Journal of Experimental Psychology: General
2007, Vol. 136, No. 4, 653–662

0096-3445/07/$12.00 DOI: 10.1037/0096-3445.136.4.653

Visual Short-Term Memory Benefit for Objects on Different 3-D Surfaces

Yaoda Xu
Harvard University and Yale University

Ken Nakayama
Harvard University

Visual short-term memory (VSTM) plays an important role in visual cognition. Although objects are located on different 3-dimensional (3-D) surfaces in the real world, how VSTM capacity may be influenced by the presence of multiple 3-D surfaces has never been examined. By manipulating binocular disparities of visual displays, the authors found that more colored objects could be held in VSTM when they were placed on 2 rather than on 1 planar 3-D surfaces. This between-surface benefit in VSTM was present only when binding of objects' colors to their 3-D locations was required (i.e., when observers needed to remember which color appeared where). When binding was not required, no between-surface benefit in VSTM was observed. This benefit in VSTM could not be attributed to the number of spatial locations attended within a given surface. It was not due to a general perceptual grouping effect either, because grouping by motion and grouping by different regions of the same surface did not yield the same benefit. This increment in capacity indicates that VSTM benefits from the placement of objects in a 3-D scene.

Keywords: visual working memory, depth perception, binocular disparity, stereo display, the binding problem

We are constantly surrounded by a huge number of objects in our environment. To recognize and categorize perceived visual objects, to sustain attended objects across saccades and other visual interruptions, and to use this information to guide behavior and thoughts, quite often we need to store object information in a short-term memory buffer, termed *visual short-term memory* (VSTM; Phillips, 1974; Phillips & Christie, 1977), before further processes can be carried out. Through the use of simple stimuli, such as colored letters or colored squares presented on a flat computer screen, studies have documented that we can encode and remember a maximum of about four objects at a time in VSTM (e.g., Irwin, 1992; Luck & Vogel, 1997; Pashler, 1988; see also Vogel, Woodman, & Luck, 2001). A few factors have been shown to influence the total number of objects that can be held in VSTM, including object complexity (Alvarez & Cavanagh, 2004), object part structure (Xu, 2002a, 2006), the nature of the object features encoded (whether they are from the same or from a different feature dimension; Wheeler & Treisman, 2002; Xu, 2002b), and the global contextual information of the display (i.e., the relationship between objects; Jiang, Olson, & Chun, 2000; Xu, 2000). In recent functional magnetic resonance imaging (fMRI) and event-related potential studies, possible neural correlates of VSTM capacity limitation have been identified (Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006).

Our visual world is 3-D. As such, objects are located at different depths, some closer and some further away. Objects may also be located on different surfaces across depth, for example, on the floor, on the table, or on the wall. Because previous VSTM studies have all presented stimuli on flat, two-dimensional (2-D) screens perpendicular to the line of sight, very little is known about how the presence of multiple 3-D surfaces may influence object representations in VSTM. In particular, it is not clear whether VSTM capacity is fixed regardless of whether objects share the same or different 3-D surfaces or whether VSTM capacity may be modulated by the presence of multiple 3-D surfaces. If the latter were true, it would provide us with a better understanding of what determines VSTM capacity. More important, it may provide us with ways to improve VSTM capacity and improve behavioral performance that depends on it.

In the past 3 decades, a number of studies have examined the effect of 3-D space on visual attention and perception and have made some interesting observations (e.g., Downing & Pinker, 1985; Enns & Rensink, 1990; He & Nakayama, 1995; Kleffner & Ramachandran, 1992; Nakayama & Silverman, 1986; Viswanathan & Mingolla, 2002). In particular, several of these studies examined the effect of planar 3-D surfaces (either perpendicular to the line of sight or slanted and expanding across multiple depths) on visual attention and perception. He and Nakayama (1995; see also Marrara & Moore, 2000) used an attention-cuing paradigm (Egly, Driver, & Rafal, 1994) and reported that the deployment of visual attention is surface-based: During target detection, it is easier to switch attention between locations within the same surface than to switch between different 3-D surfaces. Nakayama and Silverman (1986; see also Theeuwes, Atchley, & Kramer, 1998) found in a visual search study that observers could effortlessly confine their attention to one particular depth plane and could exclude distracting items from a different depth plane. In multiple object tracking, despite the limitation in observers' ability to simultaneously track a few moving target objects among identical

Yaoda Xu, Psychology Department, Yale University; Ken Nakayama, Vision Sciences Laboratory, Psychology Department, Harvard University.

This study was conducted while Yaoda Xu was a postdoctoral research fellow at the Vision Sciences Laboratory, Psychology Department, Harvard University.

This research was supported by an Air Force Office of Scientific Research grant to Ken Nakayama, by James McDonnell Foundation Program in Cognitive Neuroscience Grant JSMF #2002045 to Yaoda Xu, and by National Science Foundation Grant 0518138 to Yaoda Xu. Part of this research was presented at the 43rd Annual Meeting of the Psychonomic Society (2002), Kansas City, Missouri, and at the Third Annual Meeting of the Vision Science Society (2003), Sarasota, Florida. We thank Marvin Chun for help in collecting data for Experiment 3.

Correspondence concerning this article should be addressed to Yaoda Xu, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. E-mail: yaoda.xu@yale.edu

moving distractor objects (Pylyshyn & Storm, 1988), Viswanathan and Mingolla (2002) reported that more objects can be tracked simultaneously on multiple parallel 3-D surfaces than on the same 3-D surface. These studies have suggested that object representations on the same 3-D surface may be bundled together, separate from those on a different 3-D surface. As a result, there is more interference among object representations within the same than between different 3-D surfaces. These studies thus predict that the presence of multiple 3-D surfaces may aid VSTM, such that less interference among objects may result in more objects being retained in VSTM across multiple 3-D surfaces than within the same 3-D surface. In other words, our visual system may take advantage of the structure of the 3-D space and use it to organize object representations in VSTM.

Meanwhile, unlike studies of VSTM, previous studies involving 3-D surfaces have all used tasks that demand "on-line" processing of the stimuli, that is, all stimuli processing is carried out while the stimuli are visible (e.g., target detection, visual search, and multiple object tracking). Thus, if we ensure that objects are well perceived, the presence of 3-D surfaces may not affect how object information is retained in VSTM after objects have disappeared from view. In other words, the presence of multiple 3-D surfaces may only influence perception but not memory. As such, VSTM capacity may be fixed regardless of whether objects share the same or different 3-D surfaces.

When objects are located on the same planar surface perpendicular to the line of sight, they are also located at the same depth. Depth and surface are thus not separated in this case. However, by presenting slanted surfaces expanding across multiple depths and contrasting them with vertical surfaces at the same depth, we can separate the influence of depth and surface. With this approach, He and Nakayama (1995) showed that it is the presence of different surfaces, rather than different depths, that constrains the distribution of visual attention. In this study, we manipulated the binocular disparity of visual displays. In Experiments 1 and 2, we used planar 3-D surfaces that were perpendicular to the line of sight, did not separate surface from depth, and examined their joint effects on VSTM. In Experiment 3, we used slanted surfaces across depth and examined separate contributions from depth and surface on VSTM.

Experiment 1: The Between-Surface Benefit in VSTM

In this experiment, we used sequential presentation and examined whether encoding objects from different 3-D surfaces increases VSTM capacity. We used parallel planar 3-D surfaces perpendicular to the line of sight (see Figure 1C). Because objects on two different 3-D surfaces form two perceptual groups (a front group and a back group), we also included a condition in which a group of circles and a group of squares were presented on the same surface to examine the effect of grouping alone on VSTM.

In a pilot study, we presented six distinctive colors simultaneously either on the same 3-D surface or evenly distributed on two parallel 3-D surfaces. We found no difference in the number of colors that can be retained in VSTM between these two display conditions, $F(1, 11) < 1$. However, many observers commented afterward that it was more difficult to attend to objects on two 3-D surfaces simultaneously than on just one 3-D surface. This is consistent with the He and Nakayama (1995) study as well as with our everyday experience: When we examine objects on different 3-D surfaces in depth, we attend to objects on each surface serially rather than by spreading our attention across multiple 3-D surfaces simultaneously. The added cost of attending simultaneously to multiple 3-D surfaces might have prevented us from observing a between-surface benefit in VSTM. Thus, in all the experiments reported in this article, a sequential presentation paradigm was used: We presented objects first on one 3-D surface and then on a second 3-D surface. Within a block of trials, the presentation order was fixed so that observers could orient their attention to the different 3-D surfaces serially according to the presentation order. Observers therefore had to attend to only one 3-D surface at a time, even when multiple 3-D surfaces were present.

One may wonder whether it is valid to use sequential presentation. Would sequential presentation modulate VSTM capacity as compared with simultaneous presentation? Kumar and Jiang (2005) tested VSTM for locations, colors, and orientations and found that, in all these cases, sequential and simultaneous displays yielded identical VSTM capacity. Thus, the use of the sequential presentation paradigm in our study should not affect the total amount of information that can be stored in VSTM.

Method

Participants. Twelve observers (6 women, 6 men) from the Harvard University campus were recruited. They were between 17 and 40 years of age, had normal color vision and stereovision, and received either payment or course credits for their participation.[1]

Materials and design. We used a standard change detection paradigm (e.g., Luck & Vogel, 1997). Observers were shown a sample display of three colors followed by another sample display of three colors (a total of six colors), and after a brief delay, in a test display, they detected a possible color change to one of the colors presented in the sample displays. In two of the three display conditions, each sample display contained three colored squares. The two sample displays were presented either on the same surface perpendicular to the line of sight (one surface–squares) or on two parallel surfaces perpendicular to the line of sight (two surfaces). To examine the effect of perceptual grouping on VSTM independent of separation by 3-D surfaces, we presented in a third display condition one sample display containing three colored squares and another sample display containing three colored circles, and the two sample displays were presented on the same surface (one surface–squares and circles).

In the two surface condition, to help observers form a stable perception of the two different 3-D surfaces, we marked each surface by a 3 × 3 grid containing nine square black frames (serving as placeholders; Figure 1C). One grid was perceived to be in front of and one behind the computer monitor screen. To match with the two

[1] Observers' visual acuities were obtained from self-report. All our observers were able to read and follow instructions on the computer screen prior to the start of the experiments at the same viewing distance as in the experiments. We were therefore confident that our observers had normal or corrected-to-normal visual acuity as they claimed. Although all of our observers reported to have normal depth and 3-D perception, we did check their stereopsis in our pilot experiment and in Experiment 1. We presented observers with the same two 3-D surfaces used in the main experiment. Each surface contained either three or four colored squares, totaling seven squares across the two surfaces. Observers reported the number of colored squares present in the back surface (either three or four in different trials). All observers tested were fast and accurate in their reports with no errors. We were therefore confident that (a) our 3-D displays worked as intended and (b) observers had normal stereopsis as they claimed. Because we felt that we could trust observers' self-report, we concluded that our screening test was unnecessary and excluded it from later experiments.

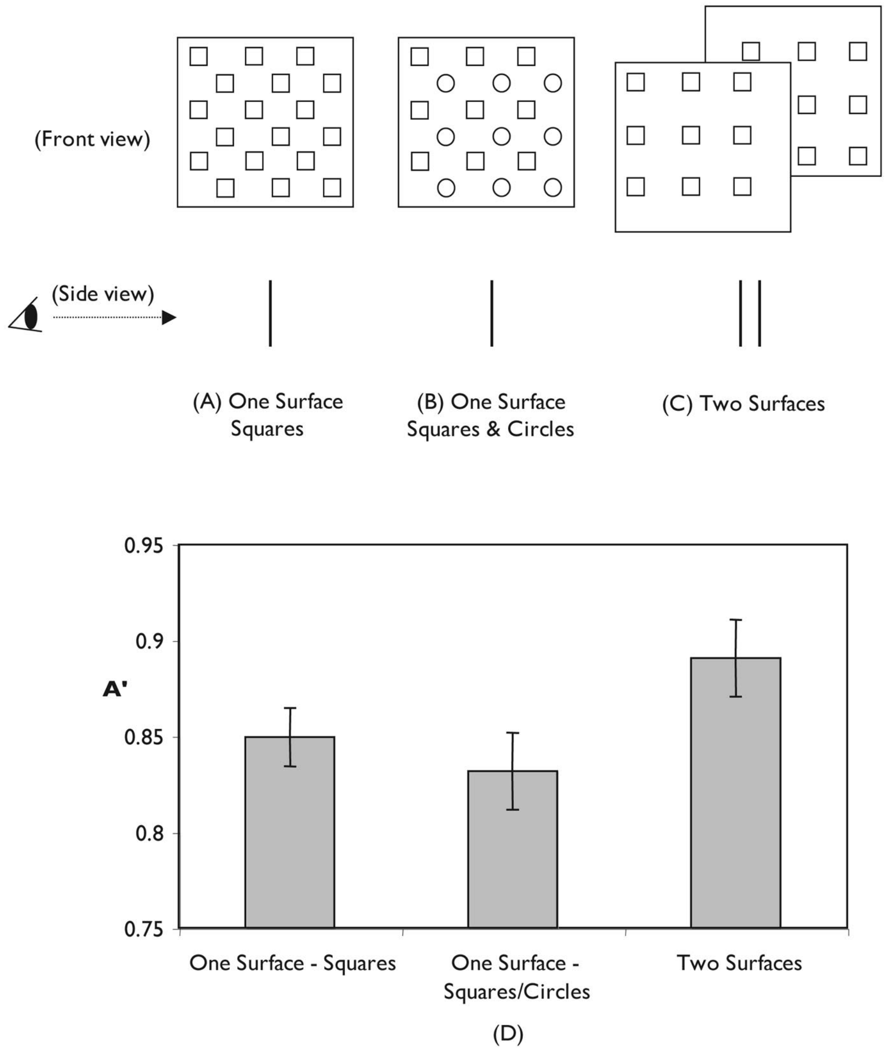

Figure 1. The displays used in Experiment 1. (A) One surface–squares: Only one vertical surface was present, marked by two overlapping 3 × 3 grids, both consisting of square placeholders. (B) One surface–squares and circles: This is the same as Panel A but with one grid consisting of square placeholders and the other consisting of circular placeholders. (C) Two surfaces: Two vertical surfaces were presented at different depths, each marked by a 3 × 3 grid consisting of square placeholders. (D) Results of Experiment 1. A between-surface benefit in visual short-term memory was observed. Error bars indicate 95% within-subject confidence intervals (see Loftus & Masson, 1994).

surfaces condition, in both the one surface–squares and the one surface–squares and circles conditions, we also presented two overlapping 3 × 3 grids, although there was no surface separation between the two grids. In one surface–squares, we marked both grids by square black frames (Figure 1A). In one surface–squares and circles, we marked one grid by square black frames and the other grid by circular black frames (Figure 1B). The grids were present throughout a block of trials, including the intervals between trials. Any color appearing in a trial was always presented inside, filling the black frame of the placeholder at a given location. Therefore, for any given trial, observers would see three colors appearing on one grid in the first sample display followed by three colors appearing on the other grid in the second sample display. The color locations on a grid were randomly chosen except that colors from the two grids could not share the same positions on the grid (i.e., a color on grid 1, row 2, column 1 excluded another color on grid 2, row 2, column 1).

In the test display, a color probe appeared at one of the locations occupied by a sample color (with a chance of 50% to be on each grid). Thus, when two 3-D surfaces were present, the probe would be located on one of the two surfaces with a 50% probability. The probe had either the same color as the sample color at that location (*no change*) or a color not present in either sample display (*change*).

Trials were blocked by display condition, with each block consisting of 4 practice and 32 experimental trials (16 change and 16 no-change trials). There were two blocks of trials for each display condition. In one trial block of the two surfaces condition, observers saw three colors appearing on the front grid and then three colors appearing on the back grid, whereas in the other trial block the presentation order was reversed. Observers could thus orient their attention to the two surfaces sequentially, according to the presentation order of the two sample displays in a given trial block. By doing so, observers had to attend to only one surface at a time, even when objects appeared on two different surfaces. When all the colors appeared on the same surface (one surface–squares and one surface–squares and circles), for a given trial block, similar to two surfaces, observers would see three colors appearing on one of the two grids first and then three colors appearing on the other grid (although there was no surface separation between the two grids).

All displays were presented on a gray background and extended approximately 10.5° × 10.5°. The offset between the two grids in each display was 1.0°. The center-to-center distance between two neighboring placeholders on a given grid was 4.1°. The width of the square and the diameter of the circle were both 1.2°. The disparities of the two 3-D surfaces in two surfaces were +0.05° and −0.05°, respectively, making the total disparity difference between the two 3-D surfaces 0.1°. These disparities were chosen to minimize the conflict between disparity and convergence cues at a viewing distance of 60 cm, while keeping a clear separation between the two surfaces. Eight colors (red, green, yellow, white, cyan, blue, violet, and brown) were used.

To prevent observers from encoding the colors verbally, in this and in all remaining experiments, we included a concurrent verbal suppression task. At the beginning of each trial, four randomly chosen digits (*1* to *9*, with replacement) were presented in green, and observers were instructed to rehearse these digits subvocally throughout a trial. At the end of a trial, another four digits were presented in blue, and observers were asked to detect a digit change that occurred with a 50% probability.

Apparatus. Our apparatus included the MacProbe Macintosh programming software, an Apple computer with a 400-MHz power PC G3 processor and a 17" monitor, and the liquid crystal shutter glasses system from Stereographics Incorporated (http://www.stereographics.com). The monitor resolution was 1024 × 768 pixels, and its refresh rate was 120 Hz/s. The actual refresh rate viewed through the liquid crystal shutter goggle was 60 Hz/s. Reaction times were recorded by the computer.

Procedure. Observers were seated in a dimly lit room approximately 60 cm away from the computer screen. They wore the shutter glasses continuously throughout the experiment, even for trials with no disparity present. Each trial consisted of the following sequential presentation: green digits for 1,000 ms, fixation dot for 500 ms, first sample display for 200 ms, blank delay for 500 ms, second sample display for 200 ms, blank delay for 1,000 ms (memory delay), test display until the observer responded, response feedback for 500 ms, and blue digits until the observer responded. The response keys were the left control key marked *different* for change trials and the enter key on the number keypad marked *same* for no change trials. Both color and digit change detections used the same response keys. Feedback for color change detection was given at the center of the screen as either a happy face for a correct response or a sad face for an incorrect response. Feedback for the digit test was given as a beep for an incorrect response. Within a block, the next trial followed automatically about half a second after the end of the previous trial. A break was given for as long as the observers needed between trial blocks. The experiment lasted about 40 min.

Results

Following Xu (2002a, 2002b), we used the measure of A' from signal detection theory to assess the accuracy of the change detection performance (Grier, 1971; Pollack & Norman, 1964). In pairwise comparisons, colors from two surfaces were better remembered than were colors from either one surface–squares and circles, $F(1, 11) = 9.69$, $p < .05$, Cohen's $d = 0.90$, or one surface–squares, $F(1, 11) = 7.44$, $p < .05$, Cohen's $d = 0.79$. (The difference between the latter two was not significant, $F(1, 11) = 1.41$, $p > .26$, Cohen's $d = 0.34$, and would require testing approximately 69 observers to reach a .05 significance level with a power of 0.80.) Results averaged over the two sample displays are plotted in Figure 1D.[2] (See the Appendix for hit and false-alarm rates for this and subsequent experiments.)

Discussion

These results show that separating objects by 3-D surfaces allowed more object information to be held in VSTM. This between-surface benefit did not result from a simple grouping effect in VSTM as a result of objects forming two perceptual groups on two different 3-D surfaces (i.e., a front and a back group). This is because when we presented objects in two perceptual groups formed by squares and circles on the same surface, we failed to observe any effect of perceptual grouping in VSTM.

It is possible that grouping by squares and circles was a much weaker form of perceptual grouping than was grouping by 3-D surfaces. Thus, the between-surface benefit in VSTM may still be due to a more general grouping effect in VSTM. To test this possibility, we examined grouping by motion in a control experiment. In one condition, the display contained the same kind of grids used before, but one grid continuously jiggled horizontally throughout the presentation of a trial, whereas the other grid stayed stationary.[3] The sample displays thus contained a sequential presentation of three colors on a moving grid and three colors on a stationary grid. The second condition was identical to the one surface–squares condition in Experiment 1. Other aspects of the experiment were the same as Experiment 1. Although the perception of grouping by motion was strong, we failed to observe any improvement in change detection performance when grouping was present compared with when grouping was absent (A's were 0.85 and 0.86, respectively, $F(1, 11) < 1$, Cohen's $d = 0.21$, and would require testing approximately 182 observers to reach a .05 significance level with a power of 0.80). Note that the jiggling motion did not disrupt color change detection, as performance did not differ between the moving and the stationary groups when grouping was present (A's were 0.85 and 0.86, respectively, $F(1, 11) < 1$, Cohen's $d = 0.15$, and would require testing about 359 observers to reach a .05 significance level with a power of 0.80). These results

[2] In all of the experiments reported here, colors from the second sample display tended to be better remembered than those from the first sample display, although this recency effect was not always statistically significant. There was no tendency for the recency effect to interact with the display manipulation (largest $F = 1.98$, smallest $p = .19$).

[3] To create the jiggling motion, we used five evenly spaced locations centered horizontally on the center location of the moving grid. These five locations were labeled as, from left to right, *L2*, *L1*, *M*, *R1*, and *R2*. The spacing between two adjacent locations extended 0.1°. Each jiggling motion consisted of presenting the grid for 100 ms at the five locations following the sequence R1, R2, R2, R1, M, L1, L2, L2, L1, and M (totaling 10 frames). Thus, a complete jiggling motion lasted 1,000 ms. For each moving trial, the moving grid started moving 5,000 ms before the onset of the fixation dot to ensure that two distinct groups could be perceived before the presentation of the sample displays. The grid motion continued throughout the trial and stopped after the probe was presented for 2,000 ms on either the moving or the static grid. The blank delay between the two sample displays was 600 ms. Other timings of the trial were identical to those of Experiment 1.

suggest that the between-surface benefit we observed in Experiment 1 was not due to a general perceptual grouping effect in VSTM.

In our study, objects were in separate regions of the space when they were on two different 3-D surfaces and not so when they were on the same surface. It is possible that when objects were on the same surface, there were higher degrees of "cluttering" and thus more difficulties in specifying locations than when objects were on different surfaces. If so, then better change detection performance should be observed when objects are placed on the left and the right sides of the same 2-D surface than when they are placed on the same side of the 2-D space. To test this, in a second control experiment, in different trial blocks, we had observers perceive, while fixating at the center fixation dot, either two groups of three colors presented sequentially on the same side of the fixation dot or one group of three colors on the left side and one group of three colors on the right side of the fixation dot. The maximum width of the displays and other aspects of the design were identical to those of Experiment 1. We found no advantage of grouping by the left and the right sides of the 2-D space in change detection (A's for grouping presence and absence were 0.81 and 0.80, respectively, $F(1, 11) < 1$, Cohen's $d = 0.06$, and would require testing more than 1,000 observers to reach a .05 significance level with a power of 0.80). It is possible, however, that the 500-ms interval between the offset of the first sample display and the onset of the second sample display was too short to allow attention to shift from one side of the space to the other. When we increased this interval to 1,000 ms, we still failed to observe any difference between these two conditions (A's for grouping presence and absence were 0.83 and 0.82, respectively, $F(1, 11) < 1$, Cohen's $d = 0.21$, and would require testing approximately 173 observers to reach a .05 significance level with a power of 0.80). The between-surface benefit in VSTM therefore could not be generalized to any two regions of the space, and it was not due to the "cluttering" of the display when objects were located on the same surface.

Alvarez and Cavanagh (2005) reported that twice as many moving objects could be simultaneously tracked when they were presented in both the left and the right hemifields compared with when they were all presented within the same hemifield. Using a similar manipulation, Delvenne (2005) found that only VSTM capacity for spatial locations, but not for colors, increased when items were presented in both hemifields as compared with when they were presented within the same hemifield. Thus, although VSTM capacity for spatial location can benefit from grouping by the left and the right sides of the 2-D space, VSTM capacity for colors cannot, which is consistent with the results of this control experiment.

When objects were presented sequentially on different 3-D surfaces, attention could be easily deployed to just one 3-D surface (Nakayama & Silverman, 1986); as such, observers could attend to the 9 locations on one 3-D surface while ignoring those on the other surface. When objects were presented sequentially on the same surface, observers could be obligated to attend to all 18 locations on that surface. Attending to more locations on a given surface might have created more interference and might have made it more difficult to maintain object information in VSTM. To examine how this hypothesis may explain the between-surface benefit in VSTM, in a third control experiment, we varied the number of locations that observers had to attend to on the same surface. We found no difference in change detection performance whether 9 or 18 locations were attended simultaneously on the same surface (A's for these two conditions were 0.80 and 0.83, respectively, $F(1, 11) = 1.09$, $p = .32$, Cohen's $d = 0.30$, and would require testing approximately 90 observers to reach a .05 significance level with a power of 0.80). If anything, the results were slightly opposite to that predicted by the location hypothesis. Thus, varying the number of locations attended to on a given surface did not affect the amount of information retained in VSTM, indicating that the between-surface benefit in VSTM could not be explained by the number of locations simultaneously attended.

Experiment 2A: Separation by Surfaces Improves Color–Location Binding in VSTM

In Experiment 1, because the probe color in the test display always appeared at a location occupied by a sample color, observers were encouraged to bind sample colors to their locations during both VSTM encoding and retention. A change in color and location binding in the test display would indicate that a change had occurred. However, because in change trials the probe color would be a new color not present in the sample displays, observers could also detect a change by remembering which colors were present in the sample displays without necessarily binding colors to locations. Without explicitly controlling for observer strategies, it is difficult to conclude whether color and location binding or color memory in VSTM benefited from the presence of multiple 3-D surfaces in Experiment 1. If the between-surface benefit in VSTM reflected observers' use of the structure of 3-D space to organize the contents of their VSTM, then this benefit should be present only when the binding of features to their locations is required. When such binding is not needed, this between-surface benefit should be absent. In this and the following experiments, we tested this idea.

In this experiment, to encourage color and location binding in the sample displays, as in Experiment 1 the probe color was always presented at a location previously occupied by a sample color. In addition, for change trials, the probe color would always be a color that had appeared elsewhere in the same sample display. Observers therefore needed to explicitly bind colors to locations in order to perform the change detection task successfully.

Method

Participants. Sixteen observers (10 women, 6 men) were recruited from the same participant pool and fulfilling the same criteria as before.

Materials, design, apparatus, and procedure. The two surfaces and one surface–squares conditions from Experiment 1 were included in this experiment. All aspects of the experiment were identical to that of Experiment 1 except that in change trials, the changed color at the probe location would be a color that appeared elsewhere in the same sample display. In other words, the changed color would never be a color from the other sample display because recency effects would make such changes easy to detect regardless of whether one or two surfaces were present. Thus, observers needed to explicitly bind colors to their locations in the sample displays in order to successfully perform the change detection task. We also explicitly instructed observers to do so. The experiment lasted about 30 min.

Results and Discussion

As before, the means of A' for change detection were calculated. Performance was higher when objects appeared on two rather than on the same 3-D surface, $F(1, 15) = 8.99$, $p < .01$, Cohen's $d = 0.75$, indicating the presence of a between-surface benefit in VSTM when observers had to bind colors to locations. Results averaged over color change detection for the two sample displays are plotted in Figure 2A.

To ensure that the between-surface benefit in VSTM obtained in this experiment could not be accounted for by other factors, we repeated the three control experiments reported in Experiment 1 with the binding of colors and locations now explicitly required. Grouping by motion did not produce any VSTM benefit (A's for the moving and the stationary conditions were 0.74 and 0.76, respectively, $F(1, 11) < 1$, Cohen's $d = 0.25$, and would require testing approximately 133 observers to reach a .05 significance level with a power of 0.80). Grouping in 2-D space did not produce any VSTM benefit either (for the shorter interval, A's for the grouped and ungrouped conditions were 0.74 and 0.75, respectively, $F(1, 11) < 1$, Cohen's $d = 0.16$, and would require testing approximately 293 observers to reach a .05 significance level with a power of 0.80; and for the longer interval, A's for the grouped and the ungrouped conditions were 0.75 and 0.75, respectively, $F(1, 11) < 1$, Cohen's $d = 0.01$, and would require testing more than 1,000 observers to reach a .05 significance level with a power of 0.80). Similarly, the number of spatial locations attended to on the same surface did not affect performance, and the effect was slightly opposite to that predicted by the location hypothesis (A's for attending to 9 and 18 locations were 0.69 and 0.73, respectively, $F(1, 11) = 1.05$, $p > .33$, Cohen's $d = 0.30$, and would require testing approximately 92 observers to reach a .05 significance level with a power of 0.80).

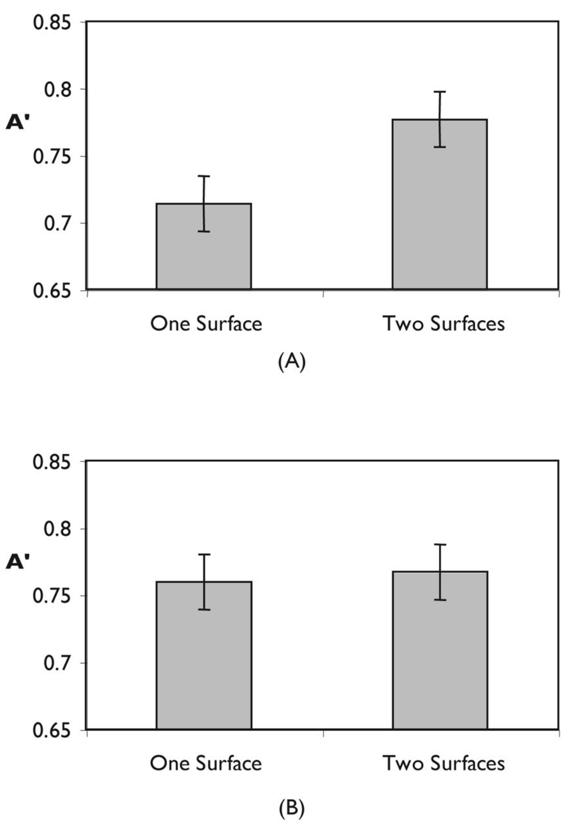

Figure 2. (A) Results of Experiment 2A. (B) Results of Experiment 2B. A between-surface benefit in visual short-term memory was present when observers had to explicitly remember the binding of colors to locations in the sample displays, as shown in Panel A. This effect was absent when binding was not required, as shown in Panel B. Error bars indicate 95% within-subject confidence intervals.

Experiment 2B: Separation by Surfaces Does Not Improve Unbound Colors in VSTM

In this experiment, we examined whether the between-surface benefit in VSTM was still present when the binding of colors and locations was not required. This was achieved by always presenting the probe color at the center of the test display (which was never occupied by a sample color) and by having the probe color be a color not present in the sample displays for the change trials. Although observers could still choose to bind colors to locations to perform the change detection task, they should have found it disadvantageous to do so because it is harder to retain color and location bindings in VSTM than to retain colors without location binding (Wheeler & Treisman, 2002).

Given that colors alone are easier to retain than the binding of colors and locations in VSTM (Wheeler & Treisman, 2002), in order to obtain comparable performance between this experiment and Experiment 2A, we asked observers to retain eight (instead of six) colors in VSTM in this experiment. This further encouraged observers to remember colors without binding.

Method

Participants. Twelve observers (8 women, 4 men) from the same participant pool and fulfilling the same criteria as before were recruited.

Materials, design, apparatus, and procedure. This experiment used the same design and displays as Experiment 2A except that each sample display contained four instead of three colors, resulting in a total of eight colors for the two sample displays. In the test display, the probe always appeared at the center of the display with zero disparity (i.e., where the fixation dot had been), and the probe location was never occupied by a color in the sample displays, even when only one surface was present. Observers were therefore instructed to remember which colors were present in the sample displays without binding colors to their locations. In addition to the eight colors used before, an orange color was added, making for a total of nine possible colors to choose from. For a no-change trial, the probe would be in one of the eight colors shown in the sample displays; and for a change trial, the probe would be in the ninth color not shown in the sample displays. Other aspects of the experiment were identical to those of Experiment 2A.

Results and Discussion

As before, the means of A' for change detection were calculated. There was no effect of the 3-D surfaces, and performance did not differ between objects on two different surfaces and objects on the same surface ($F < 1$, Cohen's $d = 0.09$, which would require testing approximately 906 observers to reach a .05 significance level with a power of 0.80). Thus, when the binding of colors and their locations was not required, there was no between-surface benefit in VSTM. Results averaged over color change detection of the two sample displays are plotted in Figure 2B.

The two-surface condition in Experiments 1 and 2B were similar in that, for change trials, the probe color in the test display was in a color not shown in the sample displays. However, the two experiments differed in that although the probe appeared at a location occupied by a sample color in Experiment 1 (which encouraged color and location binding), the probe always appeared at the center of the display in Experiment 2B (which discouraged such binding). The differences in results between these two experiments thus provided further support that the presence and absence of binding determined whether observers would use the structure of the 3-D space to organize the contents of their VSTM, such that a between-surface benefit in VSTM was observed in Experiment 1 but not in Experiment 2B.

Experiment 3: Between-Surface Versus Between-Depth Benefit in VSTM

In Experiments 1 and 2A, when objects were placed on the same planar surface perpendicular to the line of sight, they were also located at the same depth. Thus, the effect observed so far could be entirely due to the presence of multiple depths rather than to the presence of multiple 3-D surfaces per se. By presenting slanted surfaces expanding across multiple depths, Nakayama and colleagues (He & Nakayama, 1995; Nakayama, He, & Shimojo, 1995) showed, however, that it is the presence of different 3-D surfaces, rather than the presence of different depths, that influences a variety of visual perception and visual performance. To examine whether the presence of different depths or different surfaces determines the between-surface benefit that we observed in VSTM, we had observers in this experiment perceive and remember objects either from one or two vertical surfaces in depth (identical to those used in Experiment 2A; Figures 3A and 3B) or from one or two slanted surfaces across depth (Figures 3C and 3D). The slant of the surfaces relative to the line of sight was approximately 60°.

If the between-surface benefit in VSTM is determined by whether objects are placed on the same or on different surfaces, then we should observe this benefit regardless of whether surfaces are vertical or slanted across depth. On the other hand, if this benefit is determined by whether objects are placed at the same or at different depths, then when one slanted surface overlaps in depth with another slanted surface, we would expect to find little or no between-surface benefit in VSTM.

Method

Participants. Eight observers (4 women, 4 men) from the Yale University campus were recruited. They were between 17 and 40 years of age, had normal color vision and stereovision, and received course credits for their participation.

Materials, design, apparatus, and procedure. There were four display conditions. All objects appeared either on the same surface (Figures 3A and 3C) or on two different surfaces (Figures 3B and 3D); the surface was either vertical with the objects on a given surface at the same depth (Figures 3A and 3B) or slanted with the objects on a given surface at different depths (Figures 3C and 3D). The displays shown in Figures 3A and 3B were therefore replications of the two conditions in Experiment 2A. As before, the disparity difference between any two surfaces was 0.1°. The disparity difference between the top and the bottom of each slanted surface was 0.8°. The amount of depth overlap between the two slanted surfaces was about 65%. The perceived slant of the surfaces relative to the line of sight was about 60°. All other aspects of the experiment were identical to those of Experiment 2A. The experiment lasted about 55 min.

Results and Discussion

As in preceding experiments, the means of A' for change detection were calculated. There was an overall between-surface benefit in VSTM such that objects placed on two surfaces were better remembered in VSTM than were those placed on the same surface, $F(1, 7) = 9.01$, $p < .05$, Cohen's $d = 1.06$. Although there was a small difference between the slanted and the vertical surfaces, it was not significant, $F(1, 7) = 2.43$, $p > .16$, Cohen's $d = 0.55$, and would require testing approximately 28 observers to reach a .05 significance level with a power of 0.80. Neither did the interaction between the number of surfaces and surface orientation reach significance ($F < 1$, Cohen's $d = 0.21$), and it would require testing approximately 175 observers to reach a .05 significance level with a power of 0.80. In pairwise comparisons, the between-surface benefit was significant for both the vertical and the slanted surfaces, $F(1, 7) = 7.42$, $p < .05$, Cohen's $d = 0.96$, and $F(1, 7) = 10.06$, $p < .05$, Cohen's $d = 1.12$, respectively. Results averaged over color change detection of the two sample displays are plotted in Figure 3E. These results thus replicated those of Experiment 2A and indicated that the between-surface benefit seen in Experiments 1 and 2A was a true between-surface benefit and not a between-depth benefit.

General Discussion

VSTM plays an important role in visual cognition, and its limited capacity places restrictions on all behaviors dependent on it. Although objects are located on different 3-D surfaces in the real world, how the presence of multiple 3-D surfaces may influence VSTM capacity has never been examined in previous studies. By manipulating binocular disparities of visual displays, we found that more object information could be held in VSTM when objects were placed on two parallel 3-D surfaces instead of on the same 3-D surface. This result was replicated four times in the present study (once in Experiment 1, once in Experiment 2A, and twice in Experiment 3). This between-surface benefit in VSTM was present only when observers had to remember the binding of object features to locations. When such binding was not required, no between-surface benefit in VSTM was observed. This benefit in VSTM could not be attributed to the number of spatial locations attended to within a given surface. Neither was it due to a general perceptual grouping effect, because grouping by motion and grouping by different regions on the same surface (e.g., the left and the right sides of the 2-D space) did not yield the same benefit. These results suggest that separation by 3-D surfaces plays the determining role in generating the between-surface benefit in VSTM.

By definition, perceptual grouping includes any type of sensory grouping, including grouping by 3-D surfaces. What our results show is not whether perceptual grouping plays a role in VSTM (of course it does because grouping by 3-D surfaces is a type of perceptual grouping). But rather, these results highlight the role of grouping by 3-D surfaces in VSTM that is qualitatively different

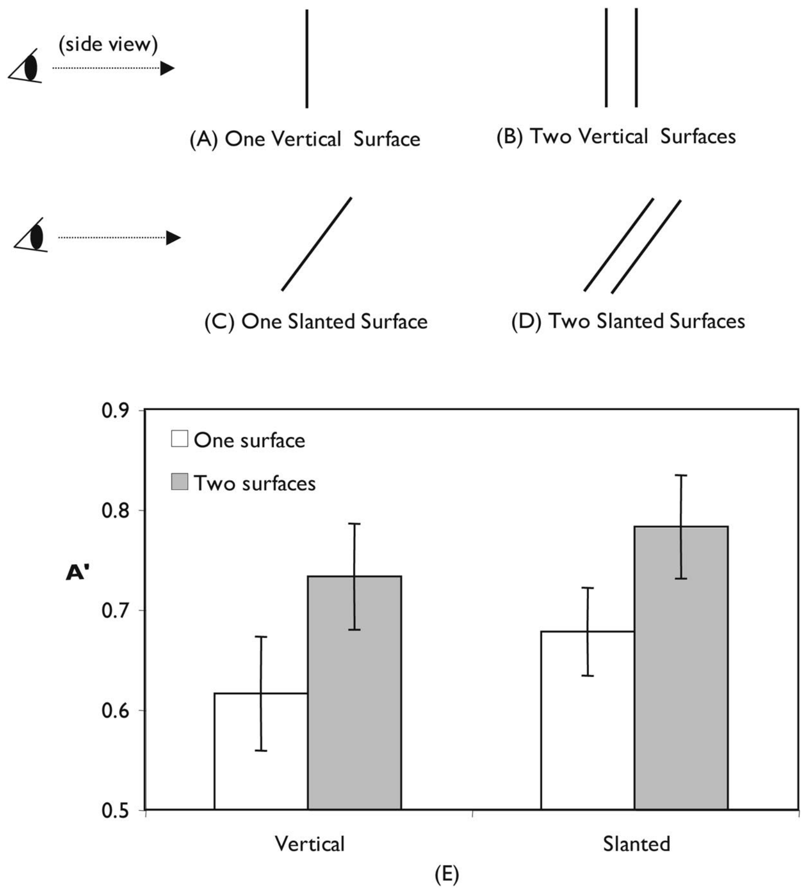

Figure 3. The displays used in Experiment 3. There were four display conditions. Objects appeared on either the same surface (A and C) or on two different surfaces (B and D). The surfaces could be either vertical with the objects on a given surface at the same depth (A and B) or slanted with the objects on a given surface at different depths (C and D). (E) Results of Experiment 3. Regardless of surface orientation, a between-surface benefit in visual short-term memory was obtained. Error bars indicate 95% within-subject confidence intervals.

from other types of perceptual grouping, such as grouping by motion or grouping by the left and the right sides of the 2-D space.

A number of studies during the last 3 decades have illustrated the importance of 3-D space on visual attention and perception (e.g., Downing & Pinker, 1985; Enns & Rensink, 1990; He & Nakayama, 1995; Kleffner & Ramachandran, 1992; Nakayama & Silverman, 1986; Viswanathan & Mingolla, 2002). Previous studies that have examined the role of 3-D surfaces in visual cognition, however, have used tasks demanding "on-line" processing of the stimuli, that is, all stimulus processing was carried out while the stimuli were visible (e.g., target detection: He & Nakayama, 1995; Marrara & Moore, 2000; visual search: Nakayama & Silverman, 1986; Theeuwes, Atchley, & Kramer, 1998; and multiple object tracking: Viswanathan & Mingolla, 2002). In our study, observers saw two sets of color objects sequentially, either on the same or on two different surfaces. This manipulation removed any bias in allocating attention to the colored objects and equated the initial perception of these objects regardless of whether one or two surfaces were present. With this manipulation, we still found that the presence of multiple 3-D surfaces has a long lasting effect on how object information is retained and stored in VSTM in the absence of visual stimulation. Thus, the presence of multiple 3-D surfaces affects not only perception but also short-term memory, indicating that our visual system can take advantage of the structure of the 3-D space and can use it to organize object representations in VSTM.

If we use Cowan's (2001) formula to estimate the number of objects held in VSTM, on the basis of the data from Experiment 2A, observers could hold about 1.91 objects when one surface was present and about 2.47 objects when two surfaces were present, roughly a 30% increase in capacity.[4] Thus, it is not the case that there is separate and independent VSTM storage for each surface, which would have doubled VSTM capacity when two surfaces were present. But rather, this is consistent with earlier studies showing that object representations within the same surface may be bundled together, similar to the "chunking" observed in verbal short-term memory (McLean & Gregg, 1967). As a result, there is more interference between object representations within than be-

[4] This VSTM capacity estimate was lower than the typical four items reported by others (e.g., Luck & Vogel, 1997; see also Cowan, 2001). It was likely due to the fact that the binding of color and location was explicitly required in Experiment 2A, which Wheeler and Treisman (2002) have also shown to be more attentionally demanding and to result in a lower VSTM capacity.

tween surfaces, resulting in more object information being clearly represented in VSTM across different 3-D surfaces. Because our studies only tested two parallel planar 3-D surfaces, further studies are needed to examine whether VSTM capacity can be further improved if more 3-D surfaces are present and if nonparallel surfaces are used. Nonetheless, given the significance of VSTM in visual cognition, our results indicate possible ways to improve VSTM capacity in situations in which a higher capacity could critically improve task performance.

Luck and Vogel (1997) argued that the total number of coherent object representations that could be simultaneously achieved by neural synchronization (Gray, König, Engel, & Singer, 1989) may determine VSTM capacity limitation. If this is indeed the case, then the present results suggest that surface separation may play an important role in determining how representations of different objects are achieved and maintained by neural synchronization.

Recent fMRI studies by Todd and Marois (2004) and by Xu and Chun (2006) have reported that activities in the posterior parietal area are correlated with the number of objects held in VSTM. In particular, Xu and Chun (2006) studied VSTM for object shapes and showed that a fixed number of objects (about three or four) are first individuated by the inferior intraparietal sulcus (IPS) via their spatial locations, and, depending on their complexity, a subset of the selected objects are then encoded and retained in great detail by the superior IPS and the lateral occipital complex. Coincidentally, brain imaging studies on humans (e.g., Gulyás & Roland, 1994; Shikata et al., 2001) and single neuron recording studies on monkeys (e.g., Taira, Tsutsui, Jiang, Yara, & Sakata, 2000; Tsutsui, Sakata, Naganuma, & Taira, 2002) have all shown the involvement of the parietal lobe in 3-D surface perception. For example, Tsutsui et al. (2002) reported that neurons in the monkey parietal region were selectively tuned to the orientation of the 3-D surface defined by either texture gradients or binocular disparity. These results thus suggest possible neural connections between surface perception and VSTM capacity limitation, consistent with the present behavioral findings.

At the moment, it is not clear when the between-surface benefit occurs in VSTM. This effect may occur at the initial object selection stage in the inferior IPS such that the interference among objects allows fewer objects to be selected within the same surface than it does between different surfaces. As a result, fewer objects are encoded and retained in VSTM when they share a surface. It is also possible, however, that the effect occurs during the encoding and maintenance of the detailed object information in the superior IPS such that objects from different surfaces may be retained with higher fidelity than those from the same surface as a result of less interference among objects located on different 3-D surfaces. Further studies are needed to understand the details of the between-surface benefit in VSTM.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15,* 106–111.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 168,* 637–643.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral Brain Sciences, 24,* 87–114.

Delvenne, J. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition, 96,* B79–B88.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In O. S. M. Marin & M. I. Posner (Eds.), *Attention and performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

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.

Enns, J. T., & Rensink, R. A. (1990, February 9). Scene-based properties influence visual search. *Science, 247,* 721–723.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989, March 23). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338,* 334–337.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Gulyás, B., & Roland, P. E. (1994). Binocular disparity discrimination in human cerebral cortex: Functional anatomy by positron emission tomography. *Proceedings of the National Academy of Sciences USA, 91,* 1239–1243.

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in 3-D space. *Proceedings of the National Academy of Sciences USA, 92,* 11155–11159.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal Experimental Psychology: Learning, Memory, and Cognition, 18,* 307–317.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 683–702.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52,* 18–36.

Kumar, A., & Jiang, Y. (2005). Visual short-term memory for sequential arrays. *Memory & Cognition, 33,* 488–499.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

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

McLean, R. S., & Gregg, L. W. (1967). Effects of induced chunking on temporal aspects of serial recitation. *Journal of Experimental Psychology, 74,* 455–459.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual cognition—An invitation to cognitive science* (2nd ed., pp. 1–70). Cambridge, MA: MIT Press.

Nakayama, K., & Silverman, G. H. (1986, March 20). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264–265.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290.

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology, 29,* 117–133.

Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125–126.

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

Shikata, E., Hamzei, F., Glauche, V., Knab, R., Dettmers, C., Weiller, C., & Büchel, C. (2001). Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: An event-related fMRI study. *Journal of Neurophysiology, 85,* 1309–1314.

Taira, M., Tsutsui, K., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of Neurophysiology, 83,* 3140–3146.

Theeuwes, J., Atchley, P., & Kramer, A. F. (1998). Attentional control within 3-D space. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1476–1485.

Todd, J. J., & Marois, R. (2004, April 15). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428,* 751–754.

Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002, October 11). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298,* 409–412.

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

Vogel, E. K., & Machizawa, M. G. (2004, April 15). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428,* 748–751.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Wheeler, M., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.

Xu, Y. (2000). Object parts in visual short-term memory and visual search. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Xu, Y. (2002a). Integrating color and shape in visual short-term memory for objects with parts. *Perception & Psychophysics, 64,* 1260–1280.

Xu, Y. (2002b). Limitations in object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 458–468.

Xu, Y. (2006). Encoding objects in visual short-term memory: The roles of location and connectedness. *Perception & Psychophysics, 68,* 815–828.

Xu, Y., & Chun, M. M. (2006, March 2). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440,* 91–95.

Appendix

Hit and False-Alarm Rates for Experiments 1–3

Condition	Hit rates	False-alarm rates
	Experiment 1	
One surface–squares	0.85	0.31
One surface–squares/circles	0.86	0.34
Two surfaces	0.91	0.29
	Experiment 2A	
One surface	0.62	0.30
Two surfaces	0.67	0.26
	Experiment 2B	
One surface	0.76	0.40
Two surfaces	0.72	0.34
	Experiment 3	
Vertical–one surface	0.56	0.39
Vertical–two surfaces	0.64	0.32
Slanted–one surface	0.60	0.31
Slanted–two surfaces	0.65	0.21

Note. The differences between the bias effects of the different conditions within each experiment were not significant.

Received October 10, 2006
Revision received May 7, 2007
Accepted May 10, 2007 ■