by locations along the horizontal meridian. Furthermore, we assume that the eccentricity effect also influences overt shifts of attention. Indeed, increasing target eccentricity has been shown to increase the number of saccades necessary to find a conjunction target (Scialfa & Joffe, 1998). In order to optimally use their fovea and parafovea, observers may tend to fixate more central clusters of items first and thus produce an eccentricity effect that is similar to the one previously found for covert shifts of attention. Moreover, it is possible that increased task difficulty will bias attentional control towards systematic scanning strategies that could deviate from the pattern predicted by the eccentricity effect (e.g., Carrasco et al., 1995, 1998) and the performance fields (e.g., Carrasco et al., 2001) described above.

Besides using eye-movement analysis to systematically study task difficulty effects on visual search dynamics, the present study explores another aspect of visual search that is often neglected – the depth dimension. Although the visual search studies described above have provided significant insight into visual processing and visual attention, these findings may be restricted because all of these studies were based on two-dimensional stimuli. Therefore, to investigate all capabilities of the visual system–which has evolved and is being trained in a three-dimensional environment–and analyze them in a more natural context, experiments employing three-dimensional scenes are essential. As a first step towards exploring 3D visual search, several studies using 2D stimuli that are interpreted by the visual system as 3D have revealed that efficient search can be based on such pictorial depth cues (Aks & Enns, 1992; Enns & Rensink, 1990; Grossberg, Mingolla & Ross, 1994; Humphreys, Keulers & Donnelly, 1994; Sousa et al., 2009).

Studies using stereoscopic displays to investigate 3D visual search have shown that efficient search – "pop out" – is possible when the target item lies in one depth plane and the distractors lie in another (Chau & Yeh, 1995; He & Nakayama, 1995; Nakayama & Silverman, 1986; O'Toole & Walker, 1997; Previc & Blume, 1993; Previc & Naegele, 2001; Snowden, 1998). However, these conclusions are based only on response time and error rate measurements.

Only one study has measured and analyzed eye movements during visual search in 3D stimuli. McSorley and Findlay (2001) presented observers with search items that were arranged in a circular pattern in the x-y plane, but at different perceived depths – induced by a shutter goggle system – around a central fixation marker. Eye tracking was used to determine saccadic latency and to identify which of the items was the target of the first eye movement. Whereas this study yielded insight into attentional processes with regard to the depth dimension, it did not focus on the analysis of 3D eye-movement trajectories. Moreover, similar to most other visual search studies, the virtual depth dimension did not reflect the continuous depth existing in the real world, but instead, only two depth levels were employed, serving as an object's binary "feature." In the natural world, however, depth is the third dimension of the search space in addition to its horizontal and vertical extent. Typically, when we search for an object, we do not know in advance its specific position in any of the three spatial dimensions, but we know its identity, at least roughly, in terms of its visual features.

The present study is the first to examine eye-movement trajectories in 3D visual search, in which the depth dimension is not a discrete search feature, but a multi-layered third dimension of the objects' locations. The most important information on 3D visual attention that we can acquire in such a scenario concerns strategy and capacity. Eccentricity effects, performance field asymmetries, and strategic scanning factors may extend into 3D space. We analyzed the strategy observers followed once the depth dimension is available; observers may tend to align their scanpaths along that dimension, rather than using their "default" 2D reading direction. This would be indicated by a bias toward consecutive items in the scanpaths being close to each other in the depth dimension. Such a finding would require us to revise our current understanding of strategic factors during search. The capacity of visual processing may also depend on the distribution of objects along the depth dimension.

We further analyzed bottom-up guidance of attention. According to the Guided Search theory (Wolfe, 1994; 1996; Wolfe et al., 1989), the more a search item differs from its neighbors, the more "conspicuous" it is, and in turn, the more likely it is to attract attention, even if it does not share any features with the search target. However, one important question is whether those neighbors have the same effect regardless of their depth position, or whether increasing the z-distance between objects reduces bottom-up activation in a way similar to that typically observed for x- and y-distances. These data provide a first assessment of the role of binocular disparity on visual search dynamics.

In the "easy" conjunction search task, observers showed a strong tendency towards scanning the display from the center outwards. Larger set size amplified the resulting eccentricity effect, reduced the selectivity of long saccades, increased initial saccadic latency, and increased fixation duration. The fixation duration effect may indicate a greater amount of information being processed per fixation and increased difficulty of saccade target selection. During the course of a search, fixation duration remained constant whereas saccade amplitude was initially shorter than later, further documenting the center-to-periphery strategy. The "difficult" shape search task, on the other hand, biased observers' scanning patterns towards their reading direction (left-to-right and top-to-bottom), with only a marginal eccentricity effect. Here, larger set size further emphasized the scanning bias but did not affect fixation duration or initial saccadic latency. Together with the finding of longer initial latency than in the easy task, these results indicate that, in the difficult task, observers initially position their gaze strategically prior to the actual search process. Furthermore, our data imply that the common use of initial saccadic latency as an indicator of task demands is problematic. Whereas previous literature assumed more difficult search tasks to induce greater latency (e.g., Boot & Brockmole, 2010; Võ & Henderson, 2010; Zelinsky & Sheinberg, 1995), our results suggest that particularly difficult search tasks can lead to systematic scanning strategies that start with a low-latency saccade towards the strategic starting position.

Experiments

Method

<u>Participants.</u> Four observers (2 females; mean age = 28.8 ys, SD = 7.9) participated in both experiments. They had normal visual acuity and, with the exception of one of the authors (M.P.) were unaware of the purpose of the study. The New York University IRB approved this study.

<u>Apparatus.</u> Stimuli were presented on a 21-inch Dell D992 monitor using a screen resolution of 1280 by 960 pixels and a refresh rate of 100 Hz. Eye movements were measured with an SR Research EyeLink-2k system (desktop mount) that–given the setup for the present study–provided an absolute error of measurement of about 0.2° of visual angle and a sampling frequency of 1000 Hz. Observers viewed the stimuli through a four-mirror stereoscope (OptoSigma Corp., Santa Ana, CA, USA) that was mounted on a chinrest placed at a distance of 48 cm from the monitor. The mirrors were adjusted so that the stimuli shown in the left and right half of the monitor display were projected onto corresponding retinal positions in the observers' left and right eye, respectively. The eye tracker camera was positioned between monitor and stereoscope so that it recorded an image of the left eye from a steep angle, in such a way that the pupil was not occluded by any of the mirrors. **Figure 1a** shows the experimental setup.

The mirror stereoscope was used due its virtual elimination of visual crosstalk, i.e., some visibility to the left eye of visual information intended for the right one, and vice versa. Crosstalk can interfere with the observer's percept and thereby create artifacts in the psychophysical data. The other common techniques for presenting stimuli with binocular disparity, shutter goggles and polarized filters, are typically affected by at least small amounts of crosstalk (Kooi & Toet, 2004; Ozolinsh, Andersson, Krumina, & Fomins, 2008). Shutter goggles block vision for one eye at a time, alternating between the eyes at a high frequency. The glasses are synchronized with the stimulus monitor in such a way that distinct stimuli can be presented to each eye. Alternatively, goggles with two orthogonally oriented, polarized filters can be used. The stimuli intended for each eye are then projected via light of the corresponding polarization. The polarization method avoids the flicker that shutter glasses require, but it is also affected by crosstalk.

The monocular eye-tracking setup allowed the assessment of gaze depth only via the known depth positions of the search items. Because items did not visually overlap, the measured 2D position allowed the identification of the currently fixated item. The virtual depth of that item was then taken as the current depth of the observer's gaze position. Author MP performed a verification procedure; he inspected 20 displays of 32 search items in a pre-specified item-by-item order. By assigning his fixations to the display item with the smallest Euclidean distance from it, 97.2% of his fixations were correctly assigned to the currently fixated item. No significant difference in this value across the eight depth levels was found. As the same error threshold for gaze measurement calibration was enforced for all four observers, the same level of accuracy can be expected for all of them. This assumption is further supported by similar values of saccadic selectivity for large set sizes – a measure that strongly depends on measurement accuracy – across observers (see **Figure 11** further below).

<u>Materials.</u> Each of the two experiments encompassed 1024 search trials plus eight practice trials. In each trial, the display showed two square-shaped frames (13.4° × 13.4°, 81.2 cd/m^2) side by side, consisting of a checkerboard pattern to facilitate the fusing of the two images into a single stereo image (see **Figures 1b and 1c**). The frames contained the 2D projections of a virtual 3D search stimulus for each of the two eyes. These stimuli consisted of 4, 8, 16, or 32 search items (0.8° diameter along their longest axis), half of which were red (CIE x = 0.68, y = 0.32) and the other half green (CIE x = 0.28, y = 0.60) at equal luminance (60.1 cd/m^2) on a black background

(0.01 cd/m^2). Every display included exactly one target item that was visually distinguishable from the other search items – the distractors.
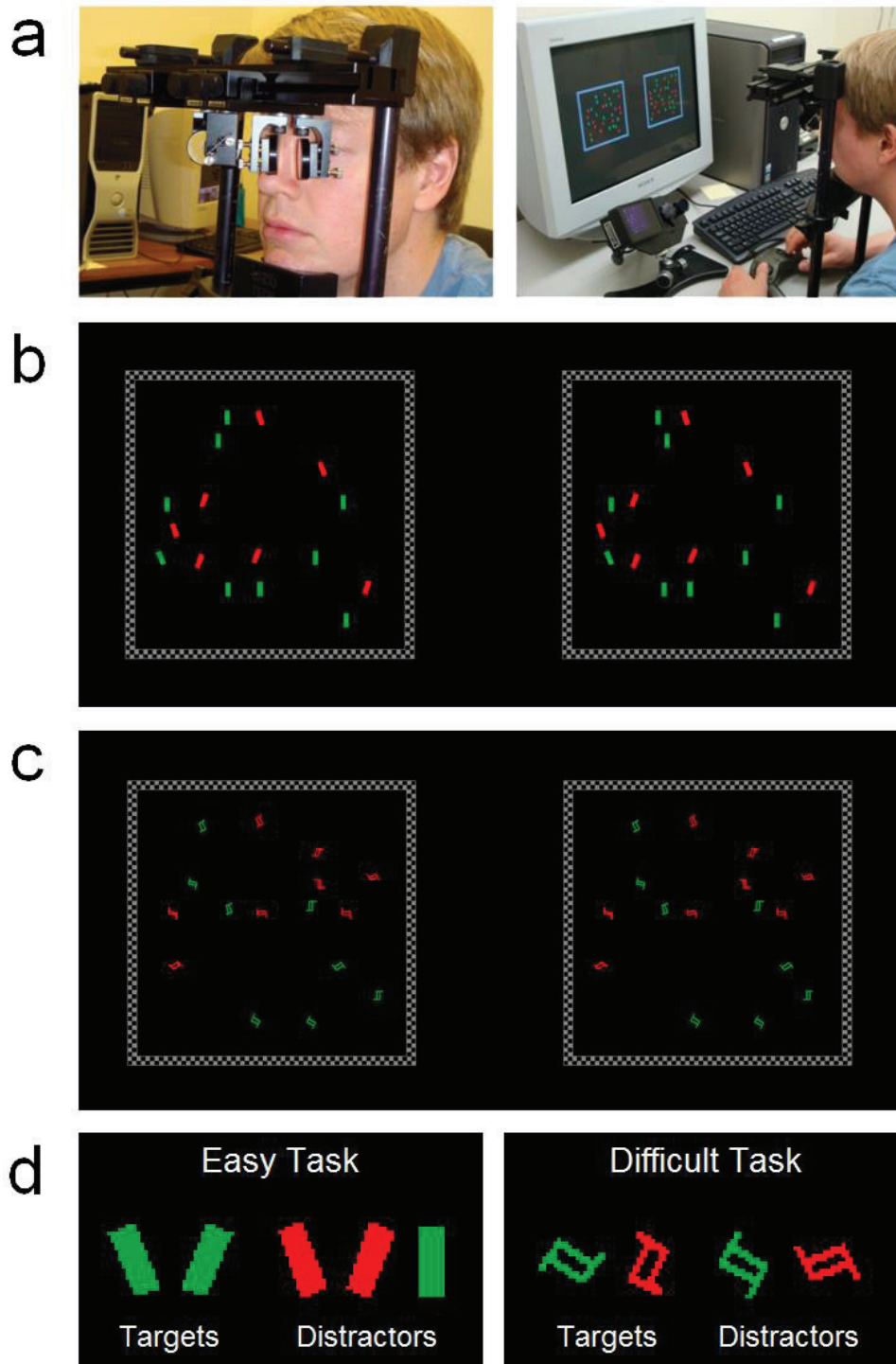

Figure 1: (**a**) Experiment setup of stereoscope and eye tracker for monocular eye tracking during the presentation of virtually three-dimensional search images; (**b**) Stimuli for the easy task – determine the tilt direction of the only tilted green bar; (**c**) Stimuli for the difficult task – determine the color of the only counterclockwise pointing item. Left and right stimulus panels are swapped to allow cross-eyed viewing of the three-dimensional stimuli. Readers may fuse the two panels by focusing their gaze on a point in front of the image in order to perceive the 3D stimulus as it was presented in the experiment. (**d**) Target and distractor items in the easy (left) and difficult (right) tasks.

In Experiment 1, the target was a green bar that was tilted 20° from vertical either to the left or to the right, whereas the distractors were vertical green bars and red bars tilted 20° from vertical to the left or right (**Figures 1b and 1d**). In Experiment 2, the target was a 'Z'-like figure of either red or green color, and the shape of the distractors was a mirror image of the target; all items were randomly oriented (see **Figures 1c and 1d**).

The virtual 3D display was divided into an invisible grid of 4×4×4 cells, and each search item was randomly assigned to one of the cells, with no cell containing more than one item. In order to minimize differences in the discriminability of items across the four depth layers, the visual angle subtended by the items did not vary with the items' virtual distance from the observer, as it would in a real 3D scenario. Consequently, the virtual grid did not form a cube but rather a pyramidal frustum – a square pyramid with its top "cut off" parallel to its floor – with the peak of the complete pyramid being located at the observer's optical center. In order to treat the three dimensions as equally as possible, the size of cells was chosen so that on the plane that divided the second and third depth layers, which coincided with the display surface, the horizontal (x) and vertical (y) cell sizes were equal to the depth (z) cell size. To prevent items from occluding each other, each cell was divided into eight (2×2×2) subcells, and each items was randomly positioned in one of these subcells. However, two items in horizontally and vertically aligned cells could not appear in subcells that matched in both their horizontal and vertical position, because it would have led to occlusion. The minimum distance between the centers of any two search items was set to 1.3° of visual angle. In the 1024 displays used in each of the two experiments, the target appeared exactly four times in each of the 4×4×4 display cells for each of the four set sizes (numbers of items).

Procedure. Each observer performed two sessions of each experiment, for a total of four 1-hr sessions. Each session started with the adjustment of the stereoscope and the calibration of the eye tracker. For this calibration, observers visually tracked a marker that appeared in nine different positions on a 3×3 grid within the stimulus frame and at the virtual depth of the physical monitor screen. Subsequently, observers were asked to repeat the procedure using a slightly different set of nine marker positions. This procedure was repeated until the average measurement error, i.e., the mean deviation between marker locations and measured gaze position fell to ≤0.2°. To verify that observers perceived the virtual depth in the stimuli, a test screen was presented prior to the search trials. This screen showed eight bar stimuli in different colors. Each stimulus was shown in one of the eight depth intervals used in the search displays, chosen randomly in such a way that each interval contained one stimulus. The observer was asked to name the colors of these stimuli in their order of depth, starting with the one closest to the observer. In the eight sessions performed in this study, incorrect orders were reported in only two cases; the mistake in both cases was a switching of the two most distant stimuli from the observer. This level of accuracy was considered to be sufficient for the study, given that the probability of reporting the correct sequence –except for switching two adjacent depth layers– purely by chance is only 1/5760. After completing eight practice trials, observers performed 512 experimental trials divided into four blocks of 128 trials. Each trial started with the presentation of a binocularly visible fixation marker in the center of the stimulus frame. Observers were instructed to fixate on that marker and press a button on a gamepad to start the trial. Once they had detected the target, they had to press one out of two buttons on the gamepad to respond.

In Experiment 1, observers reported whether the target was tilted to the left or to the right, and in Experiment 2 they reported whether the target was red or green. An acoustic tone indicated whether the response was correct. After the manual response, or if no response had been made within 30 seconds, the trial terminated and the fixation marker for the next trial was shown. The displays were identical for all observers, but their order of presentation was individually randomized.

Results

Overall performance.
As shown in **Figure 2a**, observers gave only a very small proportion of incorrect responses in both the easy task (0.9%) and the difficult task (3.2%). A two-way repeated measures ANOVA indicated a significant difference in error rate between the two tasks, $F(1; 3) = 30.1$, $p < 0.05$, but no effect of set size and no interaction of the two factors, both $Fs(3; 9) < 1.7$, $ps > 0.2$. Response time as a function of set size revealed the expected linear slope (**Figure 2b**), which was considerably steeper for the difficult task (approximately 164 ms/item) than for the easy task (approximately 29 ms/item), and the intercept was higher for the former (1.31 s) than for the latter (0.67 s). A two-way ANOVA showed significant main effects for task, $F(1; 3) = 1422.23$, $p < 0.001$, and for set size, $F(3; 9) = 297.52$, $p < 0.001$, and a task x set size interaction, $F(3; 9) = 154.16$, $p < 0.001$. **Figure 2c and d** indicate that the pattern of results was similar for the 4 observers (except for the error rate for set size 16 in the 'difficult task', consistent with the larger error bar for that data point).
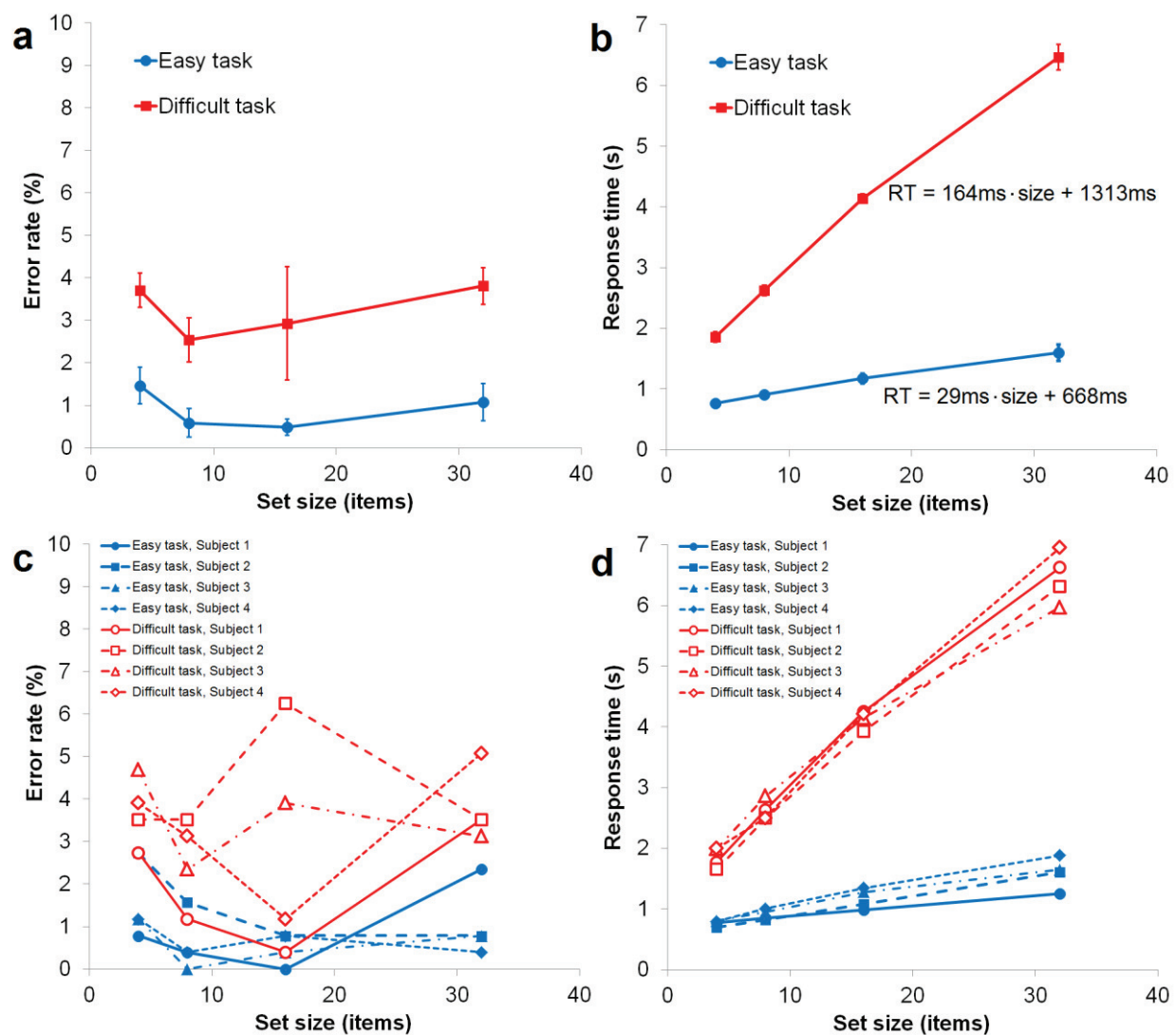

Figure 2: (**a**): Error rate and (**b**): response time as function of set size (4, 8, 16, or 32 items) in Experiment 1 (easy task) and Experiment 2 (difficult task). Error bars indicate standard error of the mean. (**c**): Error rate and (**d**): response time for individual observers.

Saccadic step size.
In order to determine the extent to which observers structured their search along each of the three spatial stimulus dimensions, we introduced a new measure named "saccadic step size." The idea underlying this measure is that a more structured scanning pattern in a given dimension will result in saccades with smaller amplitude in that dimension. For example, in the current experiment, let us assume that the observers' strategy was to scan the displays systematically from top to bottom (along the y-dimension) and not structure their search along the other two dimensions. If observers proceeded perfectly systematically, they would first scan all relevant objects in the first row, followed by all of those in the second row, and so on. This gaze behavior would obviously involve saccades that are relatively short in the vertical dimension, as compared to the other dimensions in which gaze targets are selected randomly.

To quantify this effect, we assigned each saccade to the coordinates of the closest display item. This step was necessary because our monocular eye-tracking setup allowed us to measure the depth (z-value) of the current gaze position only through the z-coordinate of the currently inspected display item. Subsequently, we computed saccadic step size as the average saccade amplitude in terms of the cells of the display grid in each spatial dimension. For example, a saccade whose start- and end- points were in the same row had a step size of zero in the y-dimension. A saccade switching to a neighboring depth layer had a step size of one in the z-dimension. It is important to note that, due the division of each cell into 2x2x2 subcells, the step size measure treated all three dimensions equally. For instance, saccades could have step sizes of zero in the x- and y-directions and at the same time a step size greater than zero in the z-direction. The maximum step size for a saccade in any dimension was three, because there were four intervals in each dimension.

The resulting values had to be compared against a baseline value for random (unstructured) saccades in order to decide whether there were significant effects of scan path structuring in a given dimension. To compute this baseline, two cases had to be distinguished: First, the starting point of the saccade could be at the extremes, that is, positions 1 or 4, in the given dimension. Then the possible step sizes were 0, 1, 2, and 3, which should occur with equal probability, leading to an expected step size value of 1.5. Second, the starting point could be near the center, that is, positions 2 or 3. In that case, the possible step sizes were (for position 2, with target positions in ascending order), 1, 0, 1, and 2, for an expected value of 1. Because for random saccades these two cases should occur equally often, the overall expected step size, i.e., the baseline value for random saccades, was 1.25. This value was confirmed by computer simulation for the displays used in the current study, with deviations below 1%.

Figure 3 shows saccadic step size as a function of set size in both visual search tasks. Whereas the x- and y-dimensions show average step sizes that are clearly below the baseline, both $t(3) > 17.25$, $ps < 0.001$, indicating structured search along these dimensions, no such effect was found in the z-dimension, $t(3) < 1$. Two-way repeated measures ANOVAs for each of the three stimulus dimensions revealed significant set size effects, all $Fs(3; 9) > 8.84$, $ps < 0.01$. This result demonstrates that saccadic step size decreased along all three dimensions with a greater number of display items, even though the absolute differences in the z-dimension were very small. Furthermore, there were significant task effects for the x- and y-dimensions, both $Fs(1; 3) > 27.69$, $ps < 0.05$, indicating that saccadic step size was smaller in the difficult task than in the easy task. For all three dimensions, significant interactions between task and set size, all $Fs(3; 9) > 7.86$, $ps < 0.05$, showed that the step size disparity between the tasks increased with greater set size. In summary, the observers' visual scan paths were more structured in the x- and y-dimensions with greater set size. This finding can be explained by the fact that systematic scanning strategies are more beneficial for larger set sizes because the search items can be aligned with such strategies more easily and revisiting previously inspected display areas is more costly.

Because step size along the z-dimension did not differ from the baseline but showed a significant set size effect and a task x set size interaction, we conducted post-hoc comparisons to determine possible conditions with significant deviations of z-step size from baseline. Only in the most demanding condition, the difficult task with 32 items, there was a tendency toward a step size below 1.25, $t(3) = 2.11$, $p = 0.11$. Inversely, in the least demanding condition, the easy task with 4 display items, step size for the z-dimension was significantly above the baseline (1.37), $t(3) = 3.23$, $p < 0.05$. The pattern of results was highly consistent for the four observers, for both the easy task (**Figure 3c**) and the difficult task **(Figure 3d).**
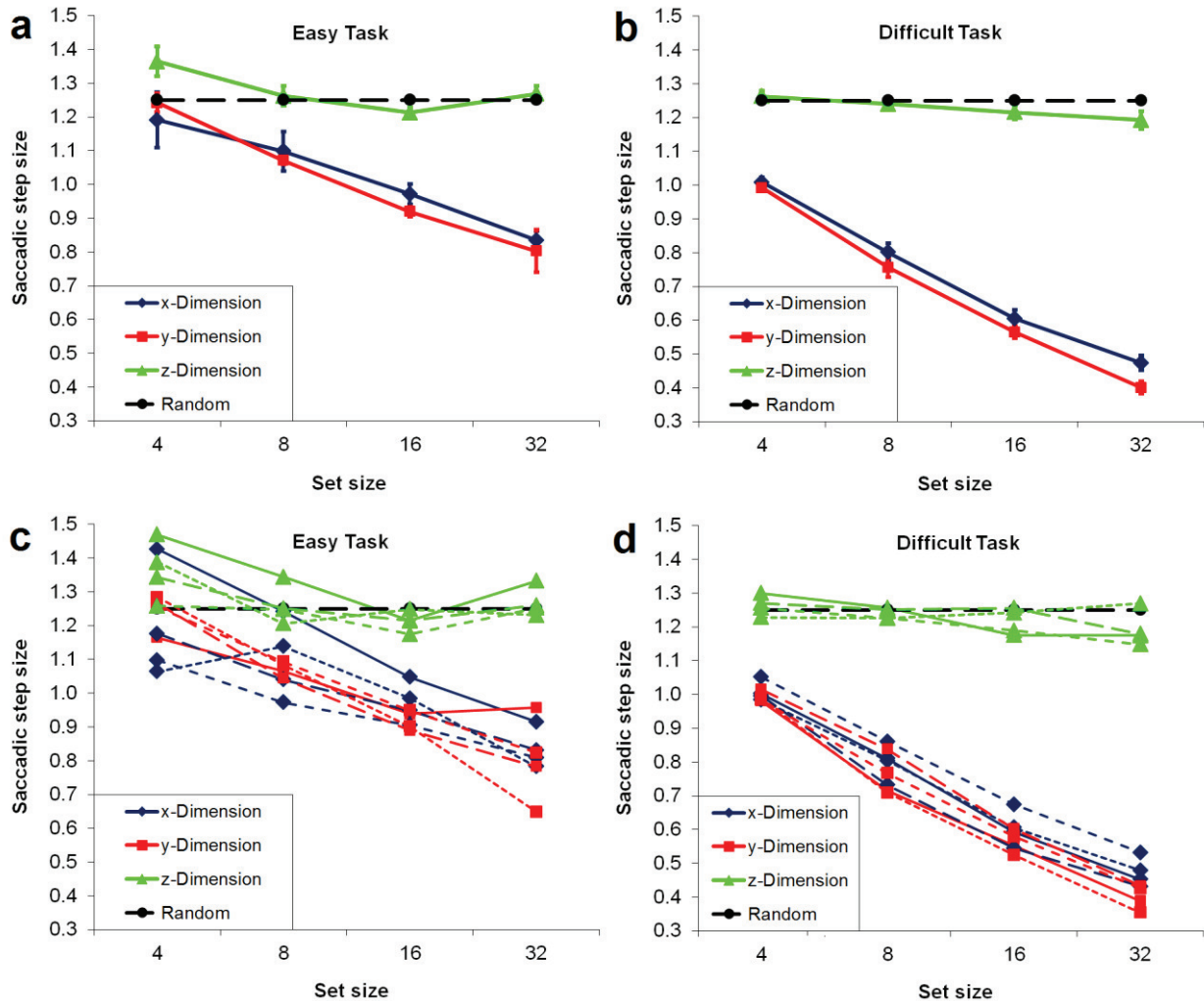

Figure 3: Saccadic step size along each of the three spatial stimulus dimensions, as compared to a baseline for random saccades. Error bars indicate standard error of the mean. (**a**) Mean step size in easy task; (**b**) mean step size in difficult task; (c) individual step size in easy task; (d) individual step size in difficult task.

Given the definition of the step size measure, a value significantly above the baseline seems counterintuitive. However, this result can be explained in terms of the two cases for baseline computation discussed above: If a saccade starts at an eccentric position, its expected step size is 1.5, and if it starts in a central position, it is 1. Consequently, if a disproportionate number of scan paths started from an eccentric z-position, i.e., either from the front layer or the back layer of cells,

an overall step size average of above 1.25 could occur. To test this hypothesis, we computed the proportion of first fixated objects in each trial in each of the four depth layers. Post-hoc analyses showed that in the easy task with four items, there was an above-chance probability of observers to first fixate an object in the front layer (30.1%), $t(3) = 4.08$, $p < 0.05$, whereas there were no such effects for other layers or set sizes in either the easy or the difficult task.

In order to show that this dependency of the saccadic step size measure on the landing point of the initial saccade did not systematically influence the results, we compared step size between the first three and all remaining saccades. Furthermore, we repeated all step size analyses including only trials in which exactly three saccades were made. None of these analyses showed any patterns that differed significantly from the pattern of results shown in **Figure 3**. To illustrate this, **Figure 4** shows the results of the same analysis as in **Figure 3** but includes mean step size for only the three initial saccades from each trial. **Figure 4c,d** indicate that the pattern of results was similar for the four observers in the easy task but showed considerable variance for the z-dimension in the difficult task, suggesting that some observers may use the z-dimension to structure their search for larger set sizes whereas others do not. It is important to note, however, that only a small proportion of trials for the difficult task and large set sizes had recordings with exactly three saccades, and therefore no statistical inference is possible.

Besides this first-fixation effect, no other influences of the virtual z-positions of search items on the search dynamics were found. For the benefit of conciseness, the corresponding analyses that were conducted will only be briefly described without detailed results. First, the effect of the z-position of the target item of response time was examined. An effect would suggest some form of systematic scanning of the display in the z-dimension. Second, RT was analyzed for different levels of the variance of search item positions along the z-dimension. If shifting spatial attention in the z-dimension requires additional time, then greater z-variance among all search items should lead to longer RT. Third, fixation duration in the easy task was analyzed as a function of the variance in z-position of the objects near fixation (in adjacent cells on the 8×8 grid) in the x-y-plane. If the visual span were limited in the z-dimension, a smaller z-variance would allow observers the processing of more objects during fixation, which should be reflected in longer fixation duration. Fixations followed by saccades shorter than two cells on the 8×8 grid were excluded from analysis to minimize potential interference of this measure with longer saccade planning processes due to greater z-variance. Fourth, the likelihood of fixations landing on "singletons," e.g., a red item surrounded by only green items, was compared between the situation when the surrounding items were on the same or on different z-planes than the singleton. Bottom-up control of attention tends to direct a disproportionate amount of attention to singletons ("pop-out effect," e.g., Wolfe, 1998), and therefore a greater effect for items on the same z-plane would indicate a perceptual z-distance limitation for pop-outs. Fifth, we compared the duration of fixations before saccades that switched between z-planes and those that did not. Longer fixation duration before z-plane switches would suggest a longer process of target selection and saccade programming. None of these five analyses yielded significant results or tendencies, all $ps > 0.2$.
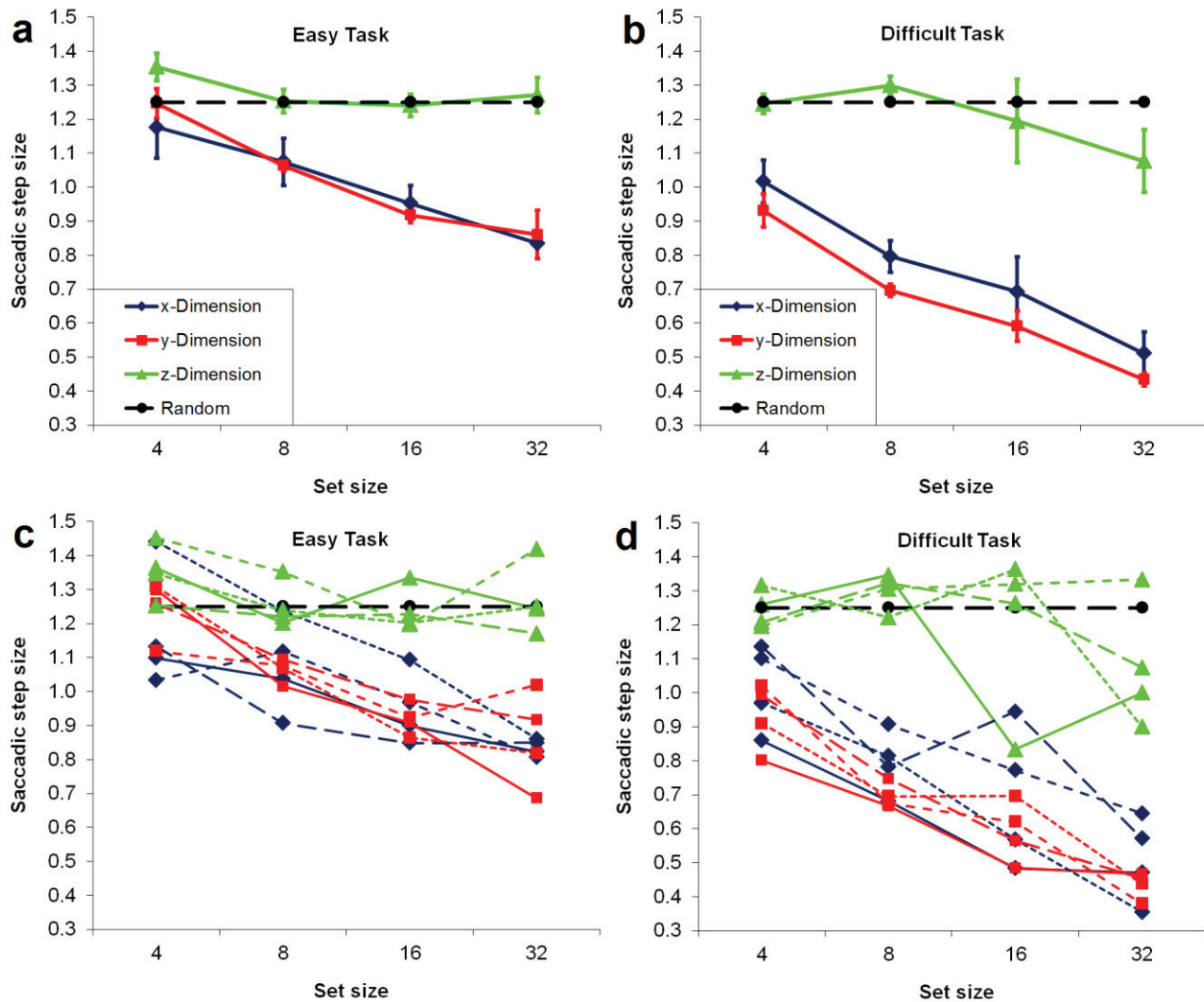
Figure 4: Step size of the first three saccades per trial along each of the three spatial stimulus dimensions, as compared to a baseline for random saccades. Error bars indicate standard error of the mean. (**a**) Mean step size in easy task; (**b**) mean step size in difficult task; (c) individual step size in easy task; (d) individual step size in difficult task.

In summary, systematic scanning behavior is clearly evident in the x- and y-dimensions, but does not occur to a significant extent in the z-dimension. The only effect of the search items' binocular disparity revealed by the current data is a stronger attraction of initial fixations by items in the front plane. This effect might be related to bottom-up control of attention, as it is only observed in the least demanding search condition, which is presumed to emphasize bottom-up effects over strategic planning (Zelinsky, 1996). Our data did not reveal any other perceptual or attentional effects of the z-dimension on the search dynamics.

Given that systematic scanning patterns were found in the x-y-plane but not along the z-dimension, further analysis – disregarding the z-dimension – was performed to characterize this behavior in more detail for the different task demands imposed by the experimental conditions. As explained above, the monocular eye tracking setup for the current study required us to perform the analysis of 3D gaze via the display item. The monocular gaze information was used to identify the currently fixated display item, whose 3D coordinates then indicated the 3D gaze position. When excluding

the z-dimension from analysis, we are no longer required to assign fixations to objects but can analyze their actual coordinates in the x-y-plane as measured directly by the eye-tracking system.

Distribution of 5 first fixations and saccade frequencies

Figure 5 presents cumulative scatter plots of the positions of all four observers' first five fixations in each trial, separated by task and set size. Besides the expected slower progression with greater set size, these scatter plots reveal a fundamental strategic difference between the easy and difficult tasks: In the easy task, search seems to move from the center (where observers fixate at stimulus onset) to the periphery. Particularly, in the 32-item condition, the central distribution of initial fixations and the ring (or "donut") shaped distribution of later fixations illustrate this point. The processing of display items seems to start during the initial, central fixation so that these central items do not usually have to be revisited later in the search. This is analogous to the eccentricity effect found with covert attention, which is also more pronounced with greater set size (Carrasco & Chang, 1995; Carrasco & Frieder, 1997; Carrasco et al., 1995, 1998; Geisler & Chou, 1995). The difficult task, in contrast, induces a bias of the initial fixations toward the upper left, while later fixations demonstrate a swipe towards the lower right. Given the steep search slope in the difficult task, it is not surprising that the progression of this systematic left-right and top-down scanning pattern greatly depends on set size.
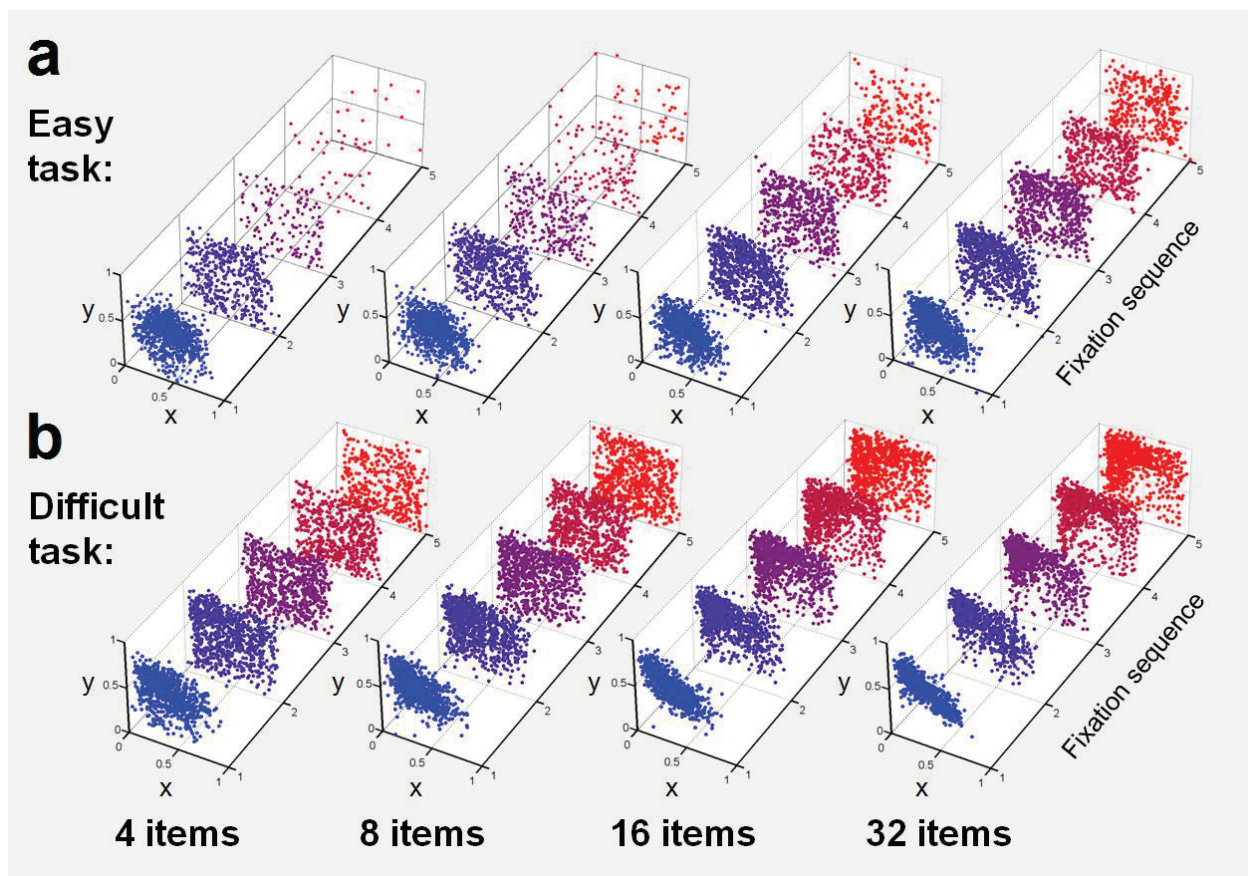

Figure 5: Distribution of the first five fixations in each trial in the x-y-plane for the easy (**a**) and difficult (**b**) tasks and the four set sizes. Note that the depth planes illustrate the fixation sequence and not the z-dimension in the stimuli. First fixations are shown in blue (lower right) and fifth fixations in red (upper left) in each sequence, with fixations number two to four shown in intermediate colors. Fixation coordinates are normalized in such a way that the stimulus area subtends the interval from zero to one in each dimension.

Furthermore, it is clearly visible in **Figure 5** that the number of fixations in a trial strongly depends on task difficulty and set size. For example, performing the easy task with four search items rarely requires five fixations, whereas searching through 32 items in the difficult task is not typically completed within five fixations. To quantify this effect, **Figure 6** shows saccade frequency histograms for all experimental conditions. As expected, distributions are progressively skewed towards larger numbers with greater set size and task difficulty.

Figure 5 also suggests that most eye movements were not directed to compensate the different sensitivity reflected in performance fields, according to which more saccades should have been directed to the upper vertical meridian, followed by the lower vertical meridian, then by the locations in the middle of the quadrants and then by locations along the horizontal meridian. Instead, we found a bias of initial saccades towards the upper-left display quadrant. In the easy task, the upper-left quadrant received 46.1% of the initial saccades, followed by the upper-right (20.9%), lower-right (17.5%), and lower-left (15.6%) quadrants. In the difficult task, 63.0% of initial saccades landed in the upper-left quadrant, followed by the upper-right (17.4%), lower-right (10.0%), and lower-left (9.7%) quadrants. A two-way ANOVA with the factors task and quadrant did not reveal an effect of task, $F(1; 3) < 1$, but showed a significant effect of quadrant, $F(3; 9) = 4.00$, $p < 0.05$, indicating a spatial bias. No interaction between the factors was found, $F(3; 9) = 1.55$, $p > 0.2$. Pairwise t-tests only revealed a trend toward more initial saccades landing in the upper-left quadrant than in the lower-left one, $t(3) = 1.55$, $p = 0.07$.
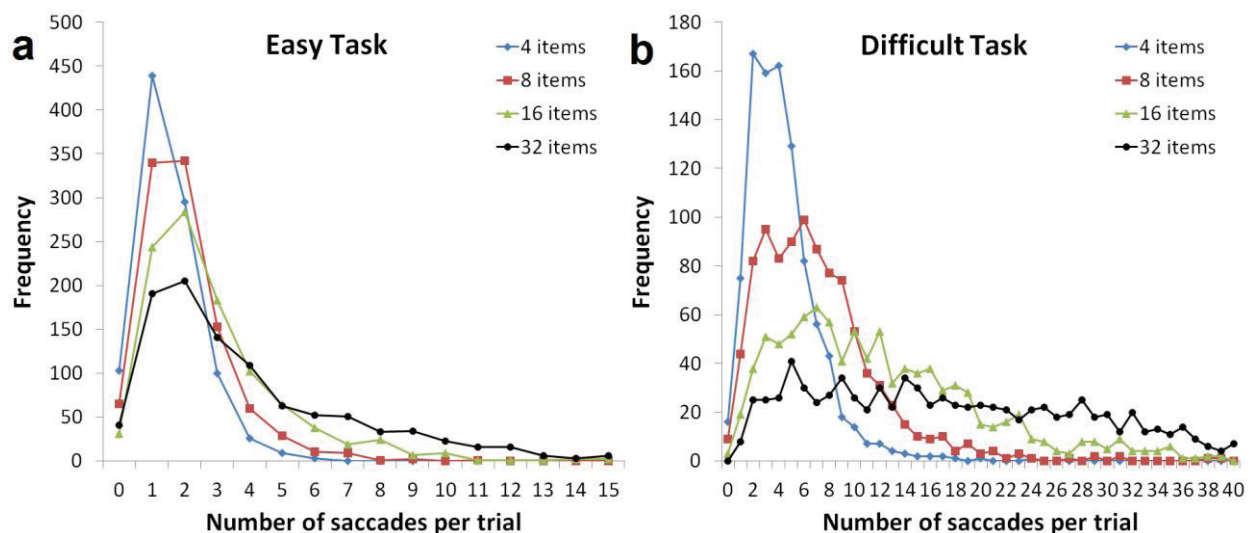

Figure 6: Saccade frequency histograms for the easy (**a**) and difficult (**b**) tasks and the four set sizes. Note the different scales in both axes of the two charts.

Response time as a function of x-y target position

In order to examine search strategies in more detail, we analyzed response time as a function of target position in the x-y-plane. As a first step, the effect of the horizontal and the vertical target position were analyzed separately. For both dimensions, the results demonstrate a tendency toward a U-shaped function in the easy task (**Figure 7a**) and an uphill slope in the difficult task (**Figure 7b**). These tendencies support the eccentricity effect observed in the easy task and the systematic left-right and top-down scanning strategy found in the difficult task. Clearly, these effects are more emphasized with larger set sizes. The "U"-shape and its interaction with set size found in the easy task are akin results obtained when observers maintained fixation (e.g., Carrasco

et al., 1995, 1998; Carrasco & Chang, 1995) and when they were allowed to move their eyes (Carrasco et al., 1995).
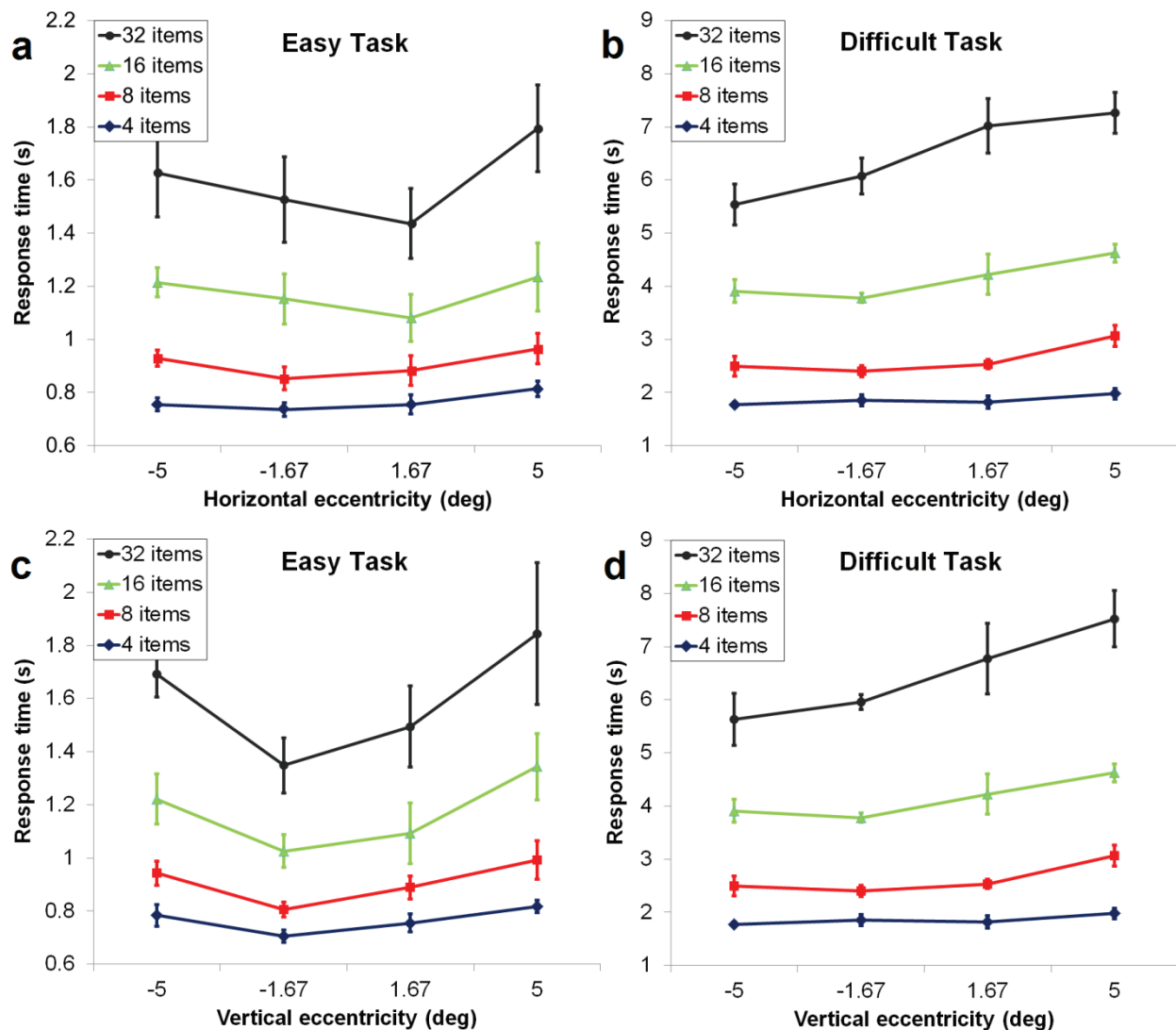

Figure 7: Response time as a function of horizontal (**a, b**) and vertical target eccentricity (**c, d**) in the easy (**a, c**) and difficult tasks (**b, d**) for different set sizes. Horizontal eccentricity represents the four horizontal positions in the display grid from left (negative) to right (positive), and vertical eccentricity represents vertical positions from top (negative) to bottom (positive). Error bars show standard error of the mean. Note the different y-scales for the two tasks.

Given that the previous response time results were similar for the x- and y-position of the target, we took two approaches to encode the two-dimensional target position to facilitate the analysis of the two different search strategies. First, target position was represented by the city-block distance – Manhattan distance – between the target and the display center; i.e., distance in vertical plus horizontal grid positions. As a result, distance 1 was assigned to the four central positions, distance 2 to the eight peripheral positions that exclude the display corners, and distance 3 to the four corner positions. Eccentricity (center-periphery) effects should be reflected by response times that increase with greater distance. Second, target position was represented by the city-block distance between the target and the upper-left display position. In other words, this distance increases with

the target being located further toward the right or the bottom of the display. Whereas the city-block distances range from 0 (upper-left corner) to 6 (lower-right corner), in our measure we collapsed distances 0 and 1 and distances 5 and 6 to increase the statistical power for these more rarely occurring values, leading to a scale from 1 to 5.

The four panels of **Figure 8** illustrate the resulting RT functions for the easy and hard tasks based on each of the two distance measures. As could be expected from the inspection of the raw data, RT shows a greater dependence on the center-based measure in the easy task, and a greater dependence on the upper-left based measure in the difficult task. Moreover, this pattern of results is more pronounced for greater set sizes. For the center-based measure, a three-way repeated measures ANOVA with the factors task, set size, and distance showed a three-way interaction trend on RT, $F(6; 18) = 2.52$, $p = 0.06$. Individual two-way ANOVAs for each of the two tasks revealed a significant set size x distance interaction for the easy task (**Figure 8a**), $F(6; 18) = 3.26$, $p < 0.05$, but not for the difficult one (**Figure 8b**), $F(6; 18) = 1.36$, $p > 0.2$. Furthermore, the easy task showed a significant main effect of distance, $F(2; 6) = 17.17$, $p < 0.005$, whereas the difficult task did not, $F(2; 6) = 1.68$, $p > 0.2$.

Similarly, for the upper-left based measure, a three-way ANOVA revealed a significant three-way interaction, $F(12; 36) = 3.19$, $p < 0.005$. Individual two-way ANOVAs for the two tasks revealed a set size x distance interaction for the difficult task (**Figure 8d**), $F(12; 36) = 3.82$, $p < 0.005$, but not for the easy task (**Figure 8c**), $F(12; 36) < 1$. Moreover, whereas the difficult task showed a main effect of distance on RT, $F(4; 12) = 5.59$; $p < 0.01$, no such an effect emerged for the easy task, $F(4; 12) = 1.35$, $p > 0.3$, . Error rate differed between the two tasks as reported above but did not reveal effects by any other factors.
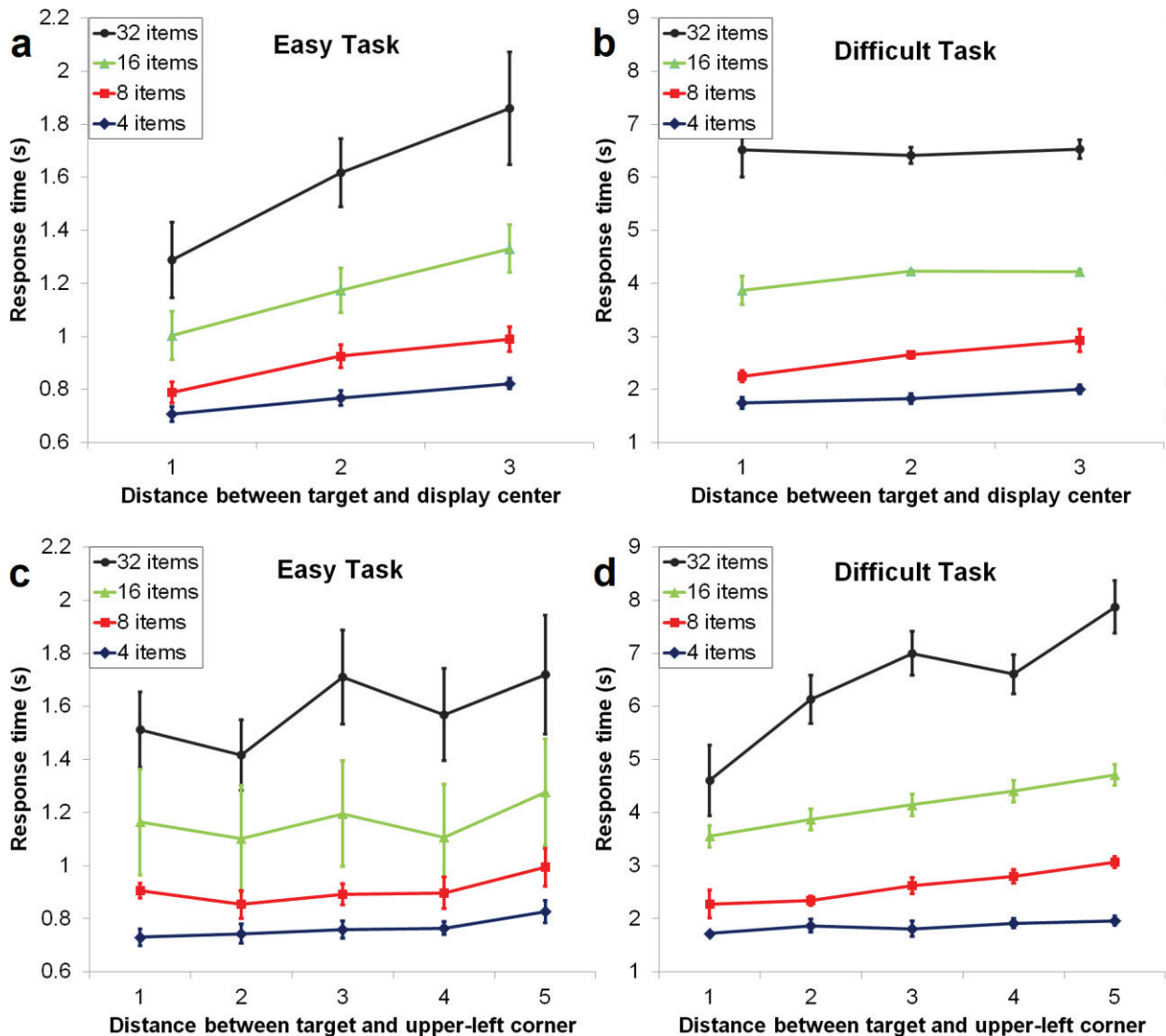
Figure 8: Response time as a function of target position in the easy task (**a, c**) and the difficult task (**b, d**). Target position is represented by the city-block distance between the target and the display center (**a, b**) or between the target and the upper-left display corner (**c, d**). Note the different y-scales for the two tasks.

For a more fine-grained investigation of the underlying mechanisms, it is useful to analyze three basic eye-movement variables – initial saccade latency, mean fixation duration, and saccade amplitude. Initial saccadic latency refers to the time from stimulus onset to the start of the first saccade. It reflects the amount of visual processing during the initial fixation as well as the difficulty of selecting the first saccade target (Bichot & Schall, 1999; Pomplun, Reingold & Shen, 2001; Zelinksy, 1996). Fixation duration during visual search mainly reflects the amount of information being processed during a fixation and the time needed for programming the following saccade (e.g., Hooge & Erkelens, 1998). In our analysis, we excluded both the initial, central fixation that started prior to stimulus onset and the final fixation that is typically influenced by target verification processes.

Initial saccadic latency and fixation duration
Initial saccadic latency was higher for the easy task (203 ms) than for the difficult one (141 ms), with only a marginally significant difference, $F_{(1, 3)} = 7.65$, $p = 0.07$ (**Figure 9a**). Furthermore, latency increased with greater set size, $F_{(3, 9)} = 5.07$, $p < 0.05$, with a more pronounced set size

effect for the easy task than for the difficult one, as indicated by a task x set size interaction, $F_{(3; 9)}$ = 8.28, $p < 0.01$. Conversely, fixation duration was longer during the difficult task (234 ms) than during the easy one (188 ms), $F_{(1; 3)} = 47.53$, $p < 0.01$ (**Figure 9b**). Greater set size led to longer fixation duration, $F_{(3; 9)} = 86.72$, $p < 0.001$, with a more pronounced effect in the easy task than in the difficult task, as revealed by a significant task x set size interaction, $F_{(3; 9)} = 9.92$, $p < 0.005$.
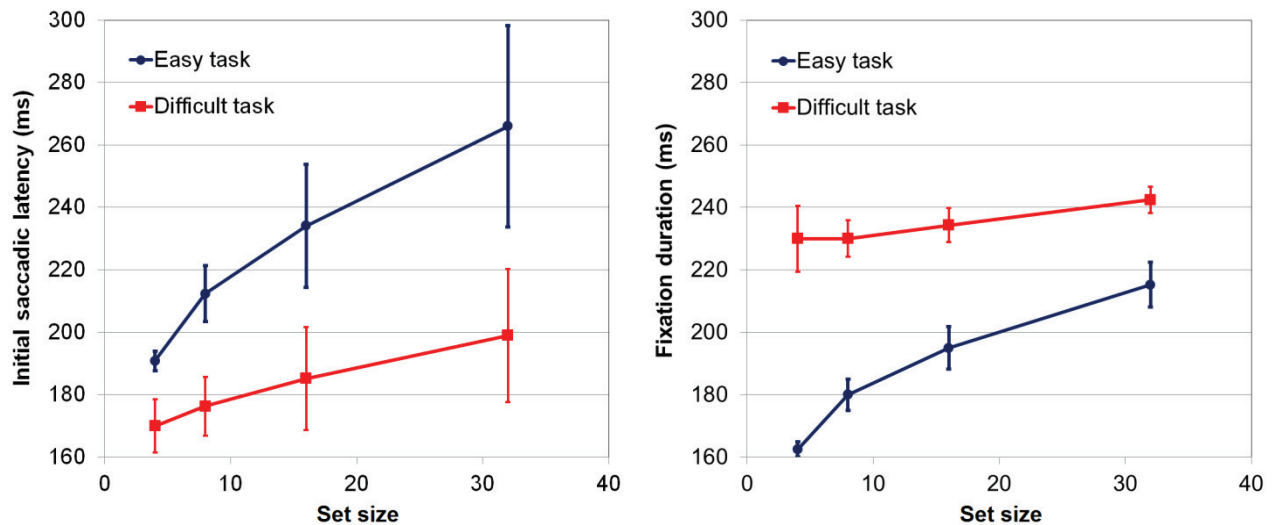

Figure 9: (**a**) Latency of the initial saccade in a trial and (**b**) mean fixation duration (right panel) as a function of set size in Experiments 1 and 2.

These results suggest that in the easy task, processing of the display already starts during the initial, central fixation. This is indicated by the large initial saccadic latency and its substantial increase with greater set size, i.e., greater density of items available for processing in the central display area. Furthermore, fixation duration in the easy task strongly increases with growing set size, suggesting that more information is processed during a single fixation when the density of search items is higher. Whereas the visual span adapts at least partially to stimulus density (see Bertera & Rayner, 2000), observers still tend to make more saccades when density is lower in order to foveate relevant search items.

In contrast, the low saccadic latency in the difficult task suggests that observers do not typically process any central items during the initial fixation but instead execute a saccade to a suitable starting point for their more systematic search strategy. The slight increase in latency with greater set size might be due to a slightly more complex selection process for the first saccade target when more display items are present. During the systematic search process, fixation duration does not significantly depend on set size, regardless of the density of search items near fixation. The generally long fixations reflect the greater difficulty of distinguishing targets from distractor items.

Saccade amplitude and fixation duration
Were these conclusions correct, we would also expect specific differences between the tasks in the time courses of eye-movement variables, especially saccade amplitude and fixation duration. In the easy task, the immediate onset of visual processing in the display center should lead to initial saccades that are shorter than later ones. Furthermore, because all fixations are assumed to involve processing of display information, their duration should not vary strongly during the course of a trial. In the difficult task, on the other hand, if the observers' initial goal is to move their gaze to a strategic starting position for the search, their initial saccade should tend to be longer than the

following ones. For the same reason, the first fixation should be shorter than later ones, as it is less likely to involve the processing of display items for target detection.

Figure 10 illustrates the results of this analysis by showing mean saccade amplitude and fixation duration for the first three fixations per trial in each task. These results mainly confirm the above assumptions. A two-way ANOVA of saccade amplitude revealed a significant task x sequence interaction (**Figure 10a**), $F_{(2; 6)} = 7.05$, $p < 0.05$. Individual one-way ANOVAs showed that in the easy task, saccade amplitude increased as the sequence progressed, $F_{(2; 6)} = 25.16$, $p < 0.005$, whereas there was no such effect in the difficult task, $F_{(2; 6)} < 1$. A corresponding two-way ANOVA of fixation duration also showed a significant task x sequence interaction (**Figure 10b**), $F_{(2; 6)} = 24.97$, $p < 0.005$. Individual one-way ANOVAs demonstrated that fixation duration increased as the sequence progressed in the difficult task, $F_{(2; 6)} = 17.22$, $p < 0.005$, but no significant sequence effect on fixation duration in the easy task, $F_{(2; 6)} = 3.34$, $p > 0.1$.
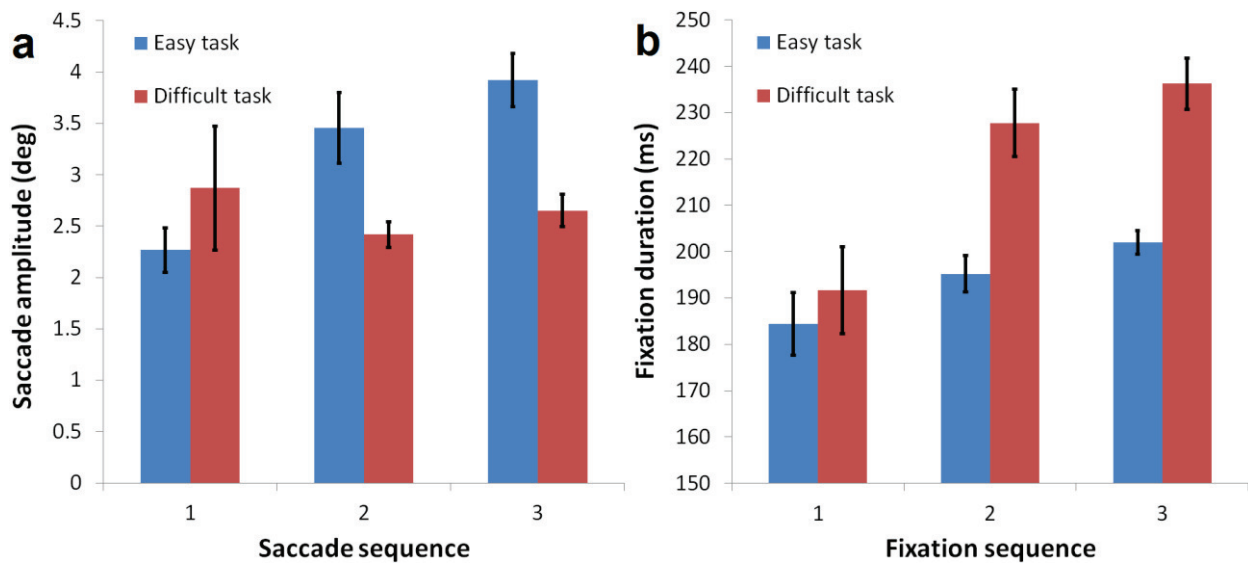

Figure 10. (**a**) Saccade amplitude and (**b**) fixation duration for the first three saccades or fixations per trial, respectively, in each of the tasks. Error bars indicate standard error of the mean.

It is possible to directly test our hypothesis whether in the easy task the observers' gaze was guided by extrafoveal stimulus features. Previous studies have shown that the color of the search target typically dominates the guidance of eye movements during search, as indicated by a disproportionate amount of saccades targeting display items that share their color with the target (Hwang, Higgins, & Pomplun, 2009; Williams & Reingold, 2001). Such guidance would elicit a bias of fixation positions towards green distractors. To investigate this, each recorded saccadic endpoint was assigned to the nearest distractor in the display. We then computed relative saccadic selectivity, i.e., the ratio of fixated green vs. red distractors relative to the occurrence of these distractor types. Guidance of eye movements by target color would be indicated by relative selectivity values above the chance level of one. Short ($\leq 4°$) and long ($> 4°$) saccades were analyzed separately to examine the retinal range of visual guidance. The cut-off point of $4°$ was chosen to obtain a sufficient amount of data points for each saccade type across set sizes. Average relative selectivity (2.9) was clearly above chance level, $t_{(3)} = 9.14$, $p < 0.005$. A two-way ANOVA revealed a trend towards a set size x saccade amplitude interaction, $F_{(3; 9)} = 3.20$, $p = 0.077$. Whereas the selectivity of long saccades decreased with greater set size, the selectivity of short saccades remained relatively constant (**Figure 11**). These results demonstrate the guidance

of saccades by target color even at far retinal eccentricity. The decreasing selectivity of long saccades with a greater number of search items could be due to increasing targeting errors or a more pronounced center-of-gravity effect. Note that relative saccadic selectivity could not be computed in the difficult task because it did not include distractors of varying similarity to the search target.
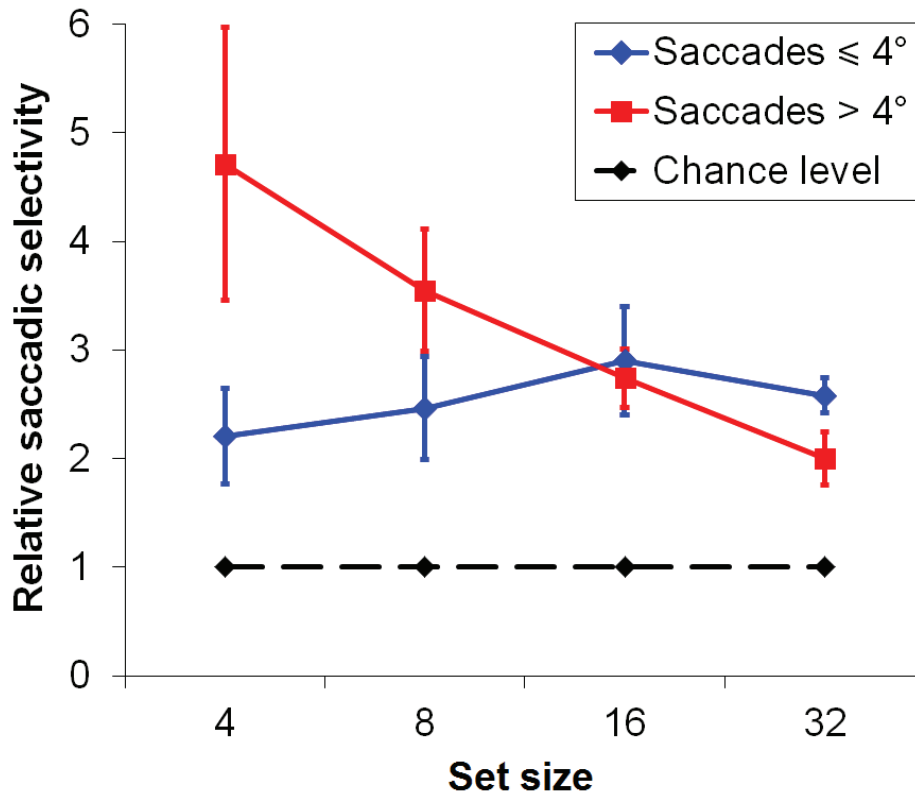

Figure 11. Relative saccadic selectivity for distractors of the same color as the search target (green), computed as the ratio of fixated green vs. red distractors divided by the ratio of green vs. red distractors occurring in the displays. Data series represent saccades $\leq 4°$, saccades $> 4°$, and chance level selectivity, which would be expected if saccades were not guided by visual distractor features. Error bars show standard error of the mean.

General Discussion

We studied visual search strategies in virtual 3D displays by combining the analysis of traditional variables –response time, error rate, fixation distribution, fixation duration, saccade amplitude, and initial saccadic latency– and novel ones –saccadic step size, fixation distance from reference point, and relative saccadic selectivity. The results revealed a consistent and cohesive pattern of visual search strategies as a function of task difficulty. Overall, this pattern was consistent across the four observers who participated in the study.

For the "easy task" –conjunction search– we found that observers tended to scan the display from the center towards the periphery, inducing an eccentricity effect, with a slight bias of initial fixations favoring the upper-left display quadrant. Increasing set size led to a more pronounced eccentricity effect, decreasing selectivity of long saccades, longer initial saccadic latency, and longer fixation

duration. These increased durations may reflect a greater amount of visual information processing and a more difficult saccade target selection when the density of items near fixation was greater. Whereas the duration of the first fixation did not vary from that of later fixations, saccade amplitude was smaller for the first saccade than for later ones, providing further evidence for a center-biased initial scanning process.

In contrast, for the "difficult task" –shape search– we found that observers scanned the display in a left-to-right and top-to-bottom manner and did not show an eccentricity effect. In larger displays, this scanning strategy was more emphasized, whereas initial saccadic latency and fixation duration were not significantly influenced. Initial saccadic latency was shorter in the difficult task than in the easy task, suggesting that observers in the difficult task initially moved their gaze to a strategic position before starting the actual search process. The saccadic step size analysis revealed that the depth dimension did not play a significant role in either of the tasks. Only in the least difficult condition – the easy task with four search items – a slight bias of initial saccades towards the front plane of the display was found.

The virtual depth dimension in the current study was created by binocular disparity alone. This binocular disparity was not found to substantially influence visual search strategy, performance, or any measures of eye movements and visual attention, such as the targeting of saccades. The only effect of the virtual depth dimension demonstrated by the present experiments was a slight increase in initial saccades directed to the depth plane closest to the observer. Given that this bias was only observed for the smallest displays in the easy task, it is likely that it is mainly driven by bottom-up control of attention, which is generally most pronounced in the initial saccades of easy search tasks. This bias was observed despite the fact that the pre-stimulus fixation marker appeared at a virtual depth between the second and third depth planes. A complete binocular eye movement to the first plane – the one closest to the observer – would thus require a vergence eye movement. However, during quick sequences of eye movements, vergence eye movements cannot be completed/executed, as this process may take up to approximately one second (Schor, 1979). In addition, the pattern of results with regard to step size showed considerable variance for the z-dimension in the difficult task, suggesting that some observers may use the z-dimension to structure their search for larger set sizes.

Whereas our virtual 3D stimuli induced the same geometrical vergence requirement for "perfect" binocular fixations as found in a natural 3D environment, our stimuli differed from natural visual scenes in many ways. It is possible that physical 3D search stimuli including monocular depth cues and the need for lens accommodation may induce depth effects that cannot be triggered by binocular depth alone. We conducted a preliminary control experiment in which we included one monocular depth cue – we reduced the size (subtended visual angle) of more distant search items according to the geometry of the virtual stimulus. No significant difference in any of the eye movement or performance measures was found. Based on these preliminary data, it seems that this monocular depth cue alone still does not lead to the depth dimension playing a significant role in visual search. More realistic search scenarios may be necessary to induce such an effect.

In the x-y-plane, on the other hand, a strong influence of search strategies, depending on task difficulty and set size, was found. In the easy task, observers' eye movements revealed a significant eccentricity effect that increased with greater displays. This pattern of results is consistent with previous studies of both feature and conjunction searches when observers maintain fixation and when they are allowed to move their eyes (e.g., Carrasco et al., 1995, 1998; Carrasco & Frieder, 1997). This effect can be modeled as a linear dependency of RT on the distance of the target from the display center. In the present experiments, error rates were low and did not vary significantly with the position of the search target. The data indicate that the eccentricity effect is induced by the observers' ability to process a significant amount of central search items during

their initial fixation and increase this amount with greater displays, i.e., higher density of items. Even if observers do not detect the target during the initial fixation, this initial processing can generate a stimulus-based hypothesis for the optimal first saccade target, which induces a center-to-periphery search strategy. In other words, observers seem to exploit their initial central gaze position for efficient guidance of their attention towards the search target. As shown by our analysis of relative saccadic selectivity, the color of distractor items, even if perceived at large retinal eccentricity, strongly contributed to this guidance. With increasing set size, selectivity decreased for long saccades but not for short ones, likely due to greater difficulty of saccade targeting and more frequent occurrence of center-of-gravity effects for long saccades. However, selectivity for both short and long saccades was clearly above chance level for all set sizes, which indicates a long retinal eccentricity range of attentional guidance. This long range facilitates the center-to-periphery search strategy, presumably making it the most efficient strategy for the easy task. Further evidence for this interpretation of the current data was provided by the analysis of saccade amplitude and fixation duration over the course of the search. Saccade amplitude was found to be shorter for the initial saccade than for the following ones, indicating that observers' gaze initially tends to remain near the display center instead of shifting to the periphery immediately. Furthermore, fixation duration did not vary significantly between the first and later fixations, suggesting that the processing of search items and the selection of the next saccade target already occur to a similar extent during the first fixation as during the following ones. In summary, the easy task does not seem to involve considerable strategic planning but starts from the initial gaze position and relies on mechanisms of attentional guidance for efficient task performance.

In the difficult task, observers seem to omit any central processing during the initial fixation for the benefit of a systematic search strategy. There are several pieces of evidence for this conclusion: First, landing points of initial saccades in the difficult task were particularly strongly biased towards the upper-left quadrant. Second, the latency of these saccades was shorter than that of later ones and did not significantly depend on set size, suggesting that the purpose of these saccades was to move the fovea to a strategically appropriate starting point rather than processing visual features of search items. Third, in contrast to the easy task, saccade amplitude in the difficult task did not increase between the first and later saccades, indicating no tendency of observers' gaze to initially remain near the display center. Fourth, RT increased linearly with the distance of the search target from the upper-left display corner, and the slope of this function was amplified by greater set size. Taken together, observers typically seem to start their search in the upper-left display corner and subsequently proceed to the right and downwards until they find the search target. Such a geometry-based, systematic search strategy is aimed at preventing the repeated scanning of the same search item, which would be expensive in a difficult task. Moreover, in a difficult task that requires foveation of search items for target detection, attentional guidance by parafoveal or peripheral information – as it is likely to occur in the easy task - is rather ineffective. In the present study, the strategies of all four observers (all of them right-handed and left-to-right readers) in the difficult task showed an initial bias toward the upper-left corner of the display. This effect can be modeled as a linear RT dependency on the distance between the upper-left display corner and the search target. Error rates in our experiment were not affected by this distance.

Neither the easy task –conjunction search– nor the difficult task –shape search– revealed that the frequency of saccades corresponded to the performance fields (e.g. Carrasco et al., 2001, 2004; Cameron et al., 2002). According to them, we would have expected more saccades to be directed to the central upper region than to the lower central region, and the least saccades directed to position close to the horizontal meridian in which sensitivity is the highest (Najemnik & Geisler, 2009). In contrast, particularly for the difficult search, we observed that the incidence of saccades was higher to the upper left quadrant than to the other three quadrants. This asymmetry is consistent with the dependency on the distance between the upper-left display corner and the search target discussed above. This finding suggests that the specific search strategies observed

in the present study – center-to-periphery guidance by extrafoveal stimulus features in the easy task, and systematic x-y scan paths in the difficult task – override the potential influence of performance fields on oculomotor control with the high-contrast, conjunction and shape stimuli we used.

Extending the interpretation of previous findings (e.g., Boot & Brockmole, 2010; Võ & Henderson, 2010; Zelinsky & Sheinberg, 1995), the current results demonstrate that initial saccadic latency is not a suitable measure for processing duration that is induced by task demands. In our study, saccadic latency was smaller in the more demanding task. The reason for the discrepancy between this and previous results (e.g., Pomplun et al., 2001; Zelinsky, 1996; Zelinsky & Sheinberg, 1995) is likely the particularly high demands by the difficult task of the present study. Whereas earlier investigations compared feature and conjunction search tasks, our experiments compared a conjunction search task and an especially demanding shape search task. For such level of difficulty, observers seem to be able to adopt an efficient routine of omitting any initial, central display analysis and quickly targeting a strategically suitable starting point for a systematic search.

The present study adds to our understanding of search strategies in 2D and 3D displays and the insight that eye-movement analysis can provide for this line of research. The limitations of our exploratory investigation with regard to the virtual 3D induced by binocular disparity illustrate the necessity for follow-up studies. Visual search in real-world scenes could inform us about the strategies that are naturally used in everyday life. Moreover, strategy measures such as saccadic step size or x-y target distance proposed by the current study need to be refined to ideally provide unified, general descriptors of search strategies. Such descriptors would not only advance the theory of visual search but also have immediate practical applications. For example, they could enable a better characterization of visual search anomalies that occur in medical conditions such as schizophrenia, Alzheimer's disease, or Parkinson's disease and thereby help to diagnose patients and better understand their conditions.

The present study also has implications for current technological developments. As 3D displays are becoming commonplace, developers are not only starting to use the third dimension to add realism to movies and games, but to develop more intuitive user interfaces that use depth to either increase user efficiency or interface intuitiveness. Understanding observers' use of search strategies in such displays will be crucial for making appropriate design choices. One particularly important area of application regards heads-up-displays in vehicles and aviation. Such displays do not use disparity to induce 3D, but are physically positioned in the viewer's line of sight. Developing user interfaces that minimize drivers' distraction while providing optimal information distribution will be a major focus for a variety of manufacturers. This process can greatly benefit from psychophysical research along the lines of the current study.

References

Abrams, J., Nizam, A., & Carrasco, M. (2012). Isoeccentric locations are not equivalent: The extent of the vertical meridian asymmetry. *Vision Research*, *52*(1), 70-78.

Aks, D.J. & Enns, J.T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception & Psychophysics*, 52, 63-74.

Bertera, J.H. & Rayner, K. (2000). Eye movements and the span of the effective visual stimulus in visual search. *Perception & Psychophysics*, 62, 576-585.

Bichot, N.P. & Schall, J.D. (1999). Saccade target selection in macaque during feature and conjunction visual search. *Visual Neuroscience*, 16, 81–89.

Boot, W. R., & Brockmole, J. R. (2010). Irrelevant features at fixation modulate saccadic latency and direction in visual search. *Visual Cognition*, 18, 481–491.

Cameron E.L., Tai, J., & Carrasco, M. (2002). Effects of covert attention on the psychometric function of contrast sensitivity. *Vision Research, 42,* 949-967.

Cameron, E.L., Tai, J.C., Eckstein, M.P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks: Identification, Yes/No detection and localization. *Spatial Vision, 17,* 295-325.

Carrasco, M., & Chang, I. (1995). The interaction of objective and subjective organizations in a localization search task. *Perception & Psychophysics*, *57,* 1134-1150.

Carrasco, M., Evert, D.L., Chang, I., & Katz, S.M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57,* 1241-1261.

Carrasco, M., & Frieder, K.S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, *37,* 63-82.

Carrasco, M., Giordano, A.M., & McElree, B. (2004). Temporal performance fields: Visual and attentional factors. *Vision Research*, *44,* 1351-1365.

Carrasco, M. & McElree, B. (2001). Covert attention speeds the accrual of visual information. *Proceedings of the National Academy of Science, 98,* 5363-5367.

Carrasco, M., McLean, T.L., Katz, S.M., & Frieder K.S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation & spatial frequency. *Vision Research*, *38,* 347-374.

Carrasco, M., Penpeci-Talgar, C., & Cameron, E.L. (2001). Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spatial Vision, 15,* 61-75.

Cave, K.R., & Wolfe, J.M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.

Chau, A.W. & Yeh, Y. (1995). Segregation by color and stereoscopic depth in three-dimensional visual space. *Perception & Psychophysics*, 57, 1032-1044.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, 9(2),111–118.

Eckstein, M.P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5):14, 1–36, http://www.journalofvision.org/content/11/5/14, doi:10.1167/11.5.14.

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

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37, 617-631.

Findlay, J. M. (2004). Eye scanning and visual search. In J. M. Henderson & F. Ferreira (Eds.), The Interface of Language, Vision, and Action: Eye Movements and the Visual World (pp. 135–159). New York: Psychology Press.

Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed), Eye guidance in reading, driving and scene perception (pp. 295-312). Oxford, Elsevier.

Geisler, W.S., & Chou, K.L. (1995). Separation of low-level and high-level factors in complex tasks: visual search. *Psychological Review*, 102(2), 356–378.

Giordano, A.M. , McElree, B. & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision,* 9 (3):30, 1–10.

Grossberg, S., Mingolla, E., & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review, 101, 470-489.*

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

Hooge, I.T.C. & Erkelens, C.J. (1998). Adjustment of fixation duration in visual search. *Vision Research, 38, 1295-1302.*

Hooge, I.T.C., & Erkelens, C.J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research*, 39, 1567-1575.

Humphreys, G.W., Keulers, N. & Donnelly, N. (1994). Parallel visual coding in three dimensions. *Perception*, 23, 453-470.

Hwang, A.D., Higgins, E.C. & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9 (5):25, 1-18.

Jacobs, A.M. (1987). Toward a model of eye movement control in visual search. In J.K. O'Regan & A. Lévy-Schoen (Eds.), Eye movements: From physiology to cognition (pp. 275-284). North-Holland: Elsevier Science Publishers.

Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays*, 25, 99–108.

Kristjánsson, A., & Sigurdardottir, H. M. (2008). On the benefits of transient attention across the visual field. *Perception*, 37(5), 747–764.

Luria, S.M., & Strauss, M.S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics*, 17, 303-308.

Mackeben, M. (1999). Sustained focal attention and peripheral letter recognition. *Spatial Vision*, 12(1), 51–72.

McElree, B. & Carrasco, M. (1999). Temporal dynamics of visual search: A speed-accuracy analysis of feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25,* 1517-1539.

McSorley, E. & Findlay, J.M. (2001). Visual search in depth. *Vision Research*, 41, 3487-3496.

Moher, J., Abrams, J., Egeth, H. E., Yantis, S., & Stuphorn, V. (2011). Distractor probability modulates oculomotor capture: Trial-by-trial adjustments of topdown set. *Psychonomic Bulletin & Review,* 19(5), 897–903.

Montaser-Kouhsari, L. & Carrasco, M. (2009). Perceptual asymmetries are preserved in short-term memory tasks. *Attention, Perception, & Psychophysics*, *71*(8), 1782-1792.

Morvan, C. & Maloney, L. T. (2012), Human visual search does not maximize the post-saccadic probability of identifying targets. *PLoS Computational Biology, 8(2), e1002342*, 1-11.

Motter, B.C., & Belky, E.J. (1998). The guidance of eye movements during active visual search. *Vision Research*, 38, 1805-1815.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3), 4, 1–14.

Najemnik, J., & Geisler, W.S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 49(10), 1286–1294.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*, 51, 1526–1537.

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

O'Toole, A. & Walker, C.L. (1997). On the preattentive accessibility of stereoscopic disparity: Evidence from visual search. *Perception & Psychophysics*, 59, 202-218.

Ozolinsh, M., Andersson, G., Krumina, G, & Fomins, S. (2008). Spectral and temporal characteristics of liquid crystal goggles for vision research. *Integrated Ferroelectrics: An International Journal*, 103, 1-9.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227–1268.

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, 46, 1886-1900.

Pomplun, M., Reingold, E.M. & Shen, J. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research*, 41, 2757 - 2769.

Pomplun, M., Reingold, E.M., & Shen, J. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science*, 27, 299-312.

Previc, F.H. & Blume, J.L. (1993). Visual search asymmetries in three-dimensional space. *Vision Research*, 33, 2697-2704.

Previc, F.H., & Naegele, P.D. (2001). Target-tilt and vertical-hemifield asymmetries in free-scan search for 3-D targets. *Perception & Psychophysics*, 63, 445-457.

Rayner, K. (1998). Eye Movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.

Rayner, K. & Fisher, D.L. (1987). Eye movements and the perceptual span during visual search. In J.K. O'Regan & A. Levy-Schoen (Eds.), Eye movements: From Physiology to Cognition. North Holland: Elsevier.

Reuter-Lorenz, P.A., & Fendrich, R. (1992). Oculomotor readiness and covert orienting: Differences between central and peripheral precues. *Perception and Psychophysics*, 52, 336-344.

Schor, C.M. (1979). The relationship between fusional vergence eye movements and fixation disparity. *Vision Research*, 19, 1359-1367.

Scialfa, C.T., & Joffe, K. (1998). Response times and eye movements in feature and conjunction search as a function of eccentricity. *Perception & Psychophysics*, 60, 1067-1082.

Shen, J., Reingold, E.M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception*, 29, 241-250.

Silva, M. F., Mateus, C., Reis, A., Nunes, S., Fonseca, P., & Castelo-Branco, M. (2010). Asymmetry of visual sensory mechanisms: Electrophysiological, structural, and psychophysical evidences. *Journal of Vision*, 10(6), 26, 1–11.

Snowden, R.J. (1998). Texture segregation and visual search: A comparison of the effects of random variations along irrelevant dimensions. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 1354-1367.

Sousa, R., Brenner, E., & Smeets, J.B. (2009). Slant cue are combined early in visual processing: Evidence from visual search. *Vision Research*, 49, 257–261.

Talgar, C.P. & Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. *Psychonomic Bulletin Review, 9,* 714-722.

Talgar, C.P., Pelli, D., & Carrasco, M. (2004). Covert attention enhances letter identification without affecting channel tuning. *Journal of Vision*, 4, 22-31.

Tavassoli, A., van der Linde, I., Bovik, A.C., & Cormack, L.K. (2007). An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics*, 69, 103–112.

Tavassoli, A., van der Linde, I., Bovik, A.C., & Cormack, L.K. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Research*, 49, 173–181.

Treisman, A. (1991a). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 652-676.

Treisman, A. (1991b). Representing visual objects. In D. Meyer & S. Kornblum (Eds.), Attention and Performance, XIV (pp. 163-175). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), Attention: Selection, awareness, and control: A tribute to Donald Broadbent (pp. 5-35). Oxford: Clarendon Press.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12, 97-136.*

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15-48.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), Attention and Performance III (pp. 280-292). Amsterdam: North-Holland.

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for guidance of eye movements when searching natural scenes. Journal of Vision, 10, 1–13.

Williams, D.E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and stimulus eccentricity. *Psychonomic Bulletin and Review*. 8, 476-488.

Williams, D.E., Reingold, E.M., Moscovitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology,* 51, 151-164*.*

Williams, L.G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27, 355-360.

Wolfe, J.M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review,* 1, 202-238.

Wolfe, J.M. (1996). Extending guided search: Why guided search needs a preattentive "item map." In A.F. Kramer, M.G.H. Coles, & G.D. Logan (Eds.), Converging operations in the study of visual attention (pp. 247-270). Washington, DC: American Psychological Association.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance,* 15, 419-433.

Zelinsky, G.J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36, 2177-2187.

Zelinsky, G.J. & Sheinberg, D.L. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In J. Findlay, R. Kentridge & R. Walker (Eds.), Eye movement research: Mechanisms, processes and applications (pp. 325-336). Amsterdam: Elsevier.

Zelinsky, G.J., & Sheinberg, D.L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 244-262.