Perception, 2002, volume 31, pages 1415 – 1437

DOI:10.1068/p3432

Dynamics of attention in depth: Evidence from multi-element tracking

---

Lavanya Viswanathan, Ennio Mingolla¶
Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA; e-mail: ennio@cns.bu.edu
Received 15 March 1999, in revised form 4 December 2001; published online 8 November 2002

Abstract. We examined the allocation of attention in depth using a multi-element tracking paradigm. Observers were required to track a predefined subset of from two to eight elements in displays containing up to sixteen identical moving elements. We first show that depth cues, such as binocular disparity and occlusion through T-junctions, improve performance in a multi-element tracking task in the case where element boundaries are allowed to intersect in the depiction of motion in a single frontoparallel plane. We also show that the allocation of attention across two perceptually distinguishable planar surfaces, either frontoparallel or receding at a slanting angle and defined by coplanar elements, is easier than allocation of attention within a single surface. The same result was not found when attention was required to be deployed across items of two-color populations rather than across items of a single color. Our results suggest that, when surface information does not suffice to distinguish between targets and distractors that are embedded in these surfaces, division of attention across two surfaces aids in tracking moving targets. A final experiment with populations of elements moving within distinct volumes produced similar results, suggesting that spatial separation in three dimensions, rather than confinement to surfaces as such, may explain the improved performance for the two-surface case.

1 Introduction

Our visual system is daily called upon not only to detect and recognize objects but also to track a number of them simultaneously through brief periods of occlusion. Sometimes the occluders may be similar in featural qualities to the objects being tracked. Objects are nearly always three-dimensional and are usually made up of several surfaces, each of which may lie at different distances from the observer. It would therefore seem natural that the visual system must be able to selectively deploy attention to noncontiguous moving objects that may lie at different distances from the observer. We investigate some conditions under which depth and surface cues may aid the allocation of attention in a multi-element tracking task, in which the visual system must simultaneously track a subset of identical moving objects.

1.1 *Theories of attention*

Two main types of theories of attention have been proposed to explain the allocation of attention in a scene: space-based and object-based theories. Space-based or location-based approaches suggest that attention may be allocated to specific regions or locations. These include the spotlight (Posner 1980), zoom lens (Eriksen and St James 1986), and spatial gradient (Downing and Pinker 1985; LaBerge and Brown 1989) models. The attentional focus might move like a spotlight, expand or contract like a zoom lens depending on task requirements, or be a fixed gradient of processing centered on the spatial location being attended to, falling off with distance from this location. In the strong form of these theories, everything inside the locus of attention must be attended to, while everything outside is disregarded.

However, substantial evidence suggests that selective attention can operate on perceptual objects and need not select on the basis of spatial location alone (Egeth and Yantis 1997; Scholl 2001). Such approaches have come to be known as object-based

¶ Author to whom all correspondence and requests for reprints should be addressed.

theories of attention. Data in favor of these theories have been obtained from two kinds of study: those in which two or more objects are displayed at the same spatial location and those in which the spatial location of one or more objects changes with time (multi-element tracking). The first kind of study shows that, in scenes with superimposed event sequences or overlapping shapes, human observers can selectively attend to one of the sequences or shapes and ignore the other, despite the physical overlap of spatial locations (Duncan 1984; Neisser and Becklen 1975; Rock and Gutman 1981). Kramer and Jacobson (1991) found that the extent to which a flanking form interfered with responses to a target form depended on whether or not the flanking form was perceived to be part of the same perceptual object as the target form, even when the physical positions of the two forms were the same in each case. Perceptual objects can be created and accessed with the help of object files (Kahneman and Treisman 1984; Kahneman et al 1992), attentional priority tags (Yantis and Johnson 1990; Yantis and Jones 1991), attentional sprites (Cavanagh et al 2001), object tokens (Chun and Cavanagh 1997; Kanwisher 1987), or 'fingers of instantiation' (Pylyshyn 1989, 1994). Object files can be generated 'preattentively' (Wolfe and Bennett 1997) or by perceptual grouping (Yantis 1992). Further evidence for object-based theories comes from studies that show that attention can spread across feature space instead of just visual space (Driver and Baylis 1989). Attention can thus operate on gradients such as shape, color, motion, and surfaces instead of just spatial gradients. Note also that inhibition-of-return (IOR) studies (Becker and Egeth 1998; Tipper et al 1994) suggest that space-based and object-based strategies must interact.

1.2 *Deploying attention in depth*

Although a number of studies have dealt with the allocation of attention in two-dimensional space, interest in the deployment of attention in three-dimensional space has been rather recent. Some researchers (Ghirardelli and Folk 1996; Iavecchia and Folk 1995; Theeuwes et al 1998) argue that attention cannot be preferentially allocated to specific locations in depth. Others disagree and suggest that, when deployed in depth, attention could be either viewer-centered (ie with a shallow gradient between the observer and the target and a steeper gradient beyond the target), object-centered (ie with the same slope on either side of the target), or action-centered as in selective reaching tasks (Tipper et al 1992). In two of the earliest studies in this area (Downing and Pinker 1985; Gawryszewski et al 1987) the movement of attention was investigated with a cuing paradigm in a real three-dimensional scene; this showed that the visual system can attend to specific locations in depth. These studies further suggested that attention was allocated in a viewer-centered manner. However, the effects they obtained could have been due to shifts in accommodation and eye convergence rather than attentional processing. Subsequent studies attempted to resolve this problem by using simulated three-dimensional scenes with binocular disparity information and yielded conflicting results. Ghirardelli and Folk (1996), Iavecchia and Folk (1995), and Theeuwes et al (1998) showed no effect of cuing in depth. Andersen (1990), Andersen and Kramer (1993), Atchley et al (1997), Hoffman and Mueller (1994), and Marrara and Moore (2000) found evidence for viewer-centered localization of attention in depth. Some prerequisites for the allocation of attention to a specific location in depth may be the attentional requirements of the task (Atchley et al 1997), whether or not an object (here, a placeholder) is present at that location (Hoffman and Mueller 1994), or the time of presentation of the attentional placeholders (Marrara and Moore 2000). All of these studies used spatial cuing paradigms. Two studies with visual search displays—Holliday and Braddick (1991) and Nakayama and Silverman (1986)—showed that attention can be allocated to a specific location defined by disparity and that, when this is done, there is no interference from distractors in other depth planes.

Other studies suggest that the deployment of attention across same-disparity loci is possible when the elements being attended to are part of a well-formed surface with locally coplanar elements (He and Nakayama 1995; Tyler and Kontsevich 1995). These studies show that it is difficult to attend to locations that span different surfaces. Instead of attention being allocated to spatial locations or to objects, it may be considered to be involuntarily "bound" to a surface like "a shroud that acts like a soap film in minimizing the curvature of the perceived depth surface consistent with the available disparity information" (Tyler and Kontsevich 1995, page 143).

However, in most of the studies of object-based or surface-based theories of attention, objects and surfaces are indistinguishable because objects do not span several depth planes or have more than one surface. If a virtual object whose vertices spanned different depths or surfaces could be constructed, and if the experimental task required that attention be preferentially allocated to this perceptual object, one could test whether or not attention across surfaces is harder than attention within a surface. By allowing the vertices of the virtual object to change their positions with time, thus precluding the chances of their being spatially contiguous, one could further measure object-based attention rather than surface-based attention. A paradigm that is ideally suited to the task described above is multi-element tracking.

1.3 *Multi-element tracking*

Pylyshyn and Storm (1988) first demonstrated that human observers are capable of tracking multiple randomly moving visual elements under a variety of conditions. In a display consisting of ten identical elements, observers could track a predefined subset of up to five elements with good accuracy. Since eye movements are not allowed, the elements must be tracked with attention. Pylyshyn and Storm (1988) concluded that tracking cannot be performed by a serial process since, if a single attentional spotlight jumps from element to element during tracking, the spotlight must move at impossible velocities. This suggests that the elements must be tracked in parallel, and hence an object-based representation rather than a space-based representation is essential. Pylyshyn and his colleagues suggested that the tracking is performed by a collection of 'fingers of instantiation' (FINSTs), one for each element being tracked (Pylyshyn 1989, 1994; Pylyshyn et al 1994). The FINSTs can be assigned to objects either through bottom – up factors such as featural salience or through top – down factors, such as priming.

In contrast to Pylyshyn's FINSTs, Yantis (1992) suggested that the elements being tracked were grouped into a virtual object which was then tracked as a single entity. He showed that performance in multi-element tracking was influenced by factors that controlled the formation and maintenance of a perceptual group formed by designated target items (ie the items to be tracked). The factors that influenced the formation of a perceptual group, such as the initial configuration of the target elements, the presentation mode of the target elements, and the instructions given to subjects, affected performance only early in practice. However, those that influenced the maintenance of a perceptual group during motion, such as dynamic constraints on the configuration of target elements during movement and the degree to which the velocities of the target and non-target elements were correlated within and between groups, affected performance throughout practice. Evidently, the perceptual grouping of items at disparate spatial locations into a virtual object can be governed by top – down processes. Whether the targets are perceptually grouped into a single virtual object, as suggested by Yantis, or assigned several object indexes or FINSTs, as suggested by Pylyshyn and colleagues, is not relevant to the purposes of the current experiments. Both theories are consistent with object-based deployment of attention in a scene.

The attentive tracking paradigm has also been used by Intriligator (1997) to measure the spatial resolution of visual attention, ie the minimum size to which attention can be focused at a given eccentricity. Tracking was easier in the lower visual field than in the upper visual field. Besides, attending to the targets does not also necessitate attending to the regions between them. Scholl and Pylyshyn (1999) found that successful tracking, ie the maintenance of perceptual objecthood, behind occluders, visible or invisible, requires the presence of accretion/deletion cues that are consistent with the presence of a fixed contour. More recently, Scholl et al (1999) showed that attentional allocation in multi-element tracking results in an encoding of spatiotemporal, but not featural, properties of objects. Culham et al (1998) found, using functional MRI, that, although an attentive tracking task produces almost no attentional enhancement in early visual areas and the MT – MST complex, bilateral activation is produced in parietal cortex and frontal cortex.

This paper addresses the following issues: Can the visual system attend to a group of objects that spans different depth planes or surfaces (or, alternatively, a virtual object whose vertices span different depth planes or surfaces)? More specifically, is performance in a task that necessitates the deployment of attention across different depths or surfaces always worse than in a task where attention only needs to be deployed to a single depth or surface? In experiment 1 we examine whether the addition of depth cues, such as binocular disparity and T-junctions signalling occlusion, to a tracking task has an effect on performance in the specific case where element boundaries are allowed to overlap in the two-dimensional projection plane of the monitor screen. In experiments 2, 3, 4, and 5 we look at whether attention can be simultaneously allocated across two depth planes, colors, forms, or surfaces. Preliminary reports of the present work have appeared in Viswanathan and Mingolla (1998a, 1998b).

2 Experiment 1

Like the majority of published work on visual attention, the multi-element tracking task used by Pylyshyn and Storm (1988) and Yantis (1992) required observers to deploy attention within a two-dimensional scene; that is, where elements moved only in an up-and-down or left-and-right fashion in a frontoparallel plane. In their experiments, the elements were surrounded by invisible 'cushions' that were not allowed to intersect throughout motion trajectories. This experimental construction was motivated by the supposition that, were the cushions around elements allowed to intersect, it would become very easy to confound the elements and lose the target to be tracked. The authors explicitly stated this assumption (Pylyshyn and Storm 1988, page 182):

> "The random motion of the objects was subject to the restriction that no two objects could be closer than 0.75 deg apart, so that the continuity of their identity was never ambiguous (as it would be if they were to collide)."

Note that the minimum distance between any two elements in the scene (0.75 deg) was more than one and a half times the size of any one element.

The goal of experiment 1 was to investigate whether depth cues, such as disparity and T-junctions, which signal occlusion, improve performance in a multi-element tracking task when elements are allowed to overlap one another dynamically during a trial. The results show that, although the tracking task does become more difficult when element boundaries are allowed to intersect, it does not become impossible, even in the purely two-dimensional case. More important, however, is the finding that when occlusion cues (disparity or T-junctions) are added to the display, human performance improves appreciably and, in fact, returns to the baseline performance levels found by Pylyshyn and Storm (1988) and Yantis (1992).

Since in the current study we used displays that mimicked those used by Pylyshyn and Storm (1988) and Yantis (1992), their paradigm is described in detail. Pylyshyn and Storm (1988) constructed two-dimensional displays that contained ten moving white crosses (ie + signs) (figure 1). Randomly chosen subsets of from one to five elements were designated as targets by flashing on and off in a static display before the movement phase. The movement phase lasted from 7 s to 15 s. At the end of the movement phase, a solid white square was flashed over (ie replaced) one of the moving elements for a small period of time. This was the probe for that trial and could be flashed either on a target or on a non-target element. The task of the observer was to specify whether the flash occurred on a target or a non-target element. A fixation square appeared at the center of the screen at all times, but eye movements were not monitored.

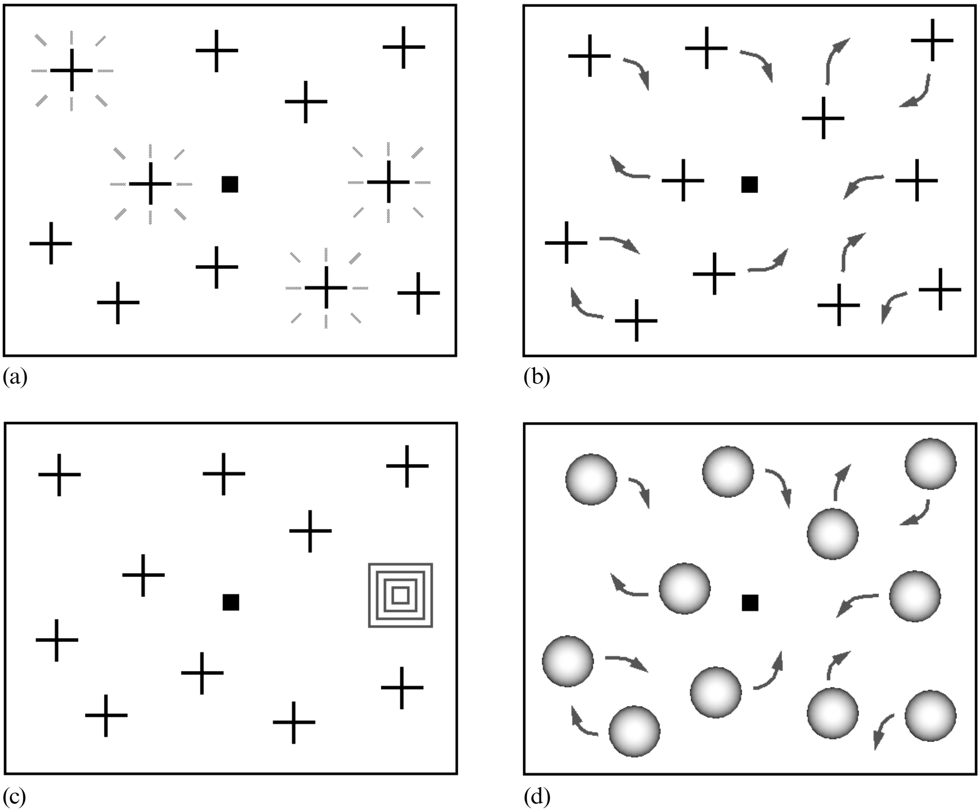

Figure 1. Schematic diagram of the types of displays used by Pylyshyn and Storm (1988) and Yantis (1992). These displays contain ten identical plus signs that move randomly. (a) Target designation phase: the targets are flashed on and off; this is diagrammed by lines radiating from the targets. (b) Movement phase: all the elements move in randomly selected directions. (c) Response phase, in which subjects are asked to move a mouse-driven cursor and click on all objects that were targets, as designated by earlier flashing. (d) The same task but for the kinds of elements used in this experiment. The curvilinear trajectories shown here were used for experiments 2, 3, 4, and 5. Trajectories in experiment 1 were linear.

2.1 *Method*

2.1.1 *Observers.* Eight naïve observers participated in six 1 h long sessions and were compensated at a rate of $8 an hour. All subjects had normal or corrected-to-normal vision and had no previous experience in visual tracking experiments, though some of them had participated in psychophysical experiments before. All observers could see depth in displays containing disparity information.

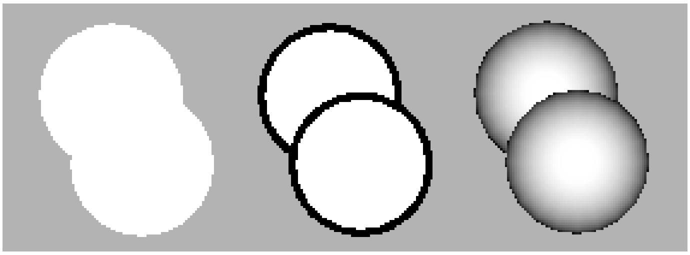
(a) (b) (c)

Figure 2. Formation of T-junctions during element intersections. (a) The elements are not shaded, so no T-junctions are formed and the display does not appear to be in depth. (b) Adding thick black outlines to flat white disks creates T-junctions during element intersections without the appearance of three-dimensionality. (c) Shading of the elements achieves two purposes: it makes the elements look three-dimensional and it creates strong contours that lead to T-junctions when the elements intersect, giving a percept of depth due to occlusion.

2.1.2 *Design.* Two depth cues were considered: binocular disparity and T-junctions signalling occlusion. T-junctions could be obtained either by adding a black outline to flat white elements, as in figure 2b, or by shading the elements, as in figure 2c. The different possible combinations of these depth cues in any given trial lead to six experimental conditions based on depth-cue manipulation. In addition, two sets of trajectories were generated for the elements. This resulted in a 6×2 design generating the following 12 experimental conditions.

Set 1. Trajectories in this set were linear. Elements change their direction of motion only when they collide either with the boundaries of the display screen or with a 'cushion' placed around the fixation square, as described in section 2.1.3. The depth conditions for this set are listed below:

Condition A: no depth cue is present;
Condition B: only outlines are present;
Condition C: only binocular disparity is present;
Condition D: only shading is present;
Condition E: two depth cues—outlines and binocular disparity—are present;
Condition F: two depth cues—shading and binocular disparity—are present.

Set 2. Trajectories in this set were also linear but with a random component. As in set 1, elements change their direction of motion when they collide either with the display screen boundaries or with the fixation square 'cushion'. They can also change their direction of motion after a random number of motion frames (see section 2.1.3 for a more detailed description of how this was implemented). The depth conditions are listed below:

Condition G: no depth cue is present;
Condition H: only outlines are present;
Condition I: only binocular disparity is present;
Condition J: only shading is present;
Condition K: two depth cues—outlines and binocular disparity—are present;
Condition L: two depth cues—shading and binocular disparity—are present.

2.1.3 *Stimuli.* Simulations were performed on a Silicon Graphics RE2 machine running an Irix 6.2 operating system. The displays were viewed through Crystal Eyes Stereographics liquid-crystal stereo glasses. The program displayed alternate images on the screen corresponding to the left-eye and right-eye images. This ensured that displays containing differing disparity information could be presented to each eye separately. The screen resolution was 1025×768 and the frame rate was 96 Hz for both eyes (ie 48 Hz per eye).

The displays used in this experiment were based on those of Pylyshyn and Storm (1988) and Yantis (1992) (figure 1). The display contained ten identical elements (disks or spheres). Two types of depth cues were used: binocular disparity and T-junctions. T-junctions can be obtained by shading the elements (figure 2). All the elements were shaded identically, with a uniform gradation between white at the center and black at the boundary. Shading the white disks makes them look more spherical and gives the impression of three-dimensional structure. Further, when two shaded spherical elements overlap, one sees a T-junction. This is a strong depth cue that tells us which element is in front of the other (Nakayama et al 1989). Non-shaded white disks do not form T-junctions when they overlap; instead, they form figure-eight regions, with no depth ordering. Another way of obtaining T-junctions, but without the appearance of three-dimensionality, is by adding thick black outlines to flat white disks as in figure 2b.

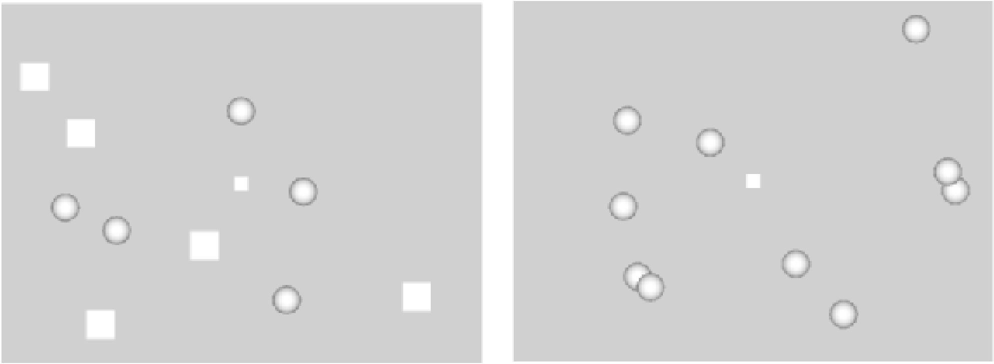
(a) (b)

Figure 3. Two frames of the multi-element tracking display. (a) The target designation phase, with the target set being 'flashed'. (b) One frame of the movement phase in which two overlaps may be seen.

Each trial consisted of three phases (figures 1 and 3):

Target designation phase: Before the onset of movement, a randomly selected subset of five out of ten elements was flashed, ie replaced by a red square of similar dimensions, five times. This defined the target set for that trial. The initial positioning of elements on the screen was done in such a way that elements did not come too close to one another, so that their identities were clearly defined at the onset of the trial. The minimum separation between the boundaries of any two elements in this frame was 0.3 deg. The target designation phase lasted approximately 4.5 s.

Movement phase: After the flashing, all the elements started moving in different randomly chosen directions. Their trajectories were restricted so that they always lay in a three-dimensional depth volume (figure 4). There were 45 possible directions of movement. The angular separation between any two adjacent directions was 8 deg. Element boundaries were allowed to intersect in the plane of the monitor screen but not in the three-dimensional depth volume. In displays without depth, this meant that two disks could intersect to form a filled figure-eight (figure 2). In displays with either disparity or T-junctions, or both, elements appeared to move in front of or behind one another. The minimum disparity difference between any two elements during an overlap was 0.05 deg. In displays with disparity, elements changed their disparities throughout the trial in a smooth fashion, so that they appeared to be moving away from or toward the observer, while simultaneously moving vertically and horizontally on the screen. The movement phase comprised 300 static frames and lasted approximately 11.5 s.

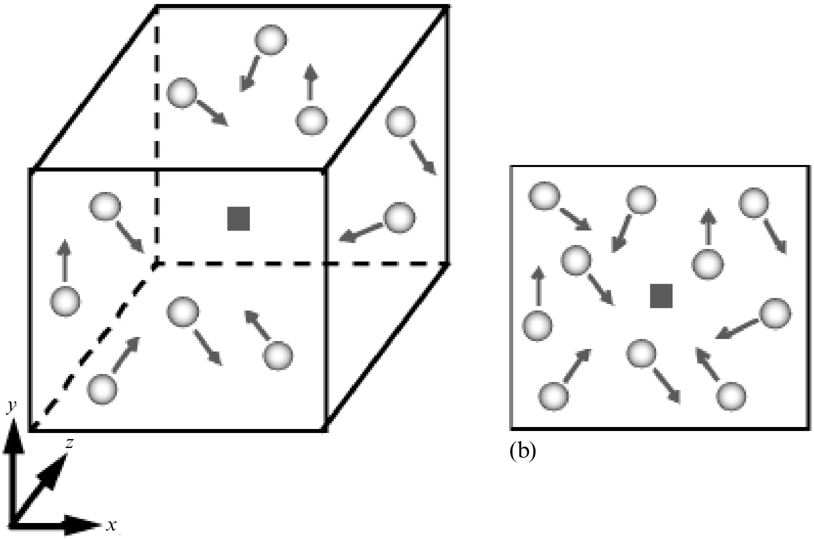

Figure 4. Schematic diagram of element trajectories in experiment 1. (a) The elements move linearly in a three-dimensional depth volume. In both set 1 and set 2 trajectories, elements bounce off the edges of the volume as well as an invisible cushion around the fixation square. Depth (z) is defined by binocular disparity. (b) View as seen on the two-dimensional projection plane of the monitor screen.

Trajectories were precomputed and chosen so that not more than three disks overlapped significantly at a single spatial location. In both set 1 and set 2 trajectories, elements bounced off the edges of the depth volume, ie their trajectories were reflected off these edges. An invisible square cushion was placed around the fixation square and elements were not allowed to intersect this cushion. If they touched the edges of this cushion, they were bounced off it. The minimum distance between any element and the fixation square was 0.5 deg. Otherwise, elements maintained their trajectories and always moved in a straight line. In addition to the changes in the direction of motion of an element due to collisions with the boundaries of the display screen and the cushion around the fixation square, set 2 trajectories also included a random component. This additional control was added to emulate the earlier work in multi-element tracking by Pylyshyn and Storm (1988) and Yantis (1992). The reason for this control is that the totally linear trajectories used in set 1 may make the motion of elements predictable, thereby reducing the attentional load of the tracking task. Following Yantis (1992), the random nature of an element's trajectory was implemented by assigning a trajectory duration to the element. This represents the number of frames for which the element retains its current direction of motion. The trajectory duration of an element was an integer chosen randomly from the set [15, 35]. At the end of this duration, a new direction of motion was randomly chosen for the element.

For set 1 trajectories, the minimum and maximum number of overlaps in a given trial were 10 and 32, respectively. For set 2 trajectories, these numbers were 12 and 36, respectively. The average number of overlaps per trial was 21 for both sets. On average, each overlap took 10 frames to complete. In set 2 trials, the minimum, maximum, and average number of times per trial that a change in the direction of motion of an element occurred during an overlap were 8, 29, and 18, respectively.

Query phase: Motion was stopped at the end of the movement phase. No overlaps were allowed in this phase so that there was no doubt about the identity of any of the elements. Observers were instructed to mark all the elements they had been tracking by clicking on them with the left mouse button (Intriligator 1997). When the observer clicked on an element, it was replaced by a red square of the same size as the element. When the observer had finished clicking on a number of elements equal to the number of targets in that trial, he/she pressed the space bar to initiate the next trial. The query phase was not timed. Observers were allowed to take as much time as they needed to respond.

The experiment was conducted under free-viewing conditions, so all reported dimensions in this paragraph are approximate. Though observers were instructed to fixate on the central square in the displays, eye movements were not monitored. The depth volume subtended 13.7 deg in width and 10.2 deg in height. The disparity of the closest surface of the volume was -0.13 deg and that of the far surface was 0.13 deg. Each element (disk or sphere) subtended a visual angle of 0.8 deg vertically and horizontally. In outlined flat disks, the thickness of the black outline was 0.05 deg. The red flashing square that replaced an element during flashes was 0.8 deg wide. The fixation square subtended a visual angle of 0.4 deg. The speed of movement of each element was 2.4 deg s^{-1}. The background was colored cyan. The spheres and the fixation square were white.

2.1.4 *Procedure.* The experiment consisted of six sessions. Each observer was required to do only one session per day. Each session contained three blocks: one practice block and two experimental blocks. The practice block lasted around 10 min and each experimental block lasted approximately 25 min. The practice block contained 12 trials and each experimental block contained 30 trials. The data from the practice block were discarded. On a single day, an observer saw 5 trials in each experimental condition and a total of 60 trials. Over the course of the whole experiment, an observer saw 30 trials in each experimental condition, ie a total of 360 trials. The starting parameters (element positions, directions of movement, target set) for 72 trials (12 practice trials and 30 trials each for set 1 and set 2) were precomputed. Other than for the practice block, each of the trajectories was then presented in every one of the experimental conditions. Thus, no experimental condition was given the unfair advantage of fewer collisions or other distinguishing factors that would make the tracking task much easier or much more difficult. The order of presentation of the trials was randomized over the whole experiment. Therefore, the probability of an observer seeing the same trajectory in two consecutive trials was very low.

2.1.5 *Instructions to observers.* The task was explained to the observers. They were told to fixate on the central square in the display and to track the target elements mentally rather than with their eyes. Observers were instructed to pick out all the elements they had been tracking by clicking on them with the left mouse button. Observers received feedback during the practice block telling them how many targets they had successfully tracked. No feedback was given during the experimental blocks. Observers were encouraged, through an on-screen display, to rest their eyes and take breaks of around 5 min between blocks, as it was essential for the experiment that they concentrate completely on the tracking task.

2.2 *Results*

The performance levels of observers are summarized in figure 5. The percentage of correctly tracked targets was computed as the ratio of the number of correctly tracked targets in a trial to the total number of targets in that trial. Chance performance was accordingly 12.5%, 25%, 37.5%, and 50% for two-, four-, six-, and eight-target trials, respectively. Five two-way repeated-measures ANOVAs were performed on the data.

For set 1, conditions A, B, C, and E together represent a 2×2 design on the presence or absence of two depth cues—binocular disparity, and element outlines. A two-way ANOVA, with binocular disparity and outlines as the two factors, reveals that

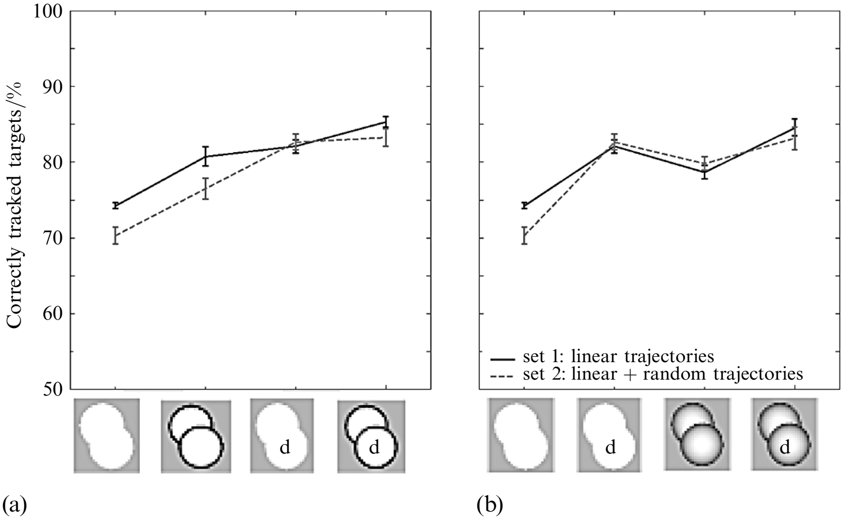

Figure 5. Experiment 1: Mean results across subjects. Solid lines represent curves for set 1 trajectories while dashed lines depict plots for set 2 trajectories. Iconic diagrams below the bottom axis show which conditions the corresponding data points belong to. A 'd' inside an element means that the condition contained binocular disparity. Bars represent standard errors. (a) Plots for the 2 × 2 binocular disparity versus outlines design. (b) Plots for the 2 × 2 shading versus binocular disparity design.

the effect of the presence of either factor on performance levels is significant (disparity: $F_{1,7} = 34.21$, $p = 0.0006$; outlines: $F_{1,7} = 21.7$, $p = 0.0023$). The effect of the interaction of both cues was also significant ($F_{1,7} = 9.72$, $p = 0.0169$). Similarly, conditions A, C, D, and F cumulatively represent a 2 × 2 design on the presence or absence of two depth cues—binocular disparity and shading. A two-way ANOVA with these two cues as the factors shows that each factor, when present, significantly betters performance (disparity: $F_{1,7} = 35.78$, $p = 0.0006$; shading: $F_{1,7} = 14.49$, $p = 0.0069$). However, in this case, the effect of the interaction of both cues was not significant ($F_{1,7} = 4.17$, $p = 0.0803$).

A similar analysis was performed for set 2. Here, conditions G, H, I, and K together represent a 2 × 2 design on the presence or absence of two depth cues—binocular disparity and element outlines. A two-way ANOVA, with binocular disparity and outlines as the two factors, reveals that the effect of the presence of either factor on performance levels is again significant (disparity: $F_{1,7} = 27.64$, $p = 0.0012$; outlines: $F_{1,7} = 15.65$, $p = 0.0055$). The effect of the interaction of both cues was also significant ($F_{1,7} = 11.1$, $p = 0.0126$). Further, conditions G, I, J, and L cumulatively represent a 2 × 2 design on the presence or absence of two depth cues—binocular disparity and shading. A two-way ANOVA with these two cues as the factors shows that each factor, when present, significantly betters performance (disparity: $F_{1,7} = 25.14$, $p = 0.0015$; shading: $F_{1,7} = 18.98$, $p = 0.0033$). The effect of the interaction of both cues was again significant ($F_{1,7} = 43.61$, $p = 0.0003$).

Finally, a two-way ANOVA with set type and condition type as the two factors was computed. No significant effect of set type on performance was observed ($F_{1,7} = 4.42$, $p = 0.0737$). However, as expected from the earlier analyses, the effects of condition type as well as the interaction were significant (condition: $F_{5,35} = 26.69$, $p < 0.0001$; interaction: $F_{5,35} = 2.89$, $p < 0.05$).

The main result to be noted from these data is that the addition of any depth cue led to a significant improvement in tracking performance. In all cases, the improved performance levels were comparable to those found by Pylyshyn and Storm (1988).

Moreover, there was no significant difference in performance between set 1 and set 2 trials. This suggests that even if the attentional load in set 1 trials was less than that in set 2, the effect is negligible. Subsequent experiments in this article, therefore, do not include a random component while computing element trajectories.

2.3 *Discussion*

The results corroborate earlier results that human observers can successfully track up to 5 targets in a display consisting of ten identical elements moving randomly (Pylyshyn and Storm 1988; Yantis 1992). During the brief durations when an element in the current displays overlaps another element, the only clues to the element's identity were occlusion cues (either disparity or T-junctions or both) that specified which element was in front of, and which behind, the other. Worse performance in the case where no depth cues are present shows that without these cues, tracking multiple moving targets in the presence of identical distractors becomes very difficult. An important conclusion that can be drawn from this study is that the addition of depth cues, or, more specifically, disparity cues or T-junctions, to a multi-element tracking paradigm makes the tracking task much easier than otherwise when element boundaries, in the two-dimensional projection plane of the monitor screen, are allowed to intersect. In fact, not only does the addition of depth cues make the tracking task easier, but performance levels improve to match the baseline performance levels found by Pylyshyn and Storm (1988) and Yantis (1992) for displays with no depth information. A recent study by Blaser et al (1999) yielded similar results. Also, Scholl and Pylyshyn (1999) found that multi-element tracking can occur in the presence of static occluders, further supporting our contention.

3 Experiment 2

The prediction implied by studies on the allocation of attention to surfaces (He and Nakayama 1995; Tyler and Kontsevich 1995) is that the deployment of attention to a virtual object whose vertices lie in different depth planes defined by disparity and that does not naturally form a surface would be more difficult than when the object is part of a surface. The current experiment was designed to test this prediction. In the one-depth case, all the elements (targets and distractors) were restricted to lie on the same depth plane. In the two-depths case, targets and distractors were distributed equally between two depth planes, so depth was not predictive of the target versus distractor distinction. There were at least as many distractors as targets on a given depth plane. An involuntary attentional mechanism that would be triggered by the presence of same-depth surfaces should interfere with the goal-driven attentional task of tracking targets across depth planes and ignoring distractors from the same depth plane. If attention to a particular location on a surface always results in the spread of attention across the entire surface, then it would be harder to track targets and ignore distractors embedded on that surface. The task would be doubly hard if one had to track targets across two surfaces while ignoring distractors at both surfaces. Thus, if it is harder to divide attention across two depth planes than within the same depth plane, performance should be worse for the two-depths case.

3.1 *Method*

3.1.1 *Observers.* Five observers participated in four sessions of approximately 1 h and were compensated at a rate of $8 an hour. Each session was conducted on a different day. All subjects had normal or corrected-to-normal vision. All observers could see depth in displays containing disparity information. Two of the subjects had participated in experiment 1.

3.1.2 *Design.* Two independent variables were examined: number of targets (2, 4, 6, or 8) and number of depth planes (one or two). A 4×2 experimental construction was used.

3.1.3 *Materials.* The experimental apparatus was the same as that used in experiment 1 with the following exceptions. Elements were always two-dimensional white disks. Each element subtended a visual angle of 0.4 deg vertically and horizontally. The speed of movement of each element was 2.8 deg s^{-1}.

The total number of elements was always sixteen. The number of targets in a trial was variable: 2, 4, 6, or 8. Figure 6 shows a schematic diagram of the depth planes used. Planes B and C were presented with disparities of −0.17 deg and 0.17 deg, respectively, one in front of and one behind the plane of the monitor screen. In the one-depth case, ie in experimental conditions A through D, all the elements were presented on depth plane A at zero disparity, so the left-eye image was identical to the right-eye image. In the two-depths case (conditions E through H), targets as well as distractors were divided equally between the front depth plane B and the back depth plane C. The fixation square was always on plane A. Disparity and depth remained constant throughout the trial. Initial element positions were generated randomly. Trials were precomputed before data collection. All observers were presented the same trials.

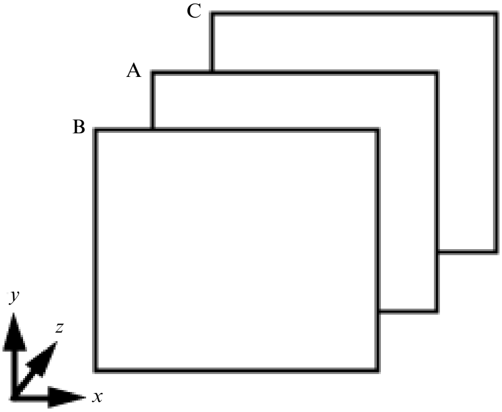

Figure 6. Schematic diagram of frontoparallel depth planes used in experiment 2. See text for details.

Elements were not allowed to occlude one another at any point during the trial. This was necessary because elements were confined to a particular depth plane throughout the trial and, for elements within a single depth plane, no depth information, in the form of either disparity or T-junctions, could be provided at intersections. Besides, it has already been shown in experiment 1 that the task becomes too hard in the absence of depth information when intersections are permitted. Occlusions were prevented by using a force-field method (Scholl and Pylyshyn 1999). Each element was repulsed by the screen boundaries, by the fixation square, and by other elements (whether on the same depth plane or on a different depth plane). The force field generated in each of these terms was inversely proportional to the square of the distance between the element and the corresponding feature in the display. The total force field acting on an element was the sum of these three constituent terms. The repulsion method caused the directions and speeds of the elements to change smoothly from frame to frame. In addition to using the repulsion method, controls were added to ensure that trajectories satisfied the following constraints: (a) no element came closer than 0.07 deg to the fixation square and (b) no two elements came closer than 0.14 deg to each other. Stimulus sequences in which these constraints were not satisfied were discarded and not shown to observers. In general, the repulsion method ensured that the distance between two elements and the distance between an element and the fixation square stayed well above these limits.

3.1.4 *Procedure.* The experiment consisted of one practice block and eight experimental blocks. The practice block contained 16 trials: 2 for each of the eight experimental conditions. The experimental blocks contained a total of 256 trials: 32 for each condition. The order of presentation of the trials was randomized. Each observer was required to do four sessions on four different days. Each session consisted of the practice block and two experimental blocks and lasted approximately 1 h.

3.1.5 *Instructions to observers.* The instructions to observers were the same as before. Subjects were not told that targets would be distributed equally between planes.

3.2 *Results*

The performance levels of observers are summarized in figure 7. A two-way repeated-measures ANOVA, with number of targets and the depth criterion, ie one depth plane versus two depth planes, as the two factors, was performed on the data. The effect of number of targets was significant ($F_{3,12} = 53.93$, $p = 0.0001$). Performance in the two-depths case was significantly better than in the one-depth case ($F_{1,4} = 15.17$, $p = 0.0176$). The effect of the interaction was not significant ($F_{3,12} = 2.19$, $p = 0.1419$).

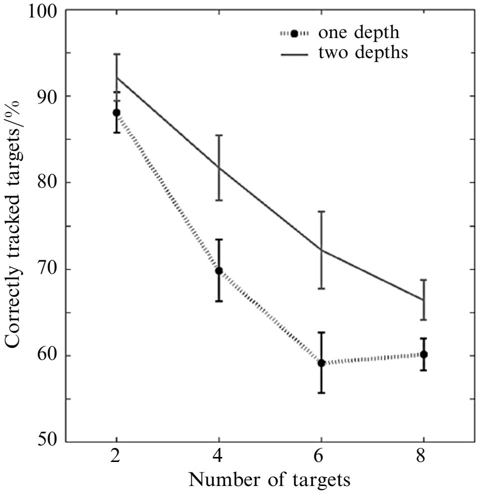

Figure 7. Experiment 2. Mean results for one depth versus two depths across subjects. Bars represent standard errors.

The proportion of variance due to the two experimental factors can be found by computing partial omega squared, $\hat{\omega}^2$ (see Keppel 1991 for a detailed discussion). This reveals that the variance due to the number of targets constitutes 64.27% of the total variance while the variance due to the depth factor (one-depth versus two-depths) constitutes 10.16% of the total variance.

3.3 *Discussion*

The results show that not only does performance in a multi-element tracking task not deteriorate when attention must be allocated across two depth planes instead of within a single depth plane, it actually improves. This suggests that the addition of depth cues can make the multi-element tracking task easier.

4 Experiment 3

An alternative explanation exists for the data obtained in experiment 2. The two-depths case may not have necessitated the deployment of attention across two depth planes. Instead, this case may be equivalent to breaking down the original tracking task into two smaller tasks, each one of which by itself would be easier to perform. If the visual system is able to track multiple elements in parallel in each one of the sub-tasks with little interference from the other, then an improvement in performance may be expected. The current experiment tests this hypothesis by using two color factors instead of two depth factors. In the one-color case, all the elements (targets and distractors) are the same color. In the two-colors case, half the targets and half the distractors are one color while the other half are a different color. This would again be equivalent to breaking down the original tracking task into two sub-tasks, each defined by a different color. If performing two smaller tracking tasks in parallel is easier than performing a single big task, then an improvement in performance should be expected for the two-colors case.

4.1 *Method*

4.1.1 *Observers.* The same five observers who participated in experiment 2 were used for this experiment.

4.1.2 *Design.* Two independent variables were examined: number of targets (2, 4, 6, or 8) and number of colors (one or two). A 4 × 2 experimental design was used.

4.1.3 *Materials.* The experimental apparatus and methods were the same as those for experiment 2 except that the elements were now presented in a completely two-dimensional display with no disparity information. In the one-color case, half the time all the elements were colored white while the rest of the time they were all colored yellow. In the two-colors case, the targets and distractors were divided evenly between the two colors. The luminances of the two colors were chosen so that neither one was more salient than the other.

4.1.4 *Procedure.* The procedure of this experiment was the same as for experiment 2. The same trajectories were reused here, with color replacing disparity.

4.1.5 *Instructions to observers.* The instructions to observers were the same as before. Subjects were not told that targets would be distributed equally between colors.

4.2 *Results*

The performance levels of observers are summarized in figure 8. A two-way repeated-measures ANOVA, with number of targets and color criterion, ie one color versus two colors, as the two factors, was performed on the data. The effect of number of targets was significant ($F_{3,12} = 59.35$, $p = 0.0001$). No significant difference between the one-color

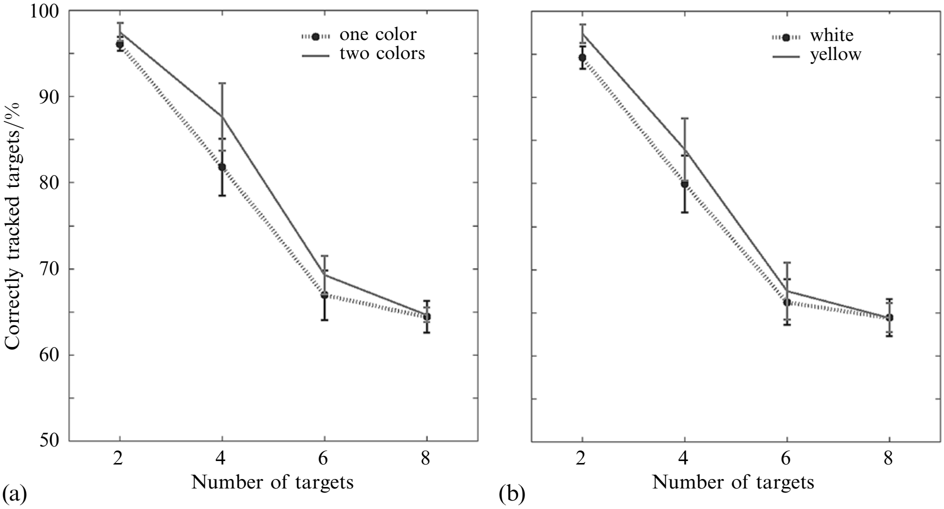

Figure 8. Experiment 3. Mean results across subjects. (a) Mean results for the one-color versus two-color factors. (b) Mean results for white versus yellow in the one-color case. Bars represent standard errors.

and two-color cases was found ($F_{1,4} = 4.53$, $p = 0.1003$). The effect of the interaction was not significant ($F_{3,12} = 2.55$, $p = 0.1043$). No significant difference was found between white and yellow in the one-color case ($F_{1,4} = 1.49$, $p = 0.2899$).

Partial $\hat{\omega}^2$ computations revealed that the variance due to number of targets and the color factor (one-color versus two-colors) constituted 72.16% and 0.81% of the total variance, respectively. Clearly, the color factor has very little effect on the total variance observed in the experiment.

4.3 *Discussion*

The results show no difference in performance between the one-color case and the two-colors case even though this experiment had the same statistical power as experiment 2. Whereas the depth factor was responsible for 10.16% of the total variance in experiment 2, the color factor was responsible for only 0.81% of the total variance in experiment 3. Based on this, it is possible to conclude that if the visual system decomposes the original tracking task into two smaller ones that may be performed in parallel independently of each other, it achieves no gain by doing this when the distinguishing element between the two sub-tasks is color rather than disparity. The results also show that tracking across two colors is not more difficult than tracking within a single color.

5 Experiment 4

As an additional control, a tracking experiment with two forms instead of two depths was performed. This experiment addresses the same issues as experiment 3. In the one-form case, all the elements (targets and distractors) have the same shape. In the two-forms case, half the targets and half the distractors are hollow squares (□) while the other half are hash signs (#). This would again be equivalent to breaking down the original tracking task into two sub-tasks, each defined by a different form. As in experiment 3, if performing two smaller tracking tasks in parallel is easier than performing a single big task, then an improvement in performance should be expected for the two-forms case.

5.1 *Method*

5.1.1 *Observers.* Five observers participated in this experiment, two of whom had also participated in experiments 1, 2, and 3.

5.1.2 *Design.* Two independent variables were examined: number of targets (2, 4, 6, or 8) and number of form (one or two). A 4 × 2 experimental design was used.

5.1.3 *Materials.* The experimental apparatus and methods were the same as those for experiment 3 except that the elements were now presented in two different forms instead of two different colors. The two forms used were hollow squares (□) and hash signs (#). Both shapes consisted of two vertical and two horizontal lines and differed only in the configuration in which they were placed. In the one-form case, half the time all the elements were squares while the rest of the time they were all hashes. In the two-forms case, the targets and distractors were divided evenly between the two shapes. The luminances of the two forms were identical.

The size of the elements, the speed of their motion, and the length of the motion sequence were changed to keep the overall difficulty of this task comparable to those in previous experiments. Each element now subtended a visual angle of 0.3 deg vertically and horizontally and moved at a speed of 3.6 deg s^{-1}. The target designation and the movement phases lasted approximately 5.2 s and 8.8 s, respectively. The movement phase comprised 200 static frames.

5.1.4 *Procedure.* The procedure of this experiment was the same as for experiment 3.

5.1.5 *Instructions to observers.* The instructions to observers were the same as before. Subjects were not told that targets would be distributed equally between forms.

5.2 *Results*

The performance levels of observers are summarized in figure 9. A two-way repeated-measures ANOVA, with number of targets and form criterion, ie one form versus two forms, as the two factors, revealed that the effect of number of targets was significant ($F_{3,12} = 90.89$, $p < 0.0001$). No significant difference between the one-form and two-form cases was found ($F_{1,4} = 4.57$, $p = 0.0993$). The effect of the interaction was not significant ($F_{3,12} = 1.15$, $p = 0.3676$). No significant difference was found between squares and hashes in the one-form case ($F_{1,4} = 0.6$, $p = 0.4815$).

Partial $\hat{\omega}^2$ computations revealed that the variance due to number of targets and the form factor (one-form versus two-forms) constituted 79.43% and 0.13% of the total variance, respectively. Clearly, the form factor has even less effect on the total variance observed in this experiment than the color factor had on the variance in experiment 3.

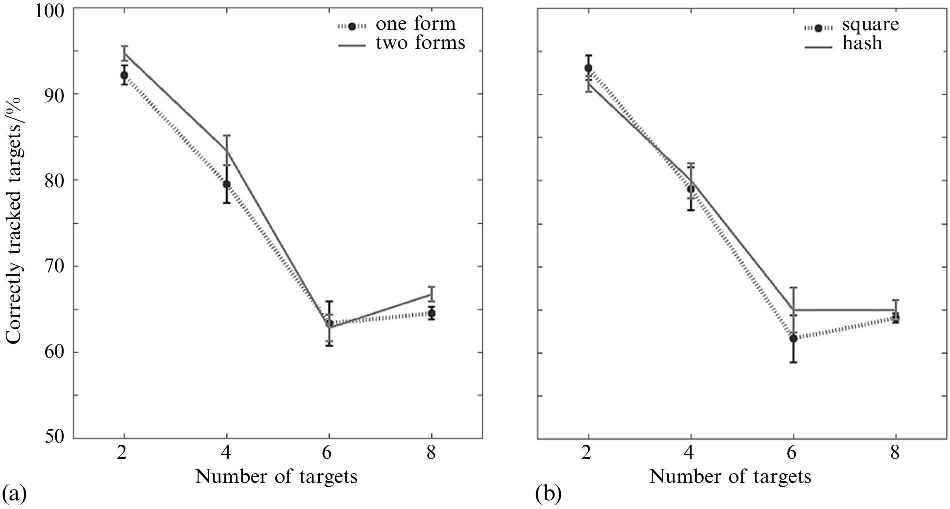

Figure 9. Experiment 4. Mean results across subjects. (a) Mean results for the one-form versus two-form factors. (b) Mean results for squares versus hashes in the one-form case. Bars represent standard errors.

5.3 *Discussion*

The results of this experiment corroborate those of experiment 3 and confirm that there is no difference in performance between the one-form case and the two-forms case just as no difference between the one-color and two-colors cases was observed in the previous experiment. To summarize, then, the depth factor was responsible for 10.16% of the total variance in experiment 2, whereas the color factor was responsible for only 0.81% of the total variance in experiment 3 and the form factor was responsible for 0.13% of the total variance in experiment 4. It can, therefore, be concluded that the trend observed in experiment 2 was a result of the depth conditions in that experiment and cannot be reproduced by substituting a color or form manipulation for binocular disparity manipulation.

6 Experiment 5

Experiment 2 shows that the deployment of attention across two depth planes is easier than the allocation of attention within a single depth plane. Does this result also extend to surfaces that are not frontoparallel? In this experiment we test whether performance in a multi-element tracking task differs when all elements are restricted to lie on the same implicit receding planar surface or divided equally between two parallel surfaces. The same experimental construction as in experiment 2 was used except with receding planar surfaces instead of frontoparallel depth planes.

6.1 *Method*

6.1.1 *Observers.* Five observers, including one of the authors, participated in four sessions of approximately 75 min and were compensated at a rate of $8 an hour. Each session was conducted on a different day. All subjects had normal or corrected-to-normal vision. All observers could see depth in displays containing disparity information.

6.1.2 *Design.* Two independent variables were examined: number of targets (2, 4, 6, or 8) and number of receding planar surfaces (one or two). A 4×2 experimental construction was used.

6.1.3 *Materials.* The experimental apparatus and methods were the same as those for experiment 2 except for the following changes. Instead of frontoparallel depth planes, planar surfaces that receded in depth were depicted (figure 10). Three parallel planar surfaces, which made an angle of 60° with the vertical, were used. Surfaces B and C were equidistant from surface A. The disparity difference between A and B was −0.05 deg while that between A and C was 0.05 deg. The disparity in plane A ranged from −8.8 deg to 8.8 deg from front (bottom) to back (top) for a viewing window of height 10.2 deg. The disparity ranges for planes B and C were −8.85 deg to 8.75 deg and −8.75 deg to 8.85 deg, respectively. Rectangular elements were used instead of disks, as these were more readily seen to be oriented along a receding planar surface. Elements were white with black outlines. The rectangular shapes of the elements were skewed to depict slant. Disparity was asymmetrical in the sense that rectangular elements were drawn in the left-eye image while horizontally shifted parallelograms were drawn in the right-eye image (see He and Nakayama 1995). Slant, as depicted by skew, and asymmetrical binocular disparity were both used to convey the impression of transparent and perceptually distinguishable receding surfaces. In addition, each surface was presented with a transparent bounding frame with texture information and grid lines to heighten the impression of a planar surface receding at a slanting angle. The surfaces contained no texture apart from that on their respective bounding frames. Elements were embedded on these transparent surfaces. During the target designation phase, the elements were overlaid on grids attached to the surfaces they lay on. The fixation square, which was also displayed with slant and disparity information, was always on surface A.

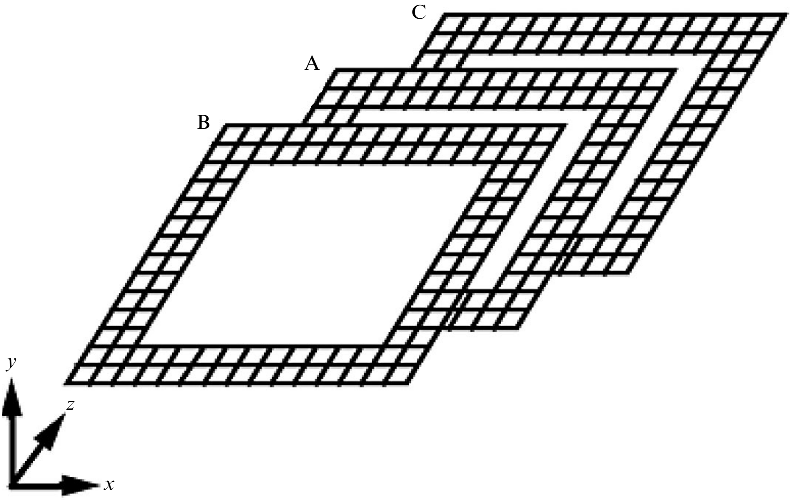

Figure 10. Schematic diagram of surfaces used in experiment 4. See text for details.

In the one-surface case, only surface A was shown. All elements were placed on this surface. In the two-surfaces case, surfaces B and C were shown and targets and distractors were divided equally between these surfaces. Element trajectories were restricted so that each element always remained on the surface that it was assigned to. The slant of all elements remained constant throughout the trial. Disparity was defined by the position of the element on its surface.

6.1.4 *Procedure.* The procedure of this experiment was the same as for experiment 2.

6.1.5 *Instructions to observers.* The instructions to observers were the same as before.

6.2 *Results*

The performance levels of observers are summarized in figure 11. A two-way repeated-measures ANOVA, with number of targets and surface criterion (one receding planar surface versus two receding planar surfaces) as the two factors, was performed on the data. The effect of number of targets was significant ($F_{3,12} = 66.55$, $p = 0.0001$). Performance in the two-surfaces case was significantly better than in the one-surface case ($F_{1,4} = 18.40$, $p = 0.0128$). The effect of the interaction was also significant ($F_{3,12} = 6.14$, $p = 0.009$).

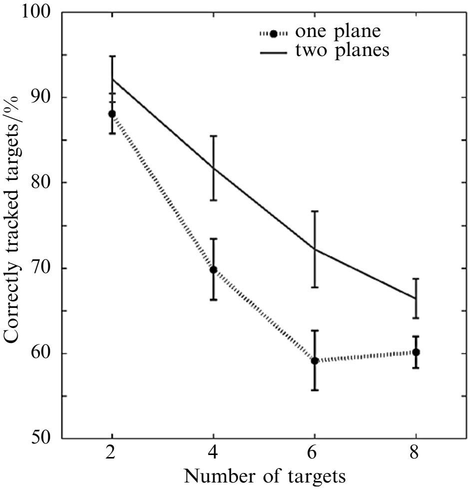

Figure 11. Experiment 5. Mean results for one plane versus two planes across subjects. Bars represent standard errors.

Partial $\hat{\omega}^2$ computations revealed that the variance due to number of targets and the surface factor (one-surface versus two-surfaces) constituted 59.62% and 19.57% of the total variance, respectively. Clearly, the presence of two surfaces instead of one had a big impact on the total variance observed in the experiment.

6.3 *Discussion*

The results show that the deployment of attention in a multi-element tracking task across two surfaces can be easier than the allocation of attention within a single surface. In contrast to the color factor in experiment 3, which was responsible for only 0.81% of total variance, and the form factor in experiment 4, which was responsible for only 0.13% of total variance, both the depth factor in experiment 2, responsible for 10.16% of total variance, and the surface factor in experiment 5, responsible for 19.57% of total variance, proved to be strong influencing factors on performance in a multi-element tracking task.

7 Experiment 6

The improvement in performance observed in the two-depths case in experiment 2 and the two-surfaces case in experiment 5 could be attributed to the additional three-dimensional spatial separation between objects in these cases as compared to the one-depth or the one-surface cases. It is conceivable that observers would be less likely to confuse targets with distractors when they are further apart in three dimensions than when they are close together.

Experiment 6 was designed to test this hypothesis. In this experiment, we compared a one-volume case, in which targets and distractors are confined to lie within a three-dimensional depth volume, with a two-volumes case, in which targets and distractors are distributed evenly between two separate depth volumes that are spatially separated in three dimensions. If spatial separation of targets and distractors is causing the observed performance in experiments 2 and 5, then we should see the same effect in experiment 6 as well.

7.1 *Method*

7.1.1 *Observers.* Five observers participated in four sessions of approximately 1 h and were compensated at a rate of $8 an hour. Each session was conducted on a different day. All subjects had normal or corrected-to-normal vision. All observers could see depth in displays containing disparity information.

7.1.2 *Design.* Two independent variables were examined: number of targets (2, 4, 6, or 8) and number of depth volumes (one or two). A 4×2 experimental construction was used.

7.1.3 *Materials.* The experimental apparatus was the same as that used in experiment 2 with a few changes. Instead of frontoparallel depth planes, depth volumes were depicted (figure 12). The near and far surfaces of volume A were presented with disparities of -0.24 deg and -0.1 deg, respectively. The disparities for the near and far surfaces of volume B were -0.07 deg and 0.07 deg. The corresponding numbers for volume C were 0.1 deg and 0.24 deg. The fixation square was always at zero disparity. Disparity and depth for a given element changed constantly throughout the trial in a controlled fashion to ensure that the element always remained within the volume that it was assigned to. Initial element positions were generated randomly. In the one-volume case, all elements were placed within volume A. In the two-volumes case, volumes B

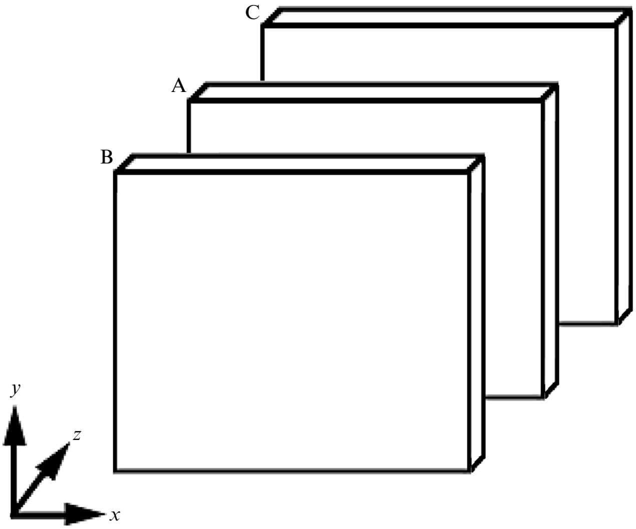

Figure 12. Schematic diagram of frontoparallel depth volumes used in experiment 6. See text for details.

and C were shown and targets and distractors were divided equally between these volumes.

7.1.4 *Procedure.* The procedure of this experiment was the same as for experiment 2.

7.1.5 *Instructions to observers.* The instructions to observers were the same as before.

7.2 *Results*

The performance levels of observers are summarized in figure 13. A two-way repeated-measures ANOVA, with number of targets and the volume criterion, ie one depth volume versus two depth volumes, as the two factors, was performed on the data. The effect of number of targets was significant ($F_{3,12} = 134.13$, $p < 0.0001$). Performance in the two-volumes case was significantly better than in the one-volume case ($F_{1,4} = 34.96$, $p = 0.0041$). The effect of the interaction was not significant ($F_{3,12} = 1.81$, $p = 0.1982$).

Proportion of variance calculations reveal that the variance due to the number of targets constitutes 83.84% of the total variance while the variance due to the volume factor (one-volume versus two-volumes) constitutes 22.03% of the total variance.

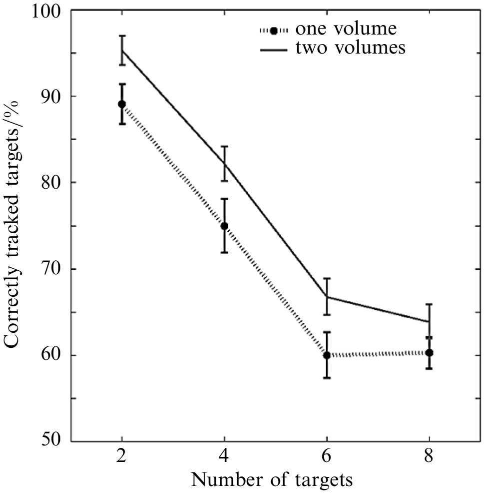

Figure 13. Experiment 6. Mean results for one volume versus two volumes across subjects. Bars represent standard errors.

7.3 *Discussion*

The results of experiment 6 suggest that it is easier to perform the multi-element tracking task for the two-volumes case than for the one-volume case. This is similar to experiment 5 where the effect of the surface factor was found to be significant. Clearly, the additional spatial segregation between targets and distractors in the two-volumes case over the one-volume case is important for improving performance in the multi-element tracking task.

8 Discussion and conclusions

We investigated the effect on performance of the addition of depth cues to a multi-element tracking task. Experiment 1 shows that depth cues, such as disparity and occlusion through T-junctions, improve performance in the special case when element boundaries are allowed to intersect on a flat monitor screen. Experiments 2 and 5 show that the allocation of attention across two depth planes or surfaces is easier than within a single depth plane or surface. Experiments 3 and 4 show that there is no difference in performance when two colors or forms are used instead of a single color or form. In none of these experiments were subjects told that targets would be equally distributed between planes, colors, or forms. Although we cannot rule out the possibility that subjects deduced this fact over the progress of trials, any resulting response

Table 1. Partial $\hat{\omega}^2$: Proportion of variance due to experimental factors.

Variance	Experiment				
	2. Depths	3. Colors	4. Forms	5. Planes	6. Volumes
	10.16%	0.81%	0.13%	19.57%	22.03%

strategies for guessing would only work against our main contention. When put together, as shown in table 1, these results suggest that, for certain experimental constructions, the allocation of attention in depth can be easier than in a completely two-dimensional scene.

The current data are in accordance with data obtained from visual search experiments that show that depth is a useful distinguishing feature between targets and distractors. Holliday and Braddick (1991) showed that a target defined by stereoscopic slant can be detected preattentively. Two more visual search studies, those of He and Nakayama (1995) and Nakayama and Silverman (1986), showed that when attention is allocated to a particular surface, there is little interference from distractors on different surfaces. The present results suggest that, when surface information alone cannot be used to distinguish between targets and distractors, division of attention between two surfaces aids in preferentially attending to targets over distractors. This might be because 'belonging' to one or the other surface provides an additional degree-of-freedom for maintaining an existing distinction between a 'target' and 'distractor' label after two objects collide along the line of sight. The results of experiment 6 suggest that spatial separation in three dimensions, rather than confinement to surfaces as such, may explain the improved performance for the two-surface case.

In most of the studies of attention in depth (Andersen 1990; Andersen and Kramer 1993; Atchley et al 1997; Downing and Pinker 1985; Gawryszewski et al 1987; He and Nakayama 1995; Hoffman and Mueller 1994; Marrara and Moore 2000) focused attention was used to measure the movement of attention in depth, and switching attention from one location to another within the same depth plane or surface was found to be easier than switching attention from one depth plane or surface to another. However, the positions of targets and distractors in these displays remained fixed. The experiments described here show that it is possible to selectively attend to targets that move in depth as well as horizontally and vertically in the presence of identical distractors that move in a similar fashion. That said, it may be that the particular demands of our task account for the two-plane advantage, while other tasks, such as those just cited, elicit no such advantage (Atchley and Kramer 2001; Egly et al 1994).

It is important to consider the implications of the present results to the theories of how object-based attention is deployed and maintained within a scene. If the FINST hypothesis of Pylyshyn and colleagues is accepted, the current results argue that the maintenance and movement of the attentional indexes of targets is easier when these targets move in depth in addition to horizontally and vertically. However, it is not intuitively clear why this should be the case. The virtual-polygon hypothesis of Yantis (1992), however, provides a plausible explanation for the trend observed in the present experiments. This hypothesis suggests that the visual system perceptually groups the target elements into a single virtual polygon and then tracks that polygon to perform the task of multi-element tracking. If this is true, then, given the randomly chosen trajectories in the current experiments, the virtual polygon is more likely to remain convex, and hence coherent as an object, when the vertices of the polygon (here, the elements being tracked) span several depths rather than lie on a single depth plane. Yantis (1992) found that 'edge collapses' of a virtual polygon, which occurred when the polygon changed from being convex to concave or vice versa, adversely affected performance in the tracking task. In the present experiments, fewer edge collapses in three-dimensional space are likely to happen when the vertices of the polygon lie at different depths from the observer than when they are constrained to lie in a single depth plane. This theory predicts a trend in performance that is in the same direction as those observed in experiments 2 and 4, namely, that multi-element tracking across two depth planes or surfaces is easier than within a single depth plane or surface.

Acknowledgments. LV was supported in part by NSF IRI-94-01659, ONR N00014-95-1-0657, and ONR N0014-95-1-0409. EM was supported in part by ONR N00014-95-1-0409, and ONR N00014-01-1-0624. The authors thank Brian Scholl and three anonymous reviewers for helpful suggestions.

References

Andersen G J, 1990 "Focused attention in three-dimensional space" *Perception & Psychophysics* **47** 112 – 120

Andersen G J, Kramer A F, 1993 "Limits of focussed attention in three-dimensional space" *Perception & Psychophysics* **53** 658 – 667

Atchley P, Kramer A F, 2001 "Object and space-based attentional selection in three-dimensional space" *Visual Cognition* **8** 1 – 32

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 and Review* **4** 524 – 529

Becker L, Egeth H, 1998 "How object motion and cued location interact to produce inhibition-of-return" *Investigative Ophthalmology & Visual Science* **39**(4) S2950

Blaser E, Pylyshyn Z W, Domini F, 1999 "Measuring attention during 3D multielement tracking" *Investigative Ophthalmology & Visual Science* **40**(4) S277

Cavanagh P, Labianca A T, Thornton I M, 2001 "Attention-based visual routines: sprites" *Cognition* **80**(1/2) 47 – 60

Chun M, Cavanagh P, 1997 "Seeing two as one: Linking apparent motion and repetition blindness" *Psychological Science* **8** 74 – 79

Culham J C, Brandt S A, Cavanagh P, Kanwisher N G, Dale A M, Tootell R B H, 1998 "Cortical fMRI activation produced by attentive tracking of moving targets" *Journal of Neurophysiology* **80** 2657 – 2670

Downing C, Pinker S, 1985 "The spatial structure of visual attention", in *Attention and Performance XI* Eds M I Posner, O S M Marin (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 171 – 187

Driver J, Baylis G C, 1989 "Movement and visual attention: The spotlight metaphor breaks down" *Journal of Experimental Psychology: Human Perception and Performance* **15** 448 – 456

Duncan J, 1984 "Selective attention and the organization of visual information" *Journal of Experimental Psychology: General* **113** 501 – 517

Egeth H, Yantis S, 1997 "Visual attention: Control, representation, and time course" *Annual Review of Psychology* **48** 269 – 297

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** 161 – 177

Eriksen C W, St James J D, 1986 "Visual attention within and around the field of focal attention: A zoom lens model" *Perception & Psychophysics* **40** 225 – 240

Gawryszewski L D G, Riggio L, Rizzolatti G, Umilta C, 1987 "Movements of attention in the three dimensions and the meaning of 'neutral' cues" *Neuropsychologia* **25** 19 – 29

Ghirardelli T G, Folk C L, 1996 "Spatial cuing in a stereoscopic display: Evidence for a 'depth-blind' attentional spotlight" *Psychonomic Bulletin & Review* **3** 81 – 86

He Z J, Nakayama K, 1995 "Visual attention to surfaces in three-dimensional space" *Proceedings of the National Academy of Sciences of the USA* **92** 11155 – 11159

Hoffman J E, Mueller S, 1994 "An in-depth look at visual attention", paper presented at the 35th annual meeting of the Psychonomic Society, St Louis, MO, November 1994

Holliday I E, Braddick O J, 1991 "Pre-attentive detection of a target defined by stereoscopic slant" *Perception* **20** 355 – 362

Iavecchia H P, Folk C L, 1995 "Shifting visual attention in stereographic displays: A timecourse analysis" *Human Factors* **36** 606 – 618

Intriligator J M, 1997 *The Spatial Resolution of Visual Attention* PhD thesis, Department of Psychology, Harvard University, Cambridge, MA

Kahneman D, Treisman A, 1984 "Changing views of attention and automaticity", in *Varieties of Attention* Eds R Parasuraman, D R Davies (New York: Academic Press) pp 29 – 61

Kahneman D, Treisman A, Gibbs B J, 1992 "The reviewing of object files: Object-specific integration of information" *Cognitive Psychology* **24** 174 – 219

Kanwisher N, 1987 "Repetition blindness: Type recognition without token individuation" *Cognition* **27** 117 – 143

Keppel G, 1991 *Design and Analysis: A Researcher's Handbook* (Englewood Cliffs, NJ: Prentice-Hall)

Kramer A F, Jacobson A, 1991 "Perceptual organization and focused attention: The role of objects and proximity in visual processing" *Perception & Psychophysics* **50** 267 – 284

LaBerge D, Brown V, 1989 "Theory of attentional operations in shape identification" *Psychological Review* **96** 101 – 124

Marrara M T, Moore C M, 2000 "Role of perceptual organization while attending in depth" *Perception & Psychophysics* **62** 786 – 799

Nakayama K, Shimojo S, Silverman G H, 1989 "Stereoscopic depth: its relation to image segmentation grouping and the recognition of occluded objects" *Perception* **18** 55 – 68

Nakayama K, Silverman G H, 1986 "Serial and parallel processing of visual feature conjunctions" *Nature* **320** 264 – 265

Neisser U, Becklen R, 1975 "Selective looking: Attending to visually specified events" *Cognitive Psychology* **7** 480 – 494

Posner M I, 1980 "Orienting of attention" *Quarterly Journal of Experimental Psychology* **32** 3 – 25

Pylyshyn Z W, 1989 "The role of location indexes in spatial perception: A sketch of the FINST spatial index model" *Cognition* **32** 65 – 97

Pylyshyn Z W, 1994 "Some primitive mechanisms of spatial attention" *Cognition* **50** 363 – 384

Pylyshyn Z W, Burkell J, Fisher B, Sears C, Schmidt W, Trick L, 1994 "Multiple parallel access in visual attention" *Canadian Journal of Experimental Psychology* **48** 260 – 283

Pylyshyn Z W, Storm R W, 1988 "Tracking multiple independent targets: Evidence for a parallel tracking mechanism" *Spatial Vision* **3** 179 – 197

Rock I, Gutman D, 1981 "The effect of inattention on form perception" *Journal of Experimental Psychology: Human Perception and Performance* **7** 275 – 285

Scholl B J, 2001 "Objects and attention: the state of the art" *Cognition* **80** 1 – 46

Scholl B J, Pylyshyn Z W, 1999 "Tracking multiple items through occlusion: Clues to visual objecthood" *Cognitive Psychology* **38** 259 – 290

Scholl B J, Pylyshyn Z W, Franconeri S L, 1999 "When are featural and spatiotemporal properties encoded as a result of attentional allocation?" *Investigative Ophthalmology & Visual Science* **40**(4) S4195

Theeuwes J, Atchley P, Kramer A F, 1998 "Attentional control within three-dimensional space" *Journal of Experimental Psychology: Human Perception and Performance* **24** 1476 – 1485

Tipper S P, Lortie C, Baylis G C, 1992 "Selective reaching: Evidence for action-centered attention" *Journal of Experimental Psychology: Human Perception and Performance* **18** 891 – 905

Tipper S P, Weaver B, Jerreat L M, Burak A L, 1994 "Object-based and environment-based inhibition of return of visual attention" *Journal of Experimental Psychology: Human Perception and Performance* **20** 478 – 499

Tyler C W, Kontsevich L L, 1995 "Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction" *Perception* **24** 127 – 153

Viswanathan L, Mingolla E, 1998a "Attention in depth: Disparity and occlusion cues facilitate multi-element visual tracking" Technical Report CAS/CNS-TR-98 – 012, Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University, Boston, MA

Viswanathan L, Mingolla E, 1998b "Disparity and occlusion cues facilitate multi-element visual tracking" *Investigative Ophthalmology & Visual Science* **39**(4) S2959

Wolfe J M, Bennett S C, 1997 "Preattentive object files: Shapeless bundles of basic features" *Vision Research* **37** 25 – 43

Yantis S, 1992 "Multielement visual tracking: Attention and perceptual organization" *Cognitive Psychology* **24** 295 – 340

Yantis S, Johnson D, 1990 "Mechanisms of attentional priority" *Journal of Experimental Psychology: Human Perception and Performance* **16** 812 – 825

Yantis S, Jones E, 1991 "Mechanisms of attentional selection: Temporally-modulated priority tags" *Perception & Psychophysics* **50** 166 – 178