1999-04

Dynamics of Attention in Depth: Evidence from Mutli-Element Tracking

https://hdl.handle.net/2144/2227

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

Lavanya Viswanathan and Ennio Mingolla

April, 1999

Technical Report CAS/CNS-99-010

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215

DYNAMICS OF ATTENTION IN DEPTH: EVIDENCE FROM MULTI-ELEMENT TRACKING

Lavanya Viswanathan[1] and Ennio Mingolla[2]
Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215

March 1999

Key Words: motion, attention, visual tracking, depth, occlusion, T-junctions, surfaces.

Running Head: Attention in Depth

Requests for reprints should be sent to:
Lavanya Viswanathan
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215

---

[1] Supported in part by the National Science Foundation (*NSF IRI-94-01659*) and the Office of Naval Research (*ONR N00014-95-1-0409* and *ONR N00014-95-1-0657*).

[2] Supported in part by the Office of Naval Research (*ONR N00014-95-1-0409*).

ABSTRACT

The allocation of attention in depth is examined using a multi-element tracking paradigm. Observers are 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 fronto-parallel plane. We also show that the allocation of attention across two perceptually distinguishable planar surfaces, either fronto-parallel 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 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.

1 INTRODUCTION

In daily life, our visual system is 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. Besides, the objects themselves are rarely equidistant from the observer. It would, therefore, seem natural that the visual system must be able to selectively deploy attention to non-contiguous moving objects that may lie at different distances from the observer. In this paper, 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 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). Such approaches have come to be known as object-based 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, Treisman and Gibbs 1992), attentional priority tags (Yantis and Johnson 1990; Yantis and Jones 1991), attentional sprites (Cavanagh 1996), object tokens (Chun and Cavanagh 1997; Kanwisher 1987) or Fingers of INSTantiation (FINST; 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 come 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, Weaver, Jerreat and Burak 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 studies (Ghirardelli and Folk 1996; Iavecchia and Folk 1995; Theeuwes, Atchley and Kramer 1998) argue that attention cannot be preferentially allocated to specific locations in depth. Other studies disagree and suggest that when deployed in depth, attention could be either viewer-centered (i.e., with a shallow gradient between the observer and the target and a steeper gradient beyond the target), object-centered (i.e., with the same slope on either side of the target) or action-centered as in selective reaching tasks (Tipper, Lortie and Baylis 1992). Two of the earliest studies in this area (Downing and Pinker 1985; Gawryszewski, Riggio, Rizzolatti and Umilta 1987) studied the movement of attention using a cuing paradigm in a real 3-D scene and showed that the visual system can attend to specific locations in depth. They 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 3-D scenes with binocular disparity information and obtained conflicting results. Ghirardelli and Folk (1996), Iavecchia and Folk (1995) and Theeuwes, Atchley and Kramer (1998) showed no effect of cuing in depth. Andersen (1990), Andersen and Kramer (1993), Atchley, Kramer, Andersen and Theeuwes (1997), Hoffman and Mueller (1994) and Marrara and Moore (1998) 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 1998). All of these studies used spatial cuing paradigms. Two studies using 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. Further, Honda and Findlay (1992) found that saccades to targets in different depth planes had longer saccadic latencies than saccades to targets in the same depth plane.

Other studies suggest that the deployment of attention across iso-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, p143).

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 could be constructed whose vertices spanned different depths or surfaces 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, Burkell, Fisher, Sears, Schmidt and Trick 1994). The FINSTs can be assigned to objects either through bottom-up factors such as attentional capture or through top-down factors.

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 (i.e., the items to be tracked). Yantis (1992) noted that 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, whereas 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 nontarget 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, i.e., 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 (1997, 1998) found that successful tracking, i.e., 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, Pylyshyn and Franconeri (1999) showed that attentional allocation in multi-element tracking results in an encoding of spatiotemporal, but not featural, properties of objects. Culham, Brandt, Cavanagh, Kanwisher, Dale and Tootell (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 will attempt to address 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 and 4, we look at whether attention can be simultaneously allocated across two depth planes, colors 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 2D scene; that is, where elements moved only in an up-and-down or left-and-right fashion in a fronto-parallel 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, p182):

> *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 signalling occlusion, improve performance in a multi-element tracking task when elements are allowed to overlap one another dynamically during a trial. We have found 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 the current study used displays that mimicked those used by Pylyshyn and Storm (1988) and Yantis (1992), we next describe their paradigm in detail. Pylyshyn and Storm (1988) constructed two-dimensional displays that contained ten moving white crosses (plus, +, signs) (figure 1). A randomly chosen subset 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 7s to 15s. At the end of the movement phase, a solid white square was flashed over (i.e., 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.

2.1 Method

2.1.1 Observers

Eight naive observers participated in two sessions of 45 minutes 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.

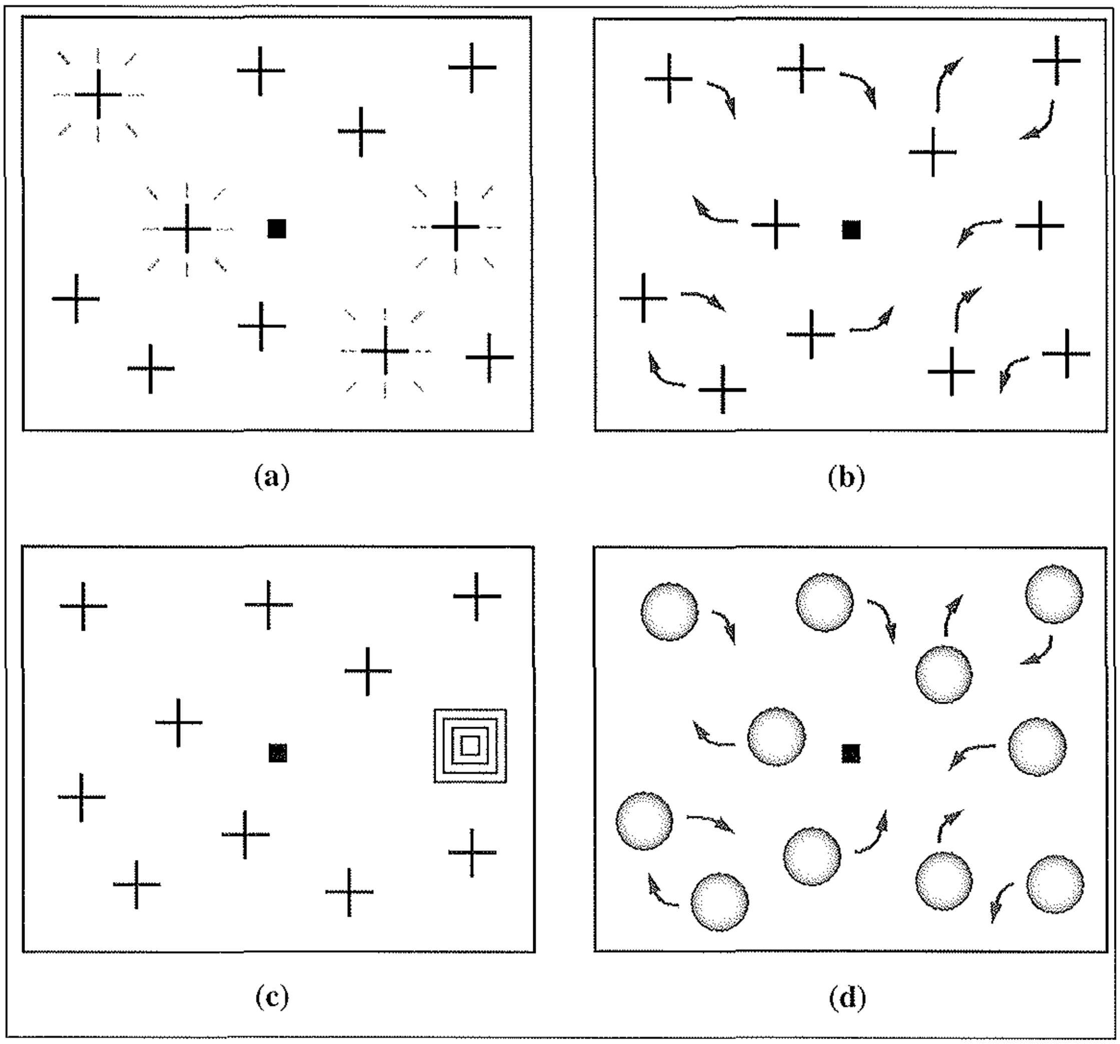

FIGURE 1. Schematic diagram of the types of displays used by Pylyshyn and Storm (1988) and Yantis (1992). These displays contain ten identical plus signs that move randomly: (a) *target designation* phase, i.e., the targets are flashed on and off; this is diagrammed by lines radiating from the targets; (b) *movement phase*, i.e., all the elements move in randomly selected directions; (c) *probe phase*, i.e., one element is randomly chosen as the probe for the given trial and this element is replaced by a concentric set of squares at the end of the movement phase. Observers must specify whether the squares appear on a target element (i.e., one that they have been tracking) or on a non-target element; (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 and 4. Trajectories in experiment 1 were linear.

2.1.2 Design

Two independent depth cues were considered: binocular disparity and T-junctions signalling occlusion. These cues were either present or absent in a trial. This led to a 2 X 2 design generating the following four experimental conditions:
Condition 1. neither disparity nor shading is present;
Condition 2. shading is absent and only disparity is present;
Condition 3. only shading is present and disparity is absent;
Condition 4. both disparity and shading are present.

2.1.3 Materials

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 and right eye images. This ensured that displays containing differing disparity information could be presented to each eye separately. The screen resolution was 1025x768 and the frame rate was 60 Hz for both eyes (i.e., 30 Hz per eye).

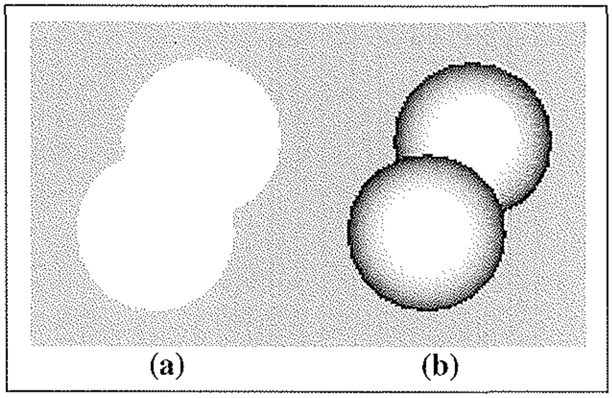

FIGURE 2. Formation of T-junctions during element intersections: (a) the elements are not shaded, so no T-junctions are formed and the display does not appear to be in depth; (b) shading of the elements creates strong contours that lead to T-junctions when the elements intersect, giving a percept of depth due to occlusion.

Our displays were based on those of Pylyshyn and Storm (1988) and Yantis (1992) (figure 1). The display contained 10 identical elements (disks or spheres). Two types of depth cues were used: binocular disparity and shading. Shading the white disks makes them look more spherical and gives the impression of three-dimensional structure (figure 2). All the elements are shaded identically, with a uniform gradation between white at the center and black at the boundary. The important result of shading for our purposes, however, is that 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, Shimojo and Silverman 1989). Non-shaded white disks do not form T-junctions when they overlap; instead, they form figure-eight regions, with no depth ordering. A control experiment presented elsewhere using flat white disks with thick black outlines instead of shaded spheres showed that occlusion through T-junctions, rather than the three-dimensional appearance of a shaded sphere, was important (Viswanathan and Mingolla 1998b).

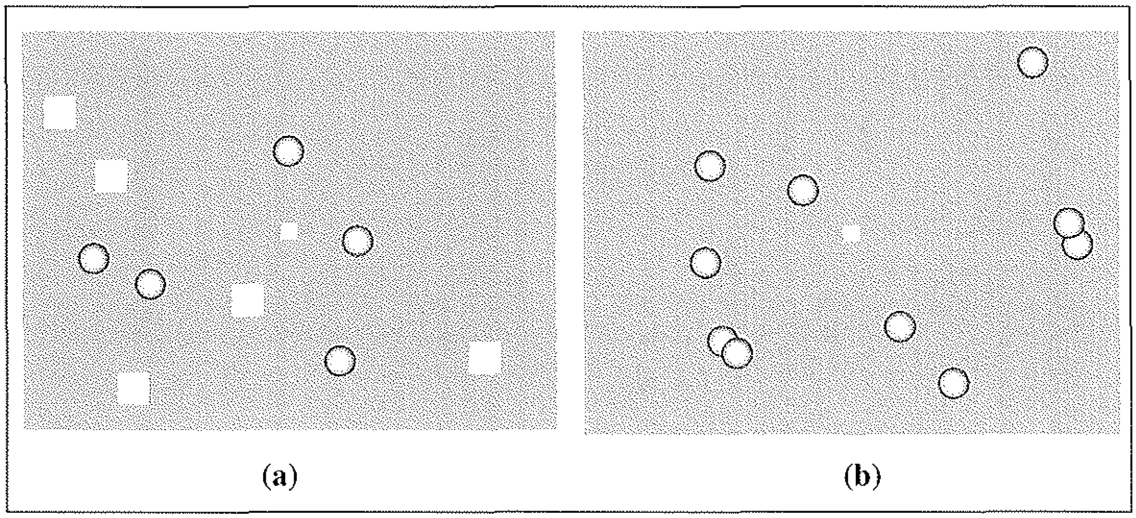

FIGURE 3. Two frames of the display: (a) the target designation phase, with the target set being "flashed"; (b) one frame of the movement phase in which two overlaps may be seen.

Each trial consisted of four 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, i.e., replaced by a white 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 0.8°, i.e., equal to the width of one element.

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 16 possible directions of movement. The angular separation between any two adjacent directions was 22.5 degrees. 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. In displays with either disparity or shading 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°. Trajectories were precomputed and chosen so that not more than three disks overlapped significantly at a single spatial location. Elements bounced off the edges of the depth volume, i.e., 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. Otherwise, elements maintained their trajectories and always moved in a straight line. 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 200 static frames displayed at a display rate of 60 Hz for two stereoscopic buffers, i.e., 30 Hz per eye. The movement phase lasted approximately 7.5 seconds.

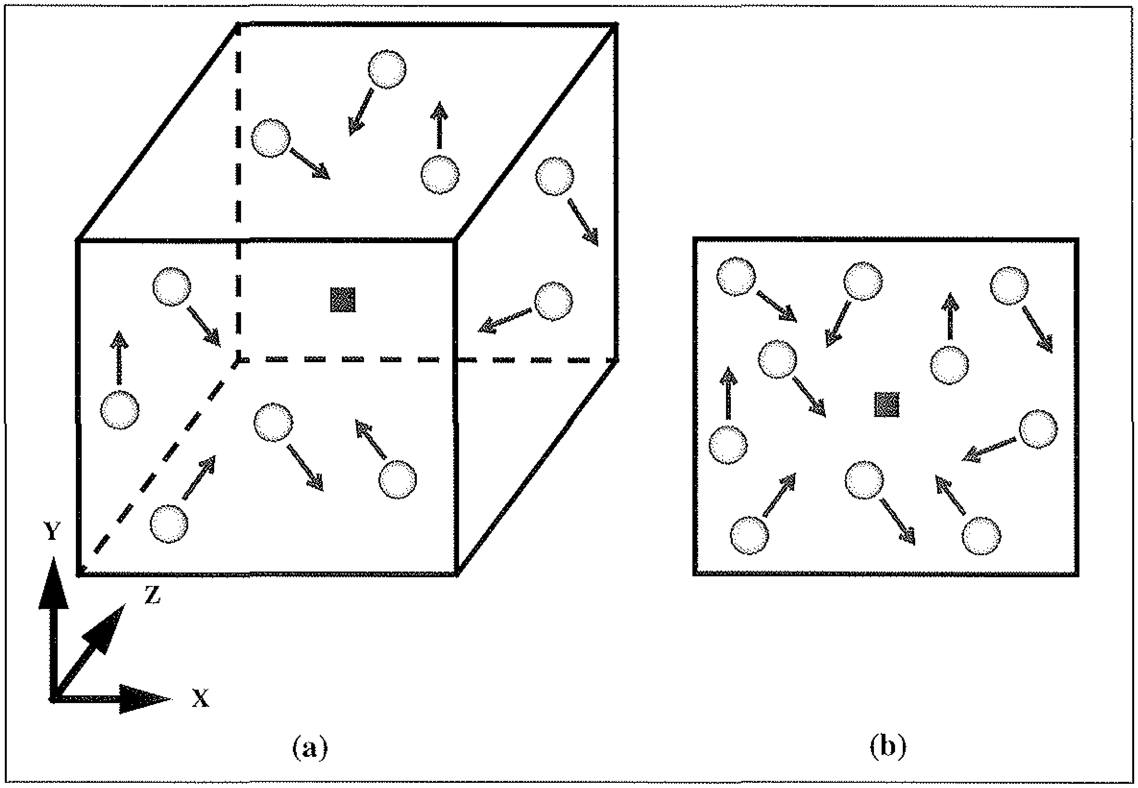

FIGURE 4. Schematic diagram of element trajectories in experiment 1: (a) the elements move linearly in a three-dimensional depth volume. Elements bounce off the edges of the volume as well as an invisible cushion around the fixation square. Depth (z) is defined by binocular disparity; (b) view as seen on the two-dimensional projection plane of the monitor screen

Probe phase: At the end of the movement phase, motion was stopped and the probe for the current trial (which was a target for 50 percent of the trials of a given experimental condition and a non-target for the remaining trials of that condition) was flashed three times, again by replacing it by a white square. No overlaps were allowed in this phase so that there was no doubt about the identity of the element being flashed or any of the other elements in the display.

Query phase: After the probe was flashed, a screen containing a query message was shown and observers were asked to press the LEFT arrow button on the keyboard if they thought the probe flash had occurred on a target element and the RIGHT arrow button if they felt that it was on a non-target element. Since the emphasis of this experiment was accuracy of tracking and not response latencies, the query screen was displayed for as long as the observer needed to make a judgement.

The depth volume subtended visual angles of 11.4^0 in width and 8.2^0 in height. The disparity of the closest surface of the volume was -0.26^0 and that of the far surface was 0.26^0. Each element (disk or sphere) subtended a visual angle of 0.8^0 vertically and horizontally. The white flashing square that replaced an element during flashes was 0.8^0 wide. The fixation square subtended a visual angle of 0.4^0. The speed of movement of each element was 2.0^0/sec. The background was colored cyan. The spheres, the flash squares and the fixation square were white. 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.

2.1.4 Procedure

The experiment consisted of five blocks: one practice block and four experimental blocks. The practice block lasted around 6 minutes and each experimental block lasted around 15 minutes. The practice block contained 16 trials and each experimental block contained 40 trials (10 for each experimental condition), for a total of 160 trials (40 per condition). The starting parameters (element positions, directions of movement, target set) for 44 trials (4 practice trials and 10 trials for each experimental block) were precomputed. Each one of the trajectories was then presented in every one of the experimental conditions. To minimize the possibility of subjects memorizing initial target configurations and element trajectories, each precomputed trial was flipped either left to right (the new X position of each element at each frame was set equal to the display width minus the old X position of the element at that frame) or top to bottom (the new Y position of each element at each frame was set equal to the screen height minus the old Y position of the element at that frame) or both. These controls had the effect of ensuring that no two trials that the observer saw were exactly the same, while also ensuring that the same trajectories were presented for each experimental condition. 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 within a block.

The probes for the trials were randomized and the only restriction imposed on their choice was that 50 percent of the probed elements for a given experimental condition and within a block be targets and the remaining be non-targets.

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. They were instructed to press the LEFT arrow button on the keyboard if they felt that the probe was on a target and the RIGHT arrow button if they thought it was on a non-target.

Observers received feedback during the practice block. If their answer was correct, a screen showing the message "CORRECT!" was displayed, otherwise a screen with the message "WRONG!" was shown. 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 five minutes between blocks, as it was essential for the experiment that they concentrate completely on the tracking task.

2.2 Results

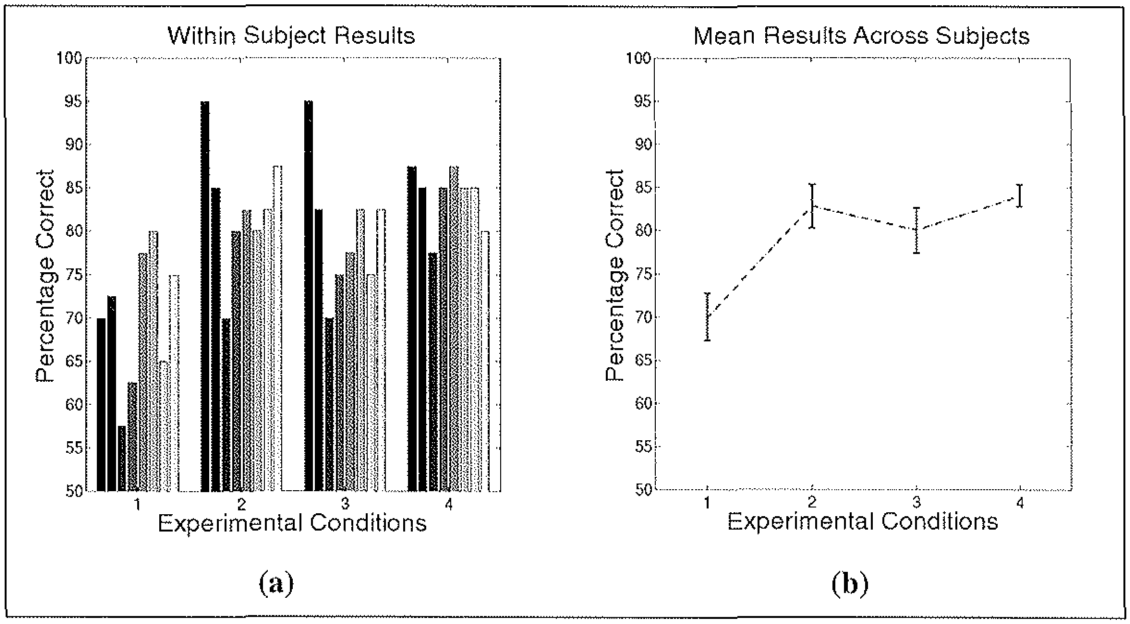

FIGURE 5. Experiment 1: (a) results for individual subjects; each bar represents the performance level of a subject for a given experimental condition; (b) mean results across subjects for each experimental condition; bars represent standard errors. Note that in condition 1, neither disparity nor shading are present; in condition 2, only disparity is present; in condition 3, only shading is present; in condition 4, both disparity and shading are present.

The performance levels of observers are summarized in figure 5. The main result to be noted from these data is that for each observer, performance was worst under experimental condition 1, i.e., when the displays contained neither disparity nor occlusion cues for depth. The addition of depth cues bettered performance for all observers. For some, T-junctions proved to be a stronger cue, while for others disparity worked better. In all cases, the improved performance levels were comparable to those found by Pylyshyn and Storm (1988). Moreover, all performance levels were above chance, suggesting that the task of accurately tracking identical elements when element boundaries are allowed to intersect is possible under certain conditions, contrary to the assumptions of Pylyshyn and Storm (1988) and Yantis (1992).

A two-way repeated-measures ANOVA, with shading and disparity as the two factors, was performed on the data. This analysis reveals that the effect of the presence of either factor on performance levels is significant (*disparity*: $F_{1,7} = 32.09$, $p = 0.0008$; *shading*: $F_{1,7} = 21$, $p = 0.0025$). But the effect of the interaction of both cues was not significant ($F_{1,7} = 4.57$, $p = 0.0698$).

2.3 Discussion

Our results corroborate earlier results that human observers can successfully track up to 5 targets in a display consisting of 10 identical elements moving randomly (Pylyshyn and Storm 1988; Yantis 1992). During the brief durations when an element in our displays overlaps another element, the only clues to the element's identity were the continuity of the element's trajectory direction and occlusion cues (either disparity or T junctions or both) that specified which element was in front of, and which behind, the other. Poor performance in the case where no depth cues are present shows that continuity of element motion alone is not sufficient for multi-element tracking. An important conclusion that can be drawn from our 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, Pylyshyn and Domini (1999) finds similar results.

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 vs. 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 iso-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 second case, i.e., the two-depths case.

3.1 Method

3.1.1 Observers

Five observers participated in four sessions of approximately one hour 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. One 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 X 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. The main difference was a change in the response strategy. Observers were now instructed to mark all the elements they had been tracking rather than respond to a single probe element (Intriligator 1997). Each trial consisted of three phases: the *target designation phase*, in which the target set was defined in a static display by flashing red squares on the targets; the *movement* phase, in which all the elements moved in randomly chosen directions; and the *query* phase, in which all elements stopped moving and a cursor appeared on the screen. Observers were instructed to pick out all the elements that they had been tracking by clicking on them with the left mouse button. The target designation phase lasted 5 seconds, the movement phase lasted 12 seconds and the query phase was not timed. Observers were allowed to take as much time as they needed to respond.

The display subtended visual angles of 11.4° in width and 8.2° in height. Elements were two dimensional white disks. Each element subtended a visual angle of 0.4° vertically and horizontally. The red flashing square that replaced an element during flashes was 0.4° wide. The fixation square, which was always presented at zero disparity, subtended a visual angle of 0.4°. The speed of movement of each element was 2.0°/sec. The background was colored cyan. The elements and the fixation square were white. The experiment was conducted under free-viewing conditions, so all reported dimensions in this paragraph are approximate.

The total number of elements was always 16. 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° and 0.17° respectively, one in front of and one behind the plane of the monitor screen. In the one-depth case, i.e., in experimental conditions A through D, all the elements were presented on depth plane A at zero disparity, i.e., 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 C and the back depth plane B. 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.

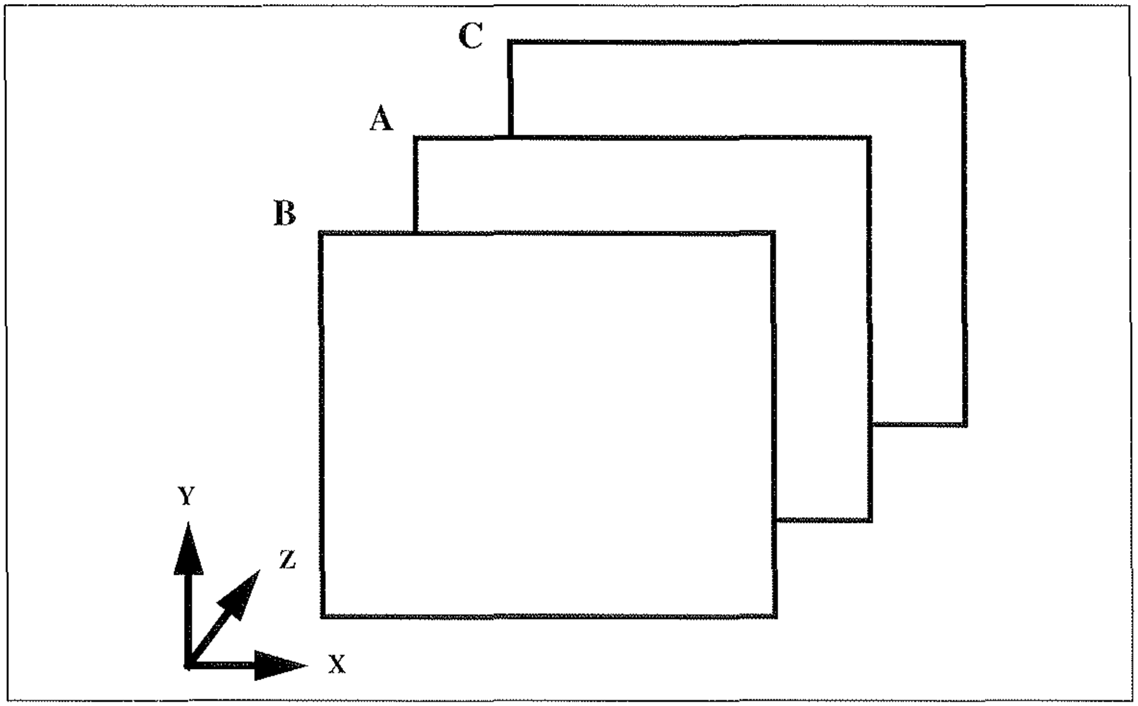

FIGURE 6. Schematic diagram of fronto-parallel depth planes used in experiment 2. See text for details.

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 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 1997). 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 0.15^{0} to the fixation square and (b) no two elements came closer than 0.3^{0} 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: two for each of the eight experimental conditions. The experimental blocks contained a total of 240 trials: 30 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 one hour.

3.1.5 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. They were instructed to click on all the elements they had been tracking when motion stopped and a cursor appeared on the screen.

Observers received feedback during the practice block. They were told how many targets they had tracked successfully. No feedback was given during the experimental blocks.

Observers were encouraged, through an on-screen display, to rest their eyes and take breaks between blocks.

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, i.e., one depth plane vs. 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 ($F_{3,12} = 2.19$, $p = 0.1419$).

The proportion of variance due to the two experimental factors can be found by computing *partial omega squared* (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 vs. two-depths) constitutes 10.16% of the total variance.

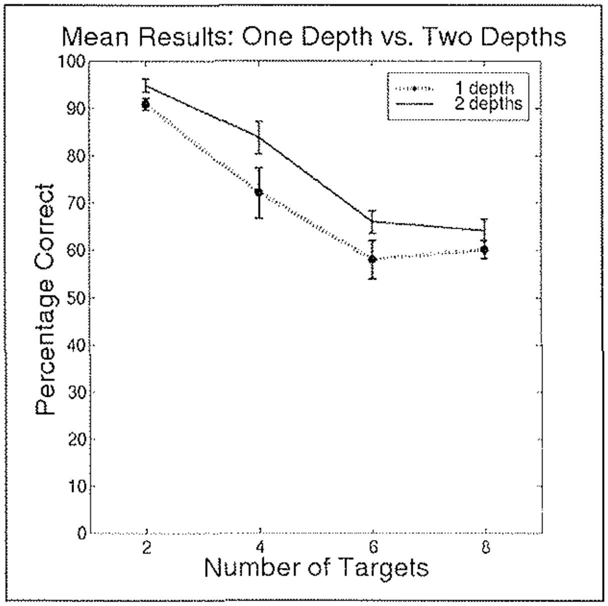

FIGURE 7. Experiment 2: Mean results across subjects; bars represent standard errors.

3.3 Discussion

Our 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 colored the same color. In the two-colors case, half the targets and half the distractors are colored one color while the other half are colored 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 X 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.

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, i.e., one color vs. 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 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 omega squared computations revealed that the variance due to number of targets and the color factor (one-color vs. 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.

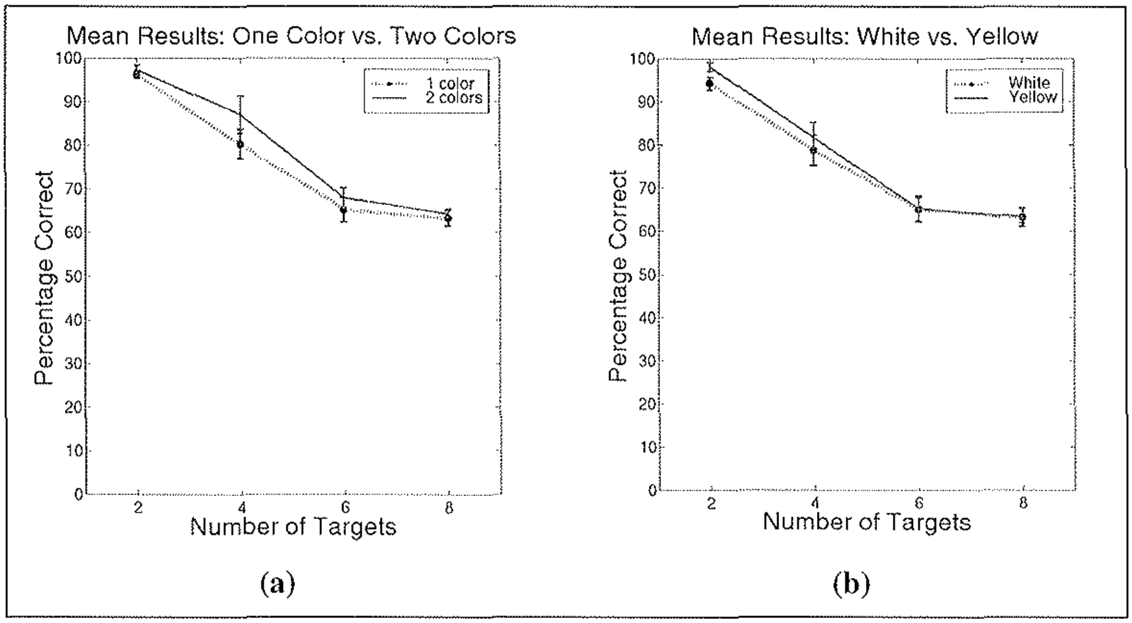

FIGURE 8. Experimental results: (a) mean results for the one color vs. two color factors; (b) mean results for white vs. yellow in the one-color case; bars represent standard errors.

4.3 Discussion

Our 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 only responsible for 0.81% of the total variance in experiment 3. We can, therefore, 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 subtasks is color rather than disparity. Our results also show that tracking across two colors is not more difficult than tracking within a single color.

5 EXPERIMENT 4

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 fronto-parallel? This experiment tests 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. We use the same experimental construction as in experiment 2 except with receding planar surfaces instead of fronto-parallel depth planes.

5.1 Method

5.1.1 Observers

Five observers, including one of the authors, participated in four sessions of approximately 75 minutes 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. One of the subjects had participated in experiments 2 and 3.

5.1.2 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 X 2 experimental construction was used.

5.1.3 Materials

The experimental apparatus and methods were the same as those for experiment 2 except for the following changes. Instead of fronto-parallel depth planes, planar surfaces that receded in depth were depicted (figure 9). Three parallel planar surfaces, which made an angle of 60^0 with the vertical, were used. Surfaces B and C were equidistant from surface A. The disparity difference between A and B was -0.1^0 while that between A and C was 0.1^0. 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 on 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. 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.

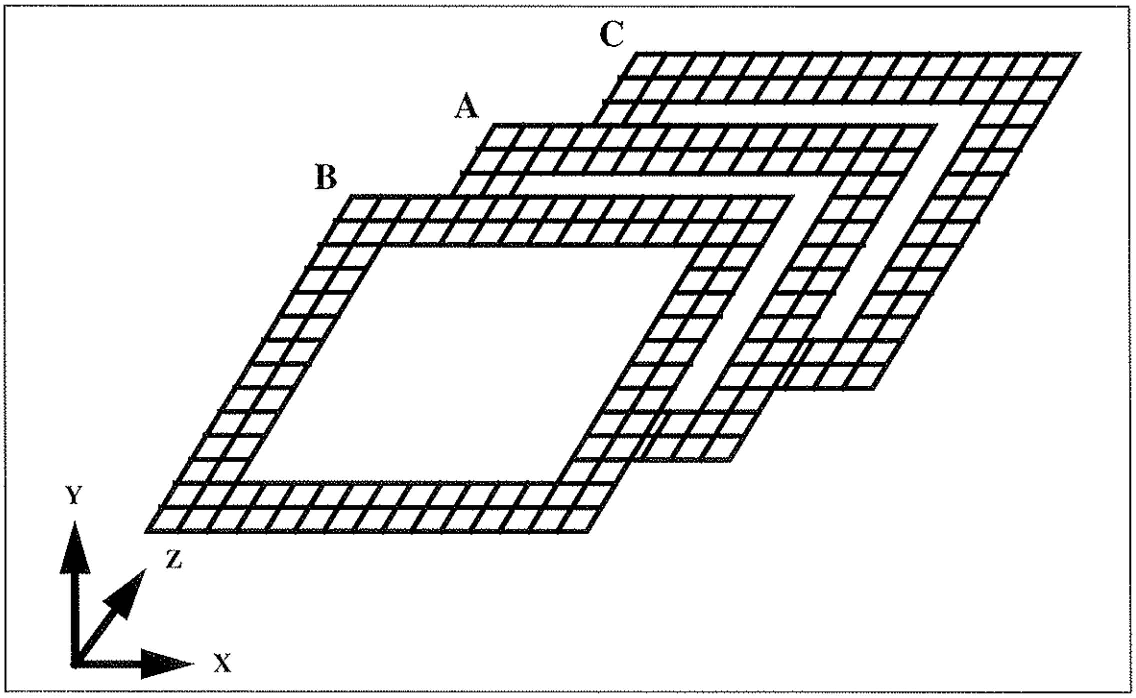

FIGURE 9. Schematic diagram of surfaces used in experiment 4. See text for details.

5.1.4 Procedure

The procedure of this experiment was the same as for experiment 2. Here, 32 trials were presented in each of the eight experimental conditions leading to a total of 256 trials in the experimental blocks. The practice block still contained 16 trials.

5.1.5 Instructions to observers

The instructions to observers were the same as before.

5.2 Results

The performance levels of observers are summarized in figure 10. A two-way repeated-measures ANOVA, with number of targets and surface criterion (one receding planar surface vs. 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 the one-surface case ($F_{1,4} = 18.40$, $p = 0.0128$). The effect of the interaction was also significant ($F_{3,12} = 6.14$, $p = 0.009$).

Partial omega squared computations revealed that the variance due to number of targets and the surface factor (one-surface vs. 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.

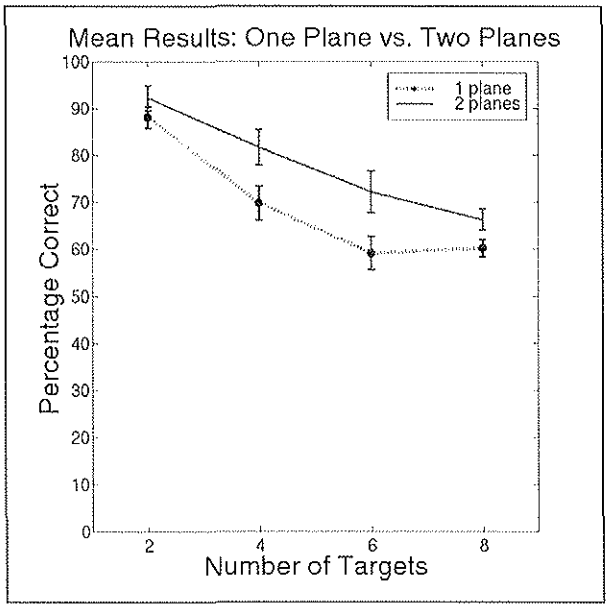

FIGURE 10. Experiment 4: Mean results across subjects; bars represent standard errors.

5.3 Discussion

Our 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 only responsible for 0.81% of total variance, both the depth factor in experiment 2, responsible for 10.16% of total variance, and the surface factor in experiment 4, responsible for 19.57% of total variance, proved to be strong influencing factors on performance in our multi-element tracking task.

6 GENERAL DISCUSSION

The present work has 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 4 show that the allocation of attention across two depth planes or surfaces is easier than within a single depth plane or surface. Experiment 3 shows that there is no difference in performance when two colors are used instead of a single color. When put together, these results suggest that the allocation of attention in depth is easier than in a completely two-dimensional scene.

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 1998) have used focussed attention to measure the movement of attention in depth and found that switching attention from one location to another within the same depth plane or surface is easier than switching attention from one depth plane or surface to another. However, the positions of targets and distractors in these displays remained fixed. We 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. In fact, the maintenance and movement of the attentional indexes of targets is easier when these targets move in depth in addition to horizontally and vertically.

Our data is 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, 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. Our 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.

7 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, 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 and Visual Sciences* **39** abstract no. 2950 (ARVO 1998: Fort Lauderdale)

Blaser E, Pylyshyn Z W, Domini F, 1999 "Measuring attention during 3D multielement tracking" *Investigative Ophthalmology and Visual Sciences* **40** abstract no. 277 (ARVO 1999: Fort Lauderdale)

Cavanagh P, 1996 "Attention-based visual processes" *Canadian Psychology* **37** 59

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: Erlbaum) pp 171-187

Driver J, Baylis G C, 1989 "Movement and visual attention: The spotlight metaphor breaks down" *Journal of Experimental Psychology: Human Perception & 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** 269-197

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 and 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 USA* 11155-11159

Hoffman J E, Mueller S, 1994 "An in-depth look at visual attention" Paper presented at the annual meeting of the Psychonomic Society (St. Louis)

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

Honda H, Findlay J M, 1992 "Saccades to targets in three-dimensional space: Dependence of saccadic latency on target location" *Perception & Psychophysics* **52** 167-174

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, Massachusetts, USA

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 (forthcoming) "Allocating attention in depth: Evidence concerning the necessity of surface information"

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 & Performance* **7** 275-285

Scholl B J, Pylyshyn Z W, 1997 "Tracking multiple items through occlusion: Clues to visual objecthood" Technical Report, Rutgers Center for Cognitive Science, TR-40 (to appear in *Cognitive Psychology*)

Scholl B J, Pylyshyn Z W, 1998 "Tracking multiple items through occlusion: Clues to visual objecthood" *Investigative Ophthalmology and Visual Sciences* **39** abstract no. 4053 (ARVO 1998: Fort Lauderdale)

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 and Visual Sciences* **40** abstract no. 4195 (ARVO 1999: Fort Lauderdale)

Theeuwes J, Atchley P, Kramer A F, 1998 "Attentional control within three-dimensional space" *Journal of Experimental Psychology: Human Perception & 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 & 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 & 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, Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University, CAS/CNS-TR-98-012

Viswanathan L, Mingolla E, 1998b "Disparity and occlusion cues facilitate multi-element visual tracking" *Investigative Ophthalmology and Visual Sciences* **39** abstract no. 2959 (ARVO 1998: Fort Lauderdale)

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 & Performance* **16** 812-825

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