Perception & Psychophysics
2002, 64 (8), 1260-1280

Encoding color and shape from different parts of an object in visual short-term memory

YAODA XU
Massachusetts Institute of Technology, Cambridge, Massachusetts

Can we find an object-based encoding benefit in visual short-term memory (VSTM) when the features to be remembered are from different parts of an object? Using object parts defined by either figure–ground separation or negative minima of curvature, results from five experiments in which the visual change detection paradigm was used showed that the object-based encoding benefit in VSTM is modulated by how features are assigned to parts of an object: Features are best retained when the color and shape features to be remembered belong to the same part of an object. Although less well retained in comparison, features from different parts of an object are still better remembered than features from spatially separated objects. An object-based feature binding therefore exists even when the color and shape features to be remembered are from different parts of an object.

Visual short-term memory (VSTM) is defined as short-term memory for nonverbal, visual information, a buffer that temporarily stores visual information before it can be further processed. VSTM thus plays an important role in visual cognition. The aim of the present study was to address the question of how features of a two-part object are encoded in VSTM. Because VSTM and visual perception are tightly linked, studying the characteristics of feature binding in VSTM may provide us with insights as to how feature binding may work in general in perception.

The Encoding of Object Features in VSTM

The encoding of multiple features of an object in VSTM has been a topic of interest for the last 3 decades. Allport (1971) presented three colored shapes for 20 msec, followed by a blank screen and then, after 0, 20, or 40 msec, by a mask. Participants were asked to report the color, the shape, or both features of all three items. Allport found that reports of both color and shape were essentially as good as reports of color or shape alone. The same pattern of results was observed when participants were asked to report the orientation and spatial frequency of a single patch of lines (Wing & Allport, 1972): Reports of a single feature, either orientation or spatial frequency, were equally accurate, whether or not the other feature also had to be reported. These observations were further confirmed by Duncan (1984), who measured accuracy of reporting stimuli consisting of a rectangle (box) with a line crossing it diagonally and extending beyond it. Participants were asked to report the tilt and texture (or brightness) of the line, the tilt and brightness of the box, or the size of the box and the position of a gap in the same box. Results were obtained that were similar to those in Allport and in Wing and Allport, which led Duncan (1984) to conclude that "when focal attention is paid to an object, all features coded as properties of the whole (size, shape, color, etc.) are perceived without mutual interference" (p. 515).

In more recent studies, multiple objects were presented, and then a single object was cued for report. Irwin (1992) presented an array of 6 or 10 letters briefly. A location was then cued, and participants were asked to report the identity of the letter in that location. Irwin found that participants could correctly report only 3–4 letters, regardless of the number of letters presented in the original letter array. Using the same paradigm in a later study, Irwin and Andrews (1996) presented colored letters and asked participants to report both the identity and the color of a letter in a specified location. The correct report of both letter identity and the color was again restricted to 3–4 items. In other words, the report of color did not affect the report of letter identity. Irwin and Andrews concluded that "all of the information at a given spatial location in the array was conjoined via attention into a unitary object file for storage in transsaccadic memory" (p. 140; transsaccadic memory here is equivalent to VSTM).

Using the change detection paradigm of Phillips (1974; see also Pashler, 1988), Luck and Vogel (1997; see also Vogel, Woodman, & Luck, 2001) presented participants with several colored oriented bars in a sample display for

This research is based on parts of a doctoral dissertation submitted to the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, under the mentorship of Mary C. Potter. This research was supported by Grant MH 47432 from the National Institute of Mental Health to M. C. Potter and by a McDonnell-Pew Investigator Initiated Grant in Cognitive Neuroscience to Y.X. Some of the research reported here was presented at the 7th Annual Workshop on Object Perception, Attention and Memory, Los Angeles, November 1999. I am grateful to Mary C. Potter, for her valuable advice, comments, and encouragement throughout the project. I also thank Jeremy Wolfe, Frank Durgin, and two anonymous reviewers for comments on earlier versions of this paper. Correspondence concerning this article should be addressed to Y. Xu, Vision Sciences Laboratory, Psychology Department, Harvard University, 33 Kirkland St., Room 744, Cambridge, MA 02138 (e-mail: yaoda@wjh.harvard.edu).

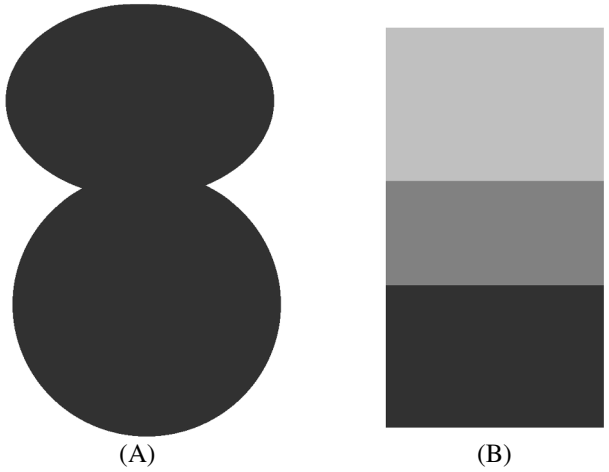


Figure 1. Examples of parsing according to negative minima of curvature (A) and according to surface color discontinuity (B).

100 msec. The display was then replaced by a 900-msec blank interval followed by a test display. The test display was identical to the sample display, except that, in 50% of the trials, 1 feature of one item would change. The participants were asked to detect this change. In one condition, only color was monitored for a possible color change (orientation never changed), and the authors estimated that the colors of about four items could be correctly encoded in VSTM. The same estimate was obtained when only orientation was monitored. In a third condition, both color and orientation features were monitored for a possible change in either feature. It was again found that both the color and the orientation of about four items could be correctly encoded in VSTM. Like Irwin and Andrews (1996), Luck and Vogel argued that this was evidence that integrated objects are encoded in VSTM in such a way that when 1 feature of an item is encoded, other features of the same item are also automatically encoded; however, only four objects can be encoded in VSTM at a time. To extend their finding, Luck and Vogel conducted another experiment in which each object had 4 features (color, orientation, size, and the presence or absence of a gap) and any one of the features could change after the delay. Performance was just as good in this experiment, leading to the conclusion that "16 features distributed across 4 objects can be retained as accurately as 4 features distributed across 4 objects" (Luck & Vogel, 1997, p. 280).

Together, with simple objects, such as colored letters (Irwin & Andrews, 1996) or colored bars (Luck & Vogel, 1997), all the studies described above found that multiple features of an object can be encoded in VSTM as accurately as just one feature of an object.

Objects and Parts

In contrast to the simple objects used in the experiments above, most objects we encounter in our everyday experience consist of multiple parts, each with its own features. How is an object with multiple parts encoded in VSTM? Is the object benefit limited to encoding features from the same part of an object? Or can it also be observed when the to-be-remembered features come from different parts of an object?

Before these questions can be addressed, a definition of *part* is needed. For three-dimensional shapes and two-dimensional shapes (silhouette) with uniform surface properties (Figure 1A), Hoffman and Richards (1984) noted that whenever two independent parts connect or interpenetrate to form a complete object, or when a part grows out of a body, the boundaries between these parts typically lie in the negative minima of curvature. Hoffman and Richards thus formulated the *minima rule*, which argues that human vision parses shapes into parts by using negative minima of curvature as boundaries between parts. More recently, the minima rule has been extended to a rule for joining boundaries to create part cuts (Singh, Seyraninan, & Hoffman, 1999) and for determining the salience of parts (Hoffman & Singh, 1997). A number of psychological studies have provided converging evidence suggesting that the human vision system indeed parses shape according to these rules (e.g., Baylis & Driver, 1994, 1995; Hoffman & Singh, 1997; Hulleman, te Winkel, & Boselie, 2000; Wolfe & Bennett, 1997; Xu & Singh, 2002).

There is, however, a different way of locating or defining parts within an object, without the need to compute negative minima of curvature. Our visual system can also use discontinuities in surface property, such as color or texture, and figure–ground separation to locate and define parts (Figure 1B). We encounter object parts defined in this manner frequently in our everyday experience. For example, there is no negative minimum of curvature present in the front view of an eye, and yet when presented in a face, we see the eye as a part of the face. Under the same principle, the eye itself can be further divided into parts, such as sclera, iris, and pupil.

Some ambiguities, however, exist in how parts should be defined. For example, because the eye has a distinct closed boundary of its own, instead of seeing the eye as part of a face, it can also be viewed as an independent object that just happens to be inserted into the eye socket. As such, when the eye is attached to the face, it becomes uncertain whether it is part of an object, or an independent object that is inserted into a larger object. As Marr (1982) once commented, "What . . . is an object, and what makes it so special that it should be recoverable as a region in an image? Is a nose an object? Is a head one? Is it still one if it is attached to a body? What about a man on horseback? . . . All these things can be an object if you want to think of them that way, or they can be a part of a larger object" (p. 270).

Despite the subjectiveness of what counts as a part or an object, however, there are some rules—for example, Gestalt principles—that govern how we perceive and organize our perceptual world. For example, according to the rule of proximity, in Figure 2, no matter how you want to see each dot as an individual object, you cannot

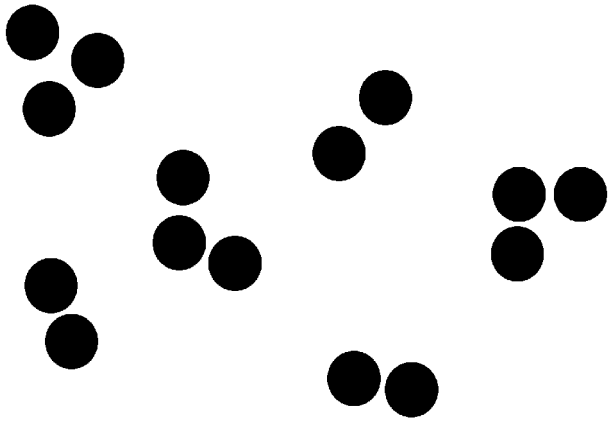


Figure 2. An example of grouping by proximity. Here, the dots form clusters, with each cluster containing either two or three dots.

help but notice that there is some structure, some hierarchy to these dots—namely, that these dots form clusters, with each cluster containing either two or three dots. Surely, one can still think of each dot as an independent/individual object, but treating each dot as part of a cluster better describes the emerging structure among these dots. *Part*, I would like to argue in this context, is a term that describes lower and intermediate level structures that make up a hierarchy, and *object* is thus a term that describes the entire hierarchy. By these definitions, depending on where our focus is on the structure/hierarchy, a structure can be an object in one context and a part in another. For example, when we focus on a person, the person becomes an object, and his/her head becomes an object part; however, when we focus on the head, the head becomes an object, and the face becomes an object part; and further, when we focus on the face, the face becomes an object, and the eyes become object parts; and so on and so forth. As such, when Marr (1982) argued that it really depends on how you want to think of it whether something is an object or a part of a larger object, what is changed is probably not your thoughts, but your focus and the level of the hierarchy that you are attending to.

In the present study, two simple geometric shapes, the parts, are attached to form a hierarchy, the object, as is shown in Figures 3A–3D; when these geometric shapes are spatially separated from each other and, thus, do not form a hierarchy, they are considered as independent objects, as is shown in Figures 3E–3G. The goal of the present study was to understand how features from different parts of an object are encoded in VSTM, using the change detection paradigm (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974).

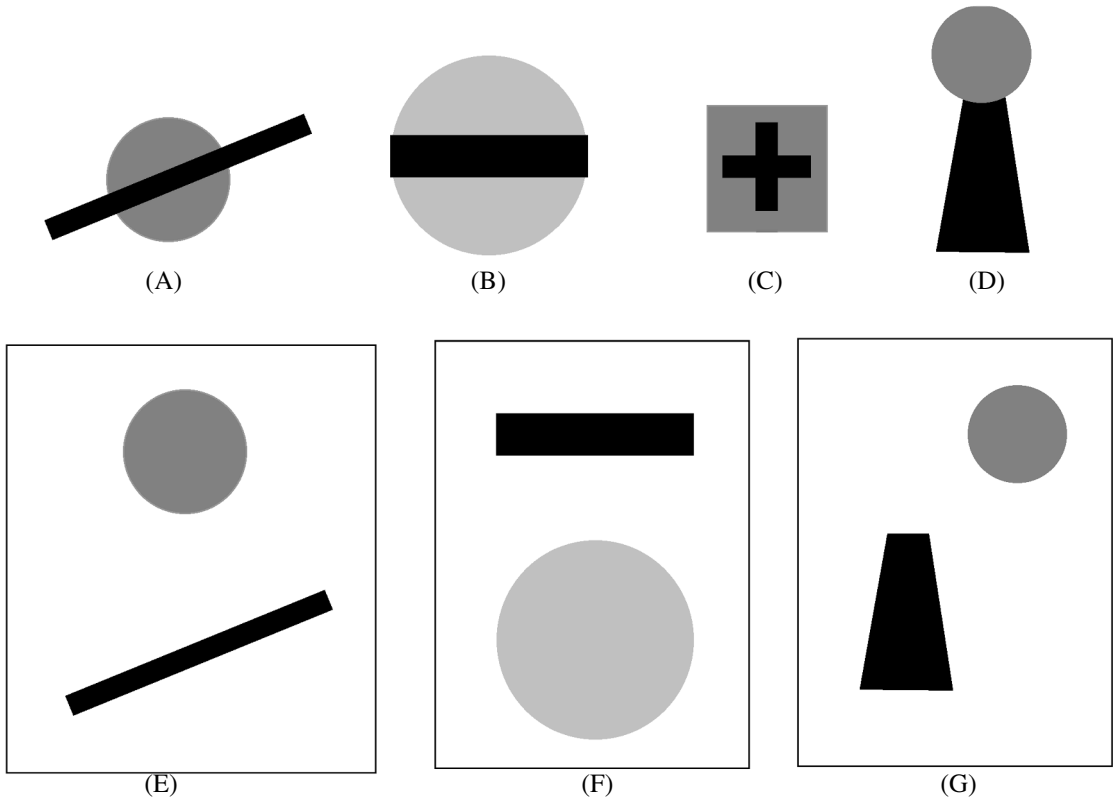


Figure 3. Panels A, B, C, and D show examples of single two-part objects; panels E, F, and G each contain two independent objects.

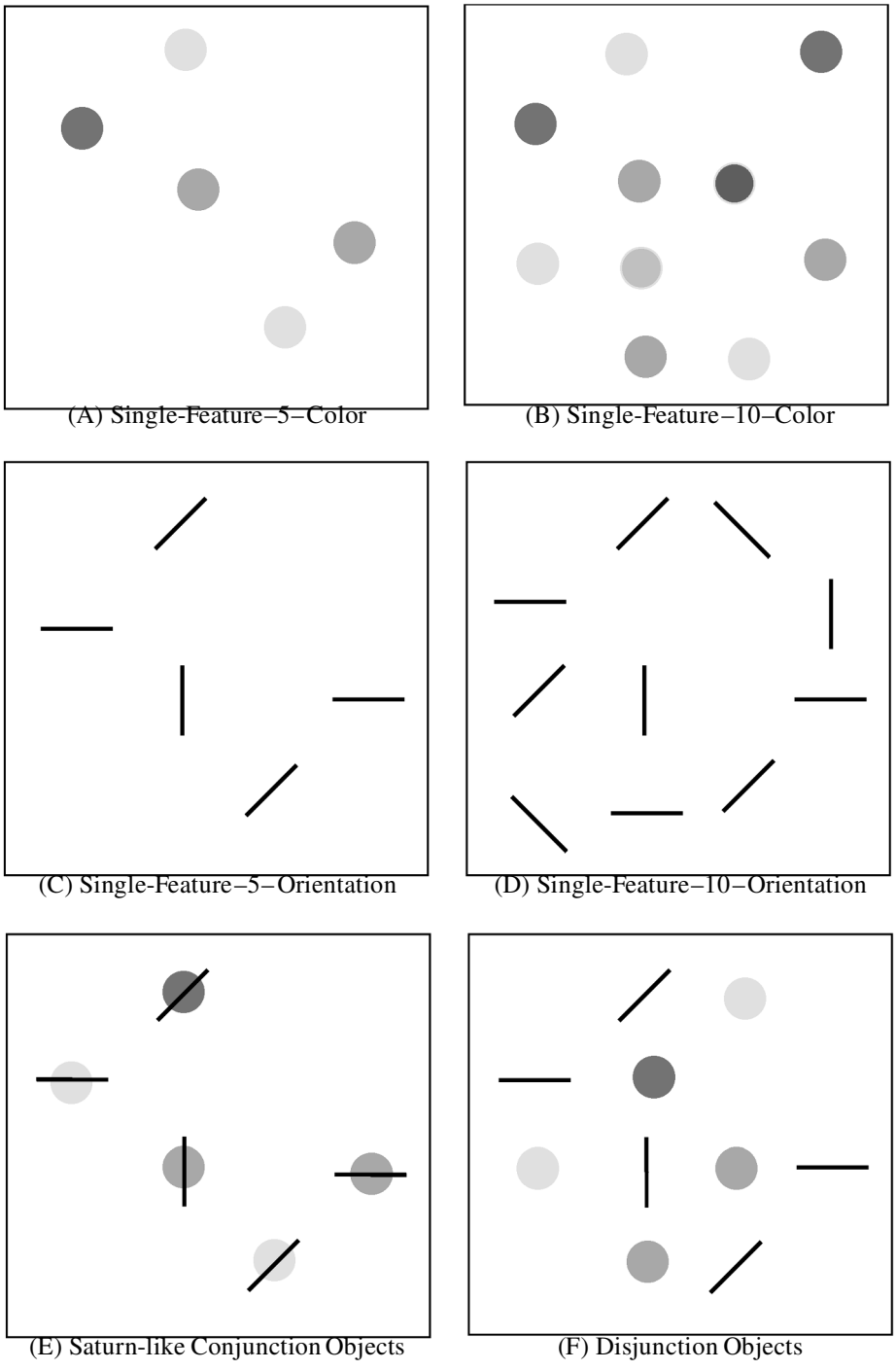


Figure 4. Examples of the stimuli used in Experiment 1. The stimuli shown in panels A and B contained 5 and 10 color features, respectively; those shown in panels C and D contained 5 and 10 orientation features, respectively; and those shown in panels E and F each contained 5 color features and 5 orientation features.

EXPERIMENT 1

To study the encoding of color and orientation from different object parts, Saturn-like objects (Figure 4E) were constructed. In these objects, the relevant color feature was carried by the color of the "planet," and the relevant orientation feature was carried by the orientation of the "ring" (a black bar). On each trial, the participants viewed an array of five such objects, followed, after 1 sec, by a second display. A change in one of the relevant fea-

tures of one of the objects would occur in 50% of the trials. The participants were asked to detect this change. If both relevant features of a Saturn-like object can be encoded as accurately as just one feature, accuracy in color change or orientation change detection should be comparable to displays containing only 5 colored circles (Figure 4A) or 5 oriented bars (Figure 4C). To control for the total number of features present in the Saturn-like object displays, there were also control displays containing one of the following: (1) 5 oriented bars and 5 colored circles that were spatially separated (Figure 4F), (2) 10 colored circles (Figure 4B), or (3) 10 oriented bars (Figure 4D).

Method

Participants. Ten volunteers from the Massachusetts Institute of Technology campus, 4 females and 6 males, were recruited. They were between 17 and 40 years of age, all had normal color vision, and they were paid for their participation.

Materials and Design. The stimuli used were black oriented bars (1.7° × 0.1°) and colored circles (diameter, 0.9°) on a light gray background. Each of the Saturn-like objects was constructed from a black oriented bar and a colored circle. Four orientations (vertical, horizontal, −45°, and +45°) and four colors (cyan, yellow, white, and pink) were used. The whole display extended 10.5° × 10.5°. The objects in a given display were separated by at least 2.6° × 2.6° (center to center).

Six display types were used. First, there were four homogenous control displays: *single-feature–5–color* and *single-feature–10–color* displays, which contained 5 and 10 colored circles, respectively, with the colors of the circles being the only relevant features (Figures 4A and 4B), and *single-feature–5–orientation* and *single-feature–10–orientation* displays, which contained 5 and 10 oriented bars, respectively, with the orientations of the bars being the only relevant features (Figures 4C and 4D). The *Saturn-like conjunction objects* display contained five Saturn-like objects (Figure 4E). In these objects, the relevant color feature was carried by the color of the "planet," and the relevant orientation feature was carried by the orientation of the "ring" (a black bar). Thus, the relevant features were on different parts of an object. The *disjunction objects* display contained 5 black oriented bars and 5 colored circles spatially separated from each other (Figure 4F). Here, the relevant color features (colors of the circles) and orientation features (orientations of the black bars) were located on separate objects. The six display types were randomly intermixed during the experiment. For displays that contained homogenous objects (black bars or colored circles), there was a total of 40 trials for each display type, with 20 change trials and 20 no-change trials. For displays that contained conjunction or disjunction objects, there was a total of 80 trials for each display type, with 20 orientation-ch ange trials, 20 color-change trials, and 40 no-change trials. There was a total of 320 experimental trials, evenly divided into four 80-trial blocks. Each participant had 32 practice trials before proceeding to the experimental trials.

Only three of the four prespecified colors and orientations were used (chosen randomly) in any given display. The orientation of the black bars and the color of the circles were then assigned randomly, with replacement from these three prechosen colors and orientations. If a change occurred, the changed feature value was randomly selected from the two remaining prechosen feature values. This sampling procedure ensured that, most of the time, the same feature values appeared in both the sample and the test displays: That is, it was rarely the case that the change introduced a color (or orientation) that had not already been present in the sample display. The objects appeared in any of 16 possible positions, in an implicit 4 × 4 matrix, with the following constraints. The 16 positions were divided into four quadrants, each containing a 2 × 2 array. The objects were distributed over the four quadrants as follows: For set size 5, 2 random positions were selected from one randomly chosen quadrant, and 1 random position was selected from each of the remaining three quadrants; for set size 10, 3 random positions were selected from each of two randomly chosen quadrants, and 2 random positions were selected from each of the remaining two quadrants. This sampling procedure ensured that object locations were evenly distributed for any given display and that the 5- and 10-object displays occupied similar envelopes.

Apparatus. The displays were generated by a Power Macintosh 7500/100 computer and MacProbe Macintosh programming software. A 17-in. AppleVision 1710 Display monitor was used to display the stimuli.

Procedure. The participants were seated in a dim-lit and quiet room, about 50 cm from the screen. They initiated each trial by pressing the space bar on the computer keyboard. Each trial began with a "+" at the center of the screen for 507 msec, followed by the sample display for 253 msec. The sample display was then replaced by the blank gray background. After 1,000 msec, the test display appeared. The test display remained on the screen until the participant made a keypress. The participants were instructed to press the key marked "Same" (key "K") when they did not detect any change and to press the key marked "Different" (key "G") when they detected a change. Feedback on response accuracy (as "Correct" or "Wrong") and feedback on the number of trials completed in a block were given after each trial. When ready, the participant pressed the space bar to start the next trial. The whole experiment lasted about 45 min. The participants were told that response accuracy was more important than response speed and that only response accuracy was analyzed in the result.

Results

Xu and Potter (2000; see also Xu, 2000) have provided evidence that feature information is represented in VSTM in a graded, rather than an all-or-nothing, manner. Therefore, only models that assume graded information representation in a detection task are appropriate for characterizing VSTM change detection performance. Among these models is the signal detection theory, which generated the d' measure and inspired the A' measure (Grier, 1971; Pollack & Norman, 1964). A' is a measure of sensitivity that estimates the area under the *receiver operating characteristic* curve. It increases from .5 for chance performance to 1.0 for perfect performance (see Macmillan & Creelman, 1991, for more detailed information regarding A'). In the present analyses, A' was used instead of d' because, in general, with yes/no paradigms, A' is more accurate in characterizing performance than is d' (Donaldson, 1993) and A' does not have the indeterminacy of d' when a participant makes no false yeses (false alarms). A' was calculated for each participant in each condition, following the formula by Grier (1971):

$$A' = .5 + (H - g) \cdot (1 + H - g)/[4 \cdot H \cdot (1 - g)],$$

where H is the hit rate and g is the guessing or false alarm rate. If guessing rate was greater than the hit rate, the following formula was used (Aaronson & Watts, 1987; Snodgrass & Corwin, 1988):

$$A' = .5 - (g - H) \cdot (1 + g - H)/[4 \cdot g \cdot (1 - H)].$$

When the participants had to monitor both features and detect a change in either color or orientation, separate false alarm rates for color and orientation changes could not be measured; as a result, A' scores could not be computed separately for color- and orientation-change detections. Performance was therefore averaged over color- and orientation-change detections before the A' score was computed for each participant in each condition. The final means of A' are plotted in Figure 5. No object-based benefit was observed for encoding the two relevant features of the Saturn-like objects in VSTM, as compared with control conditions.

Overall, there was a significant effect of display type [$F(3,27) = 21.67$, $p < .001$]. Change detection performance was much worse for the Saturn-like conjunction objects in which both relevant features of each object had to be monitored than for displays containing five colored circles or five oriented bars (the *single-feature–5* condition), in which only one feature of each object had to be monitored [$F(1,9) = 53.72$, $p < .001$]. This result suggests that VSTM cannot encode both relevant features as accurately as just one relevant feature of a Saturn-like object. The Saturn-like object displays were also compared with the disjunction object displays that contained the same number of color and orientation features on spatially separated objects. Although the mean for the former was higher than that for the latter displays, this difference was not significant [$F(1,9) = 2.50$, $p > .10$]. When all three types of displays containing10 relevant features were compared (i.e., the Saturn-like object display, the disjunction display, and the display containing 10 single features), no significant difference was found [$F(2,18) = 1.32$, $p > .25$].

Discussion

In the present experiment, the question of how features from different parts of an object are encoded in VSTM was addressed, using the Saturn-like objects in which the relevant color and orientation features were located on different parts of the object. By using the change detection paradigm, it was found that when both relevant features of the Saturn-like object had to be monitored, the amount of feature information encoded in VSTM was significantly less than that for displays containing only five colored circles or only five oriented bars. Therefore, unlike the finding for colored shapes reported by Allport (1971), colored letters reported by Irwin and Andrews (1996), and colored oriented bars reported by Luck and Vogel (1997), the 2 features of the Saturn-like objects were not as readily encoded in VSTM as 1 feature of an equivalent object. In fact, accuracy of change detection for the Saturn-like objects did not differ from that for the disjunction displays containing the same number of features located on separated objects or for displays containing10 single features (10 colors or 10 orientations). In other words, for the Saturn-like objects, although color and orientation were still features of a single object, there was little or no object-based benefit.

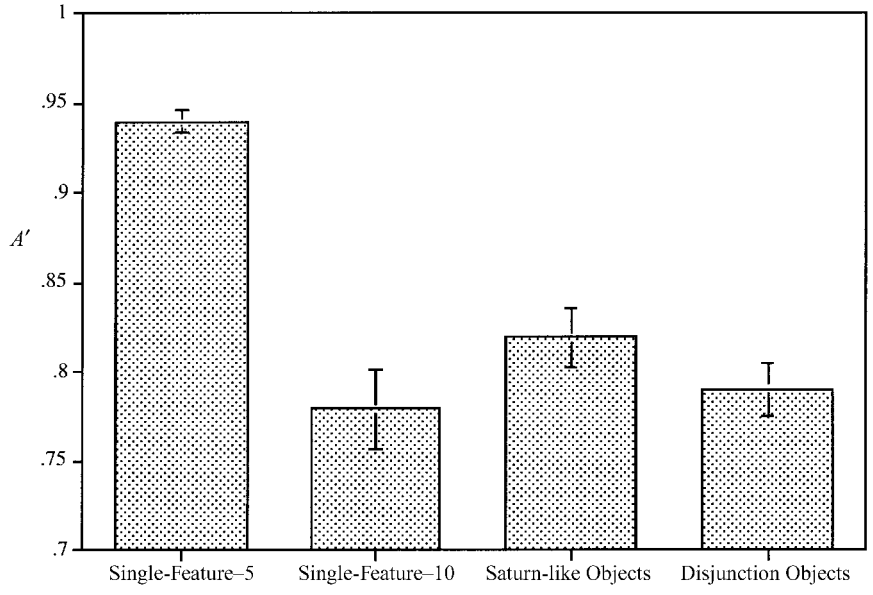


Figure 5. Results of Experiment 1. The two relevant features of a Saturn-like object were not as readily encoded in visual short-term memory (VSTM) as one feature of an equivalent object (the *single-feature–5* condition). In fact, the results for the Saturn-like objects did not differ from those for the disjunction displays, which contained the same number of features but were spatially separated, or those for displays that contained 10 colored circles or10 oriented bars (the *single-feature–10* condition). These results indicated that for the Saturn-like objects, although color and orientation were still features of a single object, there was little or no object-based encoding benefit in VSTM.

Lee and Chun (2001) have also reported this last result. These authors used stimuli similar to those used by Duncan (1984) and presented three boxes and three lines, either superimposed, forming three box–line pairs occupying three locations, or as six spatially separated objects occupying six locations. The same change detection paradigm as that in the present study was used, and participants had to remember the features of the boxes (size and location of a gap) and the lines (direction of tilt and texture) after an 800-msec delay. Lee and Chun found that performance was not affected by whether the lines and the boxes were superimposed or separated.

There are, however, two possible objections to the present conclusions. One objection is that because the different display types were randomly intermixed and, in most trials, the black bars and the colored circles were separate objects (rather than parts of the same object), the participants might have been discouraged from integrating the features of the Saturn-like objects and, instead, might have encoded each part of the Saturn-like object as two separate objects, with one lying on top of the other. If the participants indeed perceived the color and orientation features as belonging to two independent objects, it is not surprising that there was no object-based benefit.

A second objection to the present conclusion is that a Saturn-like object was visually more complex than a simple oriented bar or a colored circle. Thus, the comparison between the Saturn-like object display and the other displays might have been biased against the Saturn-like objects, canceling a possible object-based advantage for the Saturn-like objects.

EXPERIMENT 2

To address the first objection raised above, in Experiment 2, instead of randomly intermixing the different display types, trials were blocked so that, within a block, only a particular display type was used (this objection was further addressed in Experiment 3). The second objection, that objects with multiple parts are more visually complex than objects without multiple parts, was addressed in Experiment 2 by using Luck and Vogel's (1997) procedure. The participants were instructed to attend either to one feature of the object or to both features. If the encoding of one feature of the Saturn-like objects was not affected by whether or not another feature of the same object was also encoded, performance for monitoring one feature should be comparable to that for monitoring two features.

Method

Participants. Eighteen volunteers, 10 females and 8 males, from the same participant pool as that in Experiment 1 were recruited.

Materials and Design. The stimuli used included black oriented bars (1.7° × 0.1°), colored circles (diameter, 1.0°), and colored oriented bars (1.7° × 0.3°). The same four orientations and four colors as those in Experiment 1 were used here. The whole display extended 8.7° × 8.7°, with the objects in a given display separated by at least 2.2° × 2.2° (center to center).

There were three display types, which were (1) *colored oriented bars*, which contained six colored oriented bars, with the relevant color and orientation features contained in the same part of an object (Figure 6A), (2) *Saturn-like conjunction objects*, which contained six Saturn-like objects, with the relevant color features on the circles and the relevant orientation features on the bar (Figure 6B), and (3) *disjunction objects*, which contained six black bars and six colored circles spatially separated from each other and, as in (2), with the relevant color features on the circles and the relevant orientation features on the bars (Figure 6C). There were three feature monitoring conditions: (1) Only color was monitored, and a color change would occur in 50% of the trials; (2) only orientation was monitored, and an orientation change would occur in 50% of the trials; and (3) both color and orientation were monitored, and a change in either color or orientation (but never both) would occur in 50% of the trials. Trials were blocked by display type and by monitoring condition; the order of the blocks was balanced across participants. For each display type in which a single feature was monitored, there were 80 trials, with 40 change trials and 40 no-change trials, evenly distributed into two 40-trial blocks; for each display type in which both relevant features were monitored, there were 160 trials, with 40 color-change trials, 40 orientation-change trials, and 80 no-change trials, evenly distributed into four 40-trial blocks. The participants completed half of the trials of each type in one testing session and came back on a different day to complete the rest of the trials. Five practice trials preceded the experimental trials in each condition during the experiment.

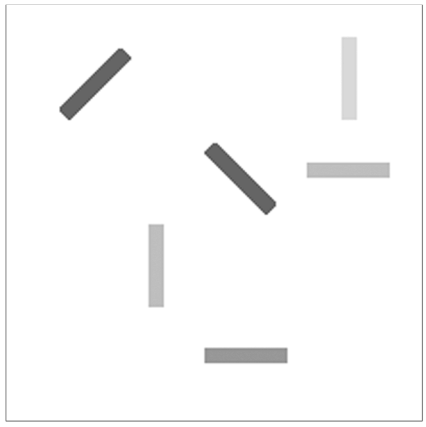
Color and orientation features used in each sample display were randomly chosen by the computer from the four prespecified colors and orientations, with the constraint that a given color or orientation could not appear more than twice. If a change occurred, the changed feature value was randomly selected from the three remaining feature values. This sampling procedure ensured that, most of the time, the same feature values appeared in both the sample and the test displays. Other aspects of the design were identical to those of Experiment 1.

Procedure. The procedure was the same as that in Experiment 1, except that (1) the sample display was on for 200 msec, instead of 253 msec, to reduce the possibility of eye movements during the viewing of the sample display and (2) the participants were asked to press the left control key with their left index fingers if they detected a change and to press the enter key on the number keypad with their right index fingers if they did not detect any changes. Feedback was given as either a happy face for a correct response or a sad face for an incorrect response. The feedback stayed on the screen for 333 msec. The whole experiment lasted about 90 min, with each of the two testing sessions lasting about 45 min.

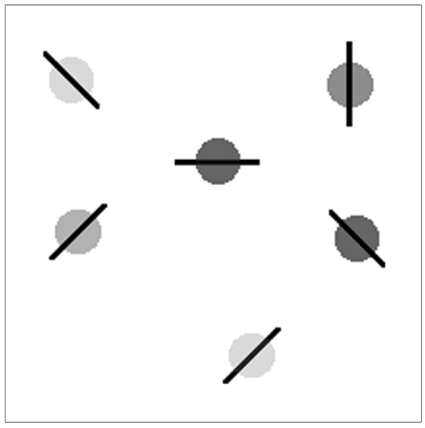
Results

As in Experiment 1, when the participants had to monitor for both features and detect a change in either color or orientation, separate A' scores could not be computed for color- and orientation-change detections. Performance was therefore averaged over color- and orientation-change detections before the A' score was computed for each participant in each condition. The final means of A' are shown in Table 1 and are plotted in Figure 7. The results indicated that color and orientation were most successfully encoded in VSTM when they were from the same part of an object, less well encoded when they were from different parts of an object, and least well encoded when they were distributed on spatially separated objects.

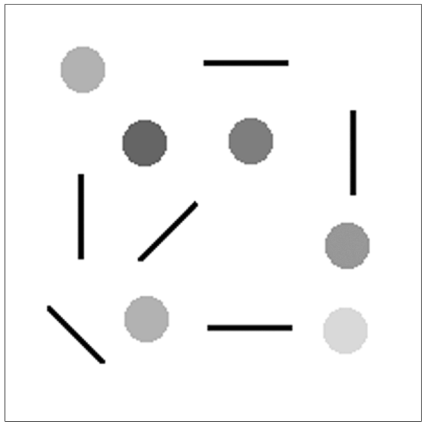
Overall, there were significant effects of display type [$F(2,34) = 25.13$, $p < .001$] and monitoring condition, so



(A) Colored Oriented Bars



(B) Saturn-like Conjunction Objects



(C) Disjunction Objects

Figure 6. Examples of the stimuli used in Experiment 2. In (A), the relevant color and orientation features were located on the same part of an object (the colored oriented bar); in (B), the relevant features were located on different parts of an object (color on the circle and orientation on the bar); and in (C), the relevant features were located on separate objects (color on the circle and orientation on the bar, with the circle and bar spatially separated from each other). During the experiment, the participants monitored either one of the two relevant features or both features.

that monitoring for one relevant feature was easier than monitoring for both relevant features [$F(1,17) = 98.63$, $p < .001$], and the two effects interacted with each other significantly [$F(2,34) = 13.02$, $p < .001$]. Even for displays containing the colored oriented bars, monitoring for one feature was easier than monitoring for both [$F(1,17) = 16.88$, $p < .01$].[1]

When the Saturn-like objects and the colored oriented bars were compared, there was a significant interaction between display type and monitoring condition [$F(1,17) =$

Table 1
Experiments 2, 3, 4, and 5: Mean A' in Each Display Condition for Monitoring One Relevant Feature and Monitoring Two Relevant Features, and the Difference Between the Two Monitoring Conditions

Display Condition	One Feature		Two Features		Difference
	M	*SE*	*M*	*SE*	
Experiment 2					
Colored oriented bars	.88	.01	.85	.01	.03
Saturn-like objects	.83	.01	.76	.01	.07
Disjunction objects	.87	.01	.76	.01	.11
Experiment 3					
Colored oriented bars	.88	.01	.84	.01	.04
Black beachball-like objects	.81	.01	.76	.01	.05
Colored beachball-like objects	.81	.01	.73	.01	.08
Disjunction objects	.87	.01	.74	.01	.13
Experiment 4					
Colored symbols on black squares	.82	.02	.76	.02	.06
Black symbols on colored squares	.84	.01	.75	.01	.09
Experiment 5					
Conjunction display with colored stems	.84	.02	.78	.01	.06
Conjunction display with black stems	.86	.02	.76	.01	.10
Disjunction display	.87	.01	.71	.02	.16

6.80, $p < .05$], indicating a bigger drop in performance for the Saturn-like objects for monitoring two features versus one feature (see Figure 7). When the Saturn-like objects and the disjunction objects were compared, there was also a significant interaction between display type and monitoring condition [$F(1,17) = 4.54$, $p < .05$], but this time it was the disjunction objects that had a bigger drop in performance for monitoring two features versus one feature (see Figure 7).

Discussion

Recall that in Experiment 1, with the total number of features held constant, performance for the Saturn-like objects did not differ from that for the disjunction ob-

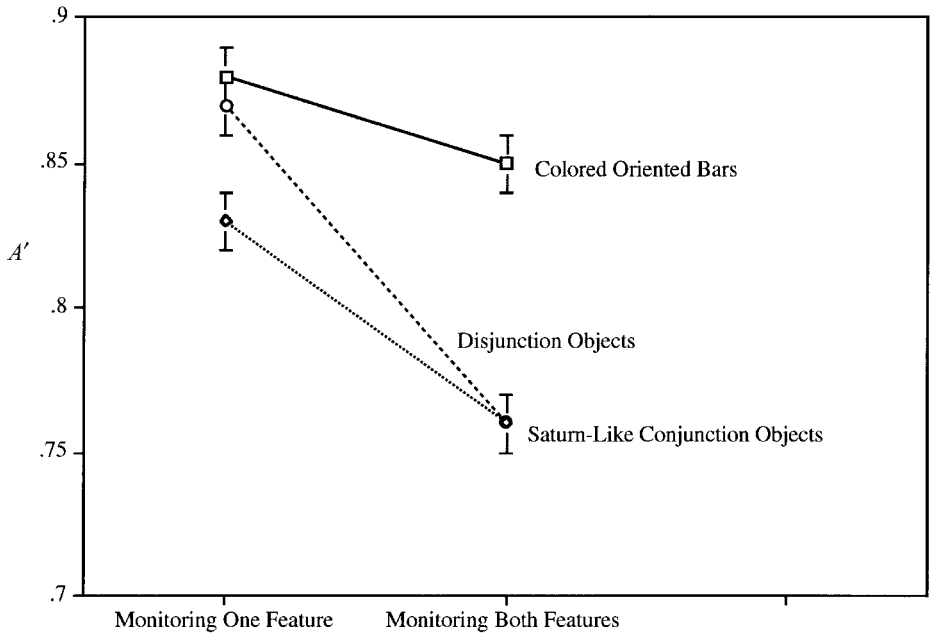


Figure 7. Results of Experiment 2. The amount of performance drop for monitoring two versus one relevant feature was smallest for the colored oriented bars, greater for the Saturn-like conjunction objects, and greatest for the disjunction objects. These results indicate that color and orientation were most successfully encoded in VSTM when they were from the same part of an object, less well encoded when they were from different parts of an object, and least well encoded when they were distributed on spatially separated objects.

jects. This result was replicated in the condition in which both features were monitored in the present experiment: The mean A's for the Saturn-like and the disjunction objects, averaged over color- and orientation-change detection, were the same (.76). However, when only one feature was monitored, performance was lower for the Saturn-like objects than it was for the other two types of objects [$F(2,34) = 12.34$, $p < .001$]. This is most likely due to higher visual complexity associated with the Saturn-like objects,[2] confirming one of the objections raised in Experiment 1. In the present experiment, the differences in visual complexity were taken into account by measuring the differences between monitoring one versus both features. By this measure, it was found that two features were easier to encode for the Saturn-like objects than for the disjunction objects.

In the study reported by Lee and Chun (2001) described earlier, differences in visual complexity for the box–line pairs and for the separated boxes and lines were not accounted for. Given the present findings, it is very likely that displays containing the box–line pairs were better encoded in VSTM than were displays containing spatially separated boxes and lines but that this advantage was canceled by the added difficulties in encoding, owing to higher visual complexity.

If how well color and orientation can be encoded in VSTM is a reflection of how well these two features can be integrated, the present results suggest that (1) the highest feature integration/binding can be achieved when color and orientation belong to the same part of an object, (2) when these features belong to different parts of an object, significantly less feature integration/binding is achieved, and (3) there is, however, still a substantial amount of feature integration/binding in (2), as compared with cases in which color and orientation are located on spatially separated objects.

There remain, however, two objections to these conclusions. Although trials from different display types were blocked in the present experiment, it is still unclear whether the Saturn-like objects were actually perceived as single objects or as two separate objects, superimposed. In the disjunction object blocks, the two features were, in fact, separate objects, which might have influenced the perception of the Saturn-like objects. In addition, the ends of the black bar in the Saturn-like object extended outside the contour of the colored circle; thus, the ensemble strongly resembled two distinct objects (one lying on top of the other), rather than a single object. If VSTM is object based—that is, if VSTM encoding is limited by the number of objects, and not by the total number of features—and the Saturn-like objects are seen as two objects, it is not surprising that colored oriented bars (unambiguously single objects) were encoded more accurately than the Saturn-like objects.

A different objection to the present conclusion would abandon the claim that VSTM is object based and would state, instead, that it is measured by the number of features in the display. Each of the Saturn-like objects contained two shape features and two color features, and although only one of the shapes and one of the colors were relevant to the change detection task, the irrelevant features may have nonetheless competed for encoding in VSTM. Thus, displays containing the Saturn-like objects would be encoded less well than displays containing the colored oriented bars, each of which had only one color and one shape feature.

EXPERIMENT 3

In Experiment 3, objects similar to the Saturn-like objects were used, but with the ends of the black oriented bar following the contour of the colored circle; the conjunction resembled a colored beachball with a black stripe (Figure 8C). Here, figure–ground interpretation assigns the orientation contour as belonging to the black bar, rather than as belonging to the two colored semicircles. The generic view principle (Rock, 1983; see also Nakayama & Shimojo, 1992) asserted that when there is more than one surface interpretation of an image, the visual system assumes that it is viewing the image from a generic, not an accidental, vantage point. According to this principle, when the ends of the black oriented bar exactly followed the contour of the colored circle, it would be much more likely for the visual system to interpret the ensemble as one object and to interpret the black bar as part of a textured surface on the object (i.e., a black stripe running across a colored beachball) than to interpret the ensemble as two distinct objects with one lying on top of the other, only accidentally in perfect juxtaposition. In addition, given that all the other circle–bar pairs presented in the display were also in perfect alignment, the interpretation of each circle–bar pair as a colored beachball with a black stripe—a single two-part object—should be even stronger. Although the generic view principle does not completely define objecthood, it does dictate how we interpret the visual world to a great extent.

To address the second objection raised above, a display matched for visual complexity of the beachball-like objects was also included. In this control display, both color and orientation information was presented in the beachball stripe; the rest of the beachball was black (Figure 8B). In other words, both the color and the orientation were now contained in the same part of the object (the colored stripe), and the color and shape of the black beachball itself were irrelevant to the change detection task. If VSTM encodes all the color and shape features present in a display, the encoding of the colored beachballs with a black stripe should be comparable to that of the black beachballs with a colored stripe. On the other hand, if the most efficient encoding of color and orientation in VSTM can be achieved when features belong to the same part of an object, the encoding of the black beachball-like objects with a colored stripe should be significantly better than that of the colored beachball-like objects with a black stripe.

As in Experiment 2, colored oriented bars and disjunction objects were also included as controls. If features are indeed encoded in VSTM in a hierarchical manner, so that features are best encoded if they are from the same part of

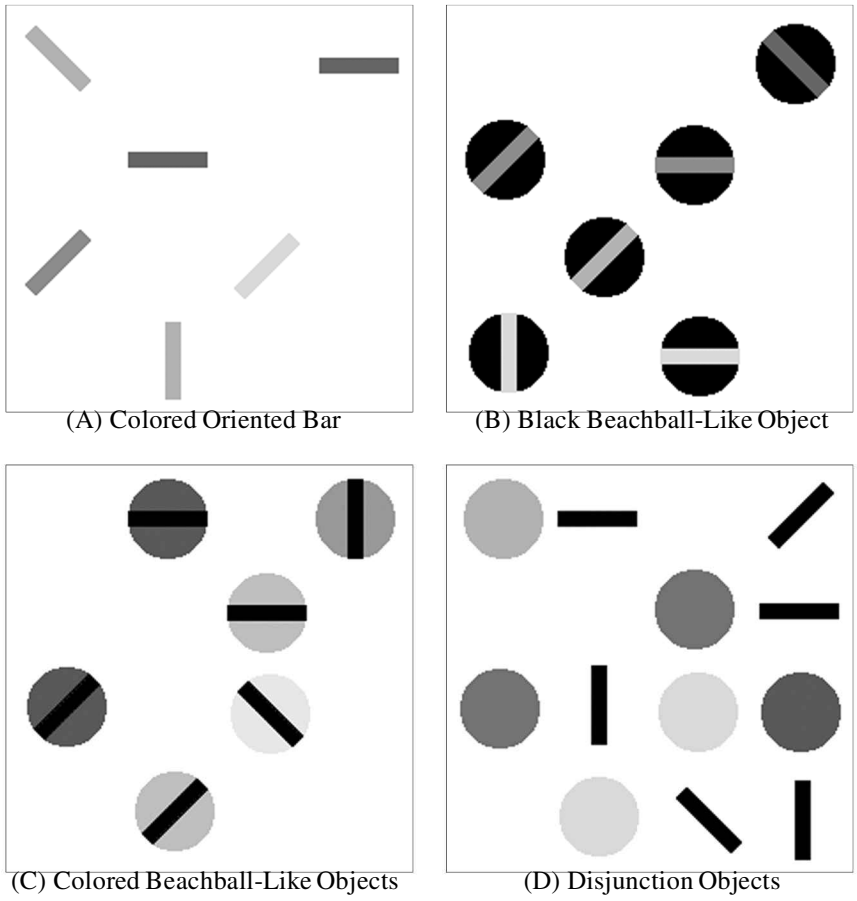


Figure 8. Examples of the stimuli used in Experiment 3. In (A) and (B), the relevant color and orientation features were located on the same part of an object (the colored orientated bar); in (C), the relevant features were located on different parts of an object (color on the circle and orientation on the bar); and in (D), the relevant features were located on separate objects (color on the circle and orientation on the bar, with the circle and bar spatially separated from each other). According to the generic view principle, each of the colored beachball-like objects in (C) should be better perceived as a single two-part object than was the Saturn-like object used in Experiment 2. Note also that both (B) and (C) contained the same number of total features, thus matching for overall visual complexity.

an object, less well encoded if they are from different parts of the same object, and least well encoded if they are from different objects, the prediction for the present experiment would be that performance drop for monitoring two features versus one feature should be the smallest for the colored oriented bars and the black beachball-like objects, greater for the colored beachball-like objects, and greatest for the disjunction objects. The task in the present experiment was identical to that in Experiment 2: The participants were asked to monitor just one or both relevant features of each display and to detect a change after a 1-sec delay.

Method

Participants. Twenty-four volunteers, 12 females and 12 males, from the same participant pool as that in the previous experiment, were recruited.

Materials and Design. The stimuli used included black and colored oriented bars (1.7° × 0.3°) and black and colored circles (diameter, 1.7°), presented separately or juxtaposed. A total of three kinds of color–orientation conjunction objects were used in the present experiment: (1) *colored oriented bars* (Figure 8A), (2) *colored beachball-like objects* in which the edges of the black bar followed the contour of the colored circle, so that the conjunction resembled a beach ball with a black stripe (Figure 8C); and (3) *black beachball-like objects*, each of which consisted of a colored oriented bar running across a black circle, resembling a black beachball with a colored stripe (Figure 8B), thus matching in visual complexity with (2). In both (1) and (2), the relevant color and orientation features were located on the same part of an object (the colored oriented bar), whereas in (3), the relevant features were on different parts of an object (color on the circle and orientation on the bar). In a fourth condition, the *disjunction objects*, the relevant color and orientation features were located on separated objects, with orientations on the black bars and colors on the circles and the bars and

circles spatially separated from each other (Figure 8D). The same four orientations and four colors as those used in Experiments 1 and 2 were used here. The whole display extended 8.7° × 8.7°, with the objects in a given display separated by at least 2.2° × 2.2° (center to center).

In each of the three conjunction conditions, there were 6 objects in each display; in the disjunction condition, there were 12 objects in each display (six oriented black bars and six colored circles). As in Experiment 2, there were three feature-monitoring conditions, one for the relevant color, one for the relevant orientation, and one for both relevant features. Trials were blocked by display type and by monitoring condition; the order of the blocks was balanced across participants. For each display type in which a single feature was monitored, there were 64 trials, with 32 change trials and 32 no-change trials, evenly distributed into two 32-trial blocks; and for each display type in which both relevant features were monitored, there were 128 trials, with 32 color-change trials, 32 orientation-change trials, and 64 no-change trials, evenly distributed into four 32-trial blocks. The participants completed half of the trials of each type in one testing session and came back on a different day to complete the rest of the trials. Five practice trials preceded the experimental trials in each condition during the experiment. Each of the two testing sessions lasted about 50 min.

In Experiment 2, it was found that color-change detection was easier than orientation-change detection. In an effort to make orientation change more salient, when an orientation changed, it was always a 90° change—that is, from 0° to 90° and vice versa or from 45° to 135° and vice versa. All other aspects of the design and procedure were identical to those in Experiment 2.

Results

As in Experiment 2, performance was averaged over color- and orientation-change detections before an A' score was computed for each participant in each condition. The final means of A' are shown in Table 1 and are plotted in Figure 9. The prediction for the present experiment was confirmed: The drop in performance for monitoring two features versus one feature (i.e., the interaction between display type and monitoring condition) was smallest for the colored oriented bars and the black beachball-like objects, greater for the colored beachball-like objects, and greatest for the disjunction objects.

Overall, as in Experiment 2, there were effects of display type [$F(3,69) = 27.95$, $p < .001$] and monitoring condition [$F(1,23) = 67.96$, $p < .001$], and the two interacted with each other significantly [$F(3,69) = 17.80$, $p < .001$]: The differences in performance for monitoring one versus both features were bigger for the colored beachball-like objects and the disjunction objects, but it was still significant for the colored oriented bars [$F(1,23) = 31.10$, $p < .001$].

When the colored bars and the black beachball-like objects (both had features located on the same part of the object) were compared, the interaction between display type and monitoring condition was not significant ($F < 1$), indicating that the drop in change detection performance for monitoring both features versus one relevant feature did not differ between these two conditions. The interaction of display type and monitoring condition, however, reached significance when the colored beachball-like objects were compared with the colored oriented bars [$F(1,23) = 6.61$, $p < .05$] and when the colored beach-

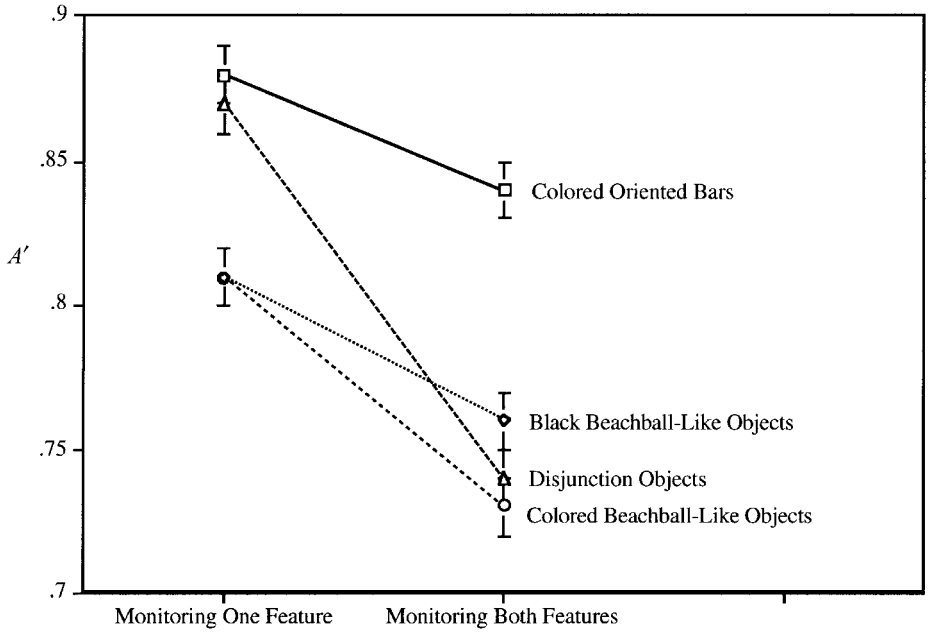


Figure 9. Results of Experiment 3. As in Experiment 2, the amount of performance drop for monitoring two versus one relevant feature was smallest for the colored oriented bars and the black beachball-like objects, greater for the colored beachball-like objects, and greatest for the disjunction objects. These results thus replicated the finding of Experiment 2 and showed again that features were most successfully encoded in visual short-term memory when they were from the same part of an object, less well encoded when they were from different parts of an object, and least well encoded when they were distributed on spatially separated objects.

ball-like objects were compared with the black beachball-like objects [$F(1,23) = 5.58, p < .05$], indicating that it was harder to encode the two relevant features (as compared with just one relevant feature) of the colored beachball-like objects than it was for the other two conditions. This drop in performance, or the interaction between display type and monitoring condition, was even greater for the disjunction objects than for the colored beachball-like objects [$F(1,23) = 10.10, p < .01$].

Discussion

The magnitude of the effect observed in the present experiment was quite comparable to that observed in Experiment 2 (see Table 1), although each of the colored beachball-like objects was better perceived as a single two-part object (according to the generic view principle) than was the Saturn-like object used in Experiment 2. Thus, the difficulty in encoding the relevant features of the Saturn-like objects versus the colored bars in Experiment 2 must not have occurred because the Saturn-like objects were treated as two separate objects, with one lying on top of the other; rather, it reflected some genuine difficulties in encoding two features from different parts of an object, as compared with features from the same part of an object.

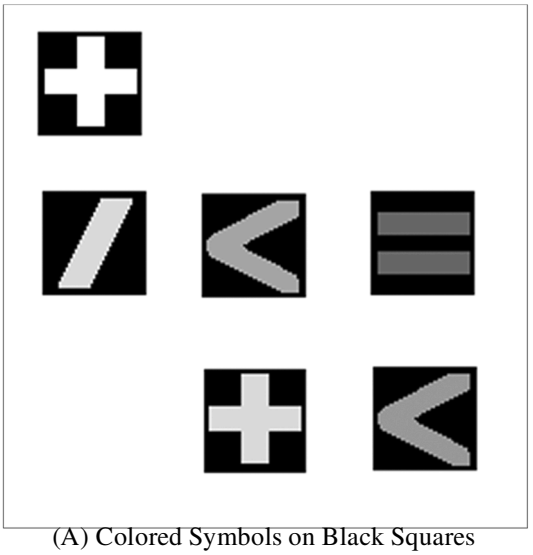
In the present experiment, when the features of a colored beachball-like object were rearranged so that the relevant color and orientation features were contained within the same part of the object (the stripe) and the object was turned into a black beachball-like object with a color stripe, monitoring for both relevant features of the object became easier. Note that each of these black beachball-like objects still contained the same number of shape and color features as the colored beachball-like objects, thus controlling for object complexity. And yet, depending on whether the two relevant features were contained in the same part or different parts of an object, the amount of information encoded in VSTM differed significantly.

In summary, the present results replicated those of Experiment 2 and showed again that two features were best encoded in VSTM when they were from the same part of an object, less well when they were from different parts of an object, and least well when they were located on spatially separated objects.

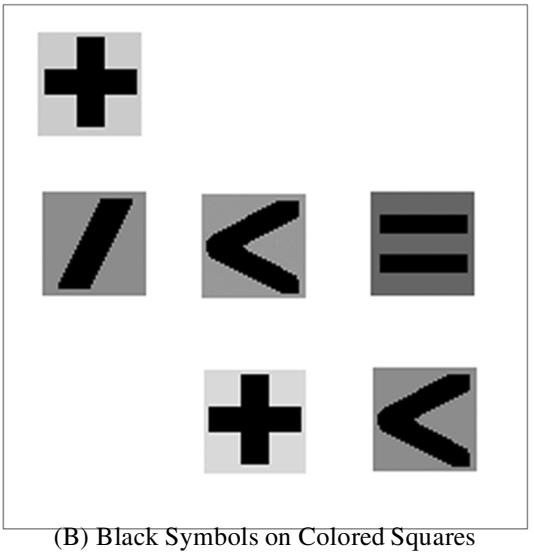
EXPERIMENT 4

The purpose of this experiment was to verify whether the difference observed in feature encoding between the black beachball-like objects and the colored beachball-like objects in Experiment 3 would be replicable with a different set of stimuli. To do so, shapes (math symbols +, = , /, and <) replaced orientations as one of the two features in this experiment. There were two kinds of displays; one contained colored symbols on black squares (Figure 10A, equivalent to the black beachball-like objects in Experiment 3), and the other contained black symbols on colored squares (Figure 10B, equivalent to the colored beachball-like objects in Experiment 3).

One might argue that these objects are not as good as single two-part objects as the beachball-like objects used in Experiment 3, because one can easily think of the symbol as detachable from the square, forming two independent objects. If change detection performance is indeed influenced by subjective interpretation of the display, the difference between the two displays used in the



(A) Colored Symbols on Black Squares



(B) Black Symbols on Colored Squares

Figure 10. Examples of the stimuli used in Experiment 4. In (A), the relevant color and shape features were located on the same part of an object (the colored symbol), and in (B), the relevant features were located on different parts of an object (color on the square and shape on the symbol). This was a different set of stimuli from that used in the previous two experiments and was used here in an attempt to replicate the findings of Experiments 2 and 3.

present experiment for monitoring one versus both features ought to be bigger than those observed in Experiment 3 for the black beachball-like objects and the colored beachball-like objects. On the other hand, if objects and parts are defined by where in the hierarchy we focus our attention, the symbol–square unit used in the present experiment was as good as the bar–circle unit used in Experiment 3. As such, the use of symbols and squares in the present experiment should not change the magnitude of the effect. There are examples in nature that resemble the stimuli used in the present experiment, such as the spots on a leopard, the black dots on the back of a lady bug, and the freckles on a face, that we naturally think of as a part of a leopard, a lady bug, and a face, respectively.

Method

Participants. Twenty-four volunteers, 12 females and 12 males, from the same participant pool as that in the previous experiments were recruited.

Material and Design. The stimuli used were colored math symbols (maximum extent, 1.5° × 1.5°) on black squares (1.7° × 1.7°) and black math symbols on colored squares, as is shown in Figure 10. The symbols used were +, = , /, and < . The colors used were red, cyan, white, and yellow. The whole display extended 8.7° × 8.7°, with the objects in a given display separated by at least 2.2° × 2.2° (center to center).

There were two display types: (1) *colored symbols on black squares*, which contained six colored math symbols on black squares, with the relevant color and shape features contained in the same part of an object (Figure 10A), and (2) *black symbols on colored squares*, which contained six black math symbols on colored squares, with the relevant color and shape features contained in different parts of the object (Figure 10B). These two display types were presented in separate blocks. For each display, color and shape features were chosen randomly, with the constraint that each of the four feature values appeared at least once, but no more than twice, in both the sample and the test displays. As in Experiments 2 and 3, there were three monitoring conditions: monitor for color, shape, or both color and shape. The experiment lasted about 50 min. Other aspects of the design were identical to those in Experiment 3.

Results

As in Experiments 2 and 3, performance was averaged over color and shape change detections before A' was computed for each participant in each condition. The final means of A' are shown in Table 1 and are plotted in Figure 11. The results replicated the findings of Experiment 3, so that the performance drop between the two display conditions for monitoring both features versus one feature was quite comparable to that for the two types of beachball-like objects used in Experiment 3 (see Table 1).

There was no overall difference between the two display types ($F < 1$). Monitoring for one relevant feature was significantly easier than monitoring for both relevant features [$F(1,23) = 64.39$, $p < .001$, for the overall effect; $F(1,23) = 76.62$, $p < .001$, for black symbols on colored squares; and $F(1,23) = 20.86$, $p < .001$, for colored symbols on black squares]. Display types interacted significantly with monitoring conditions [$F(1,23) = 5.53$, $p < .05$], so that the drop in performance for monitoring two versus one relevant feature was greater for black symbols on colored squares than it was for colored symbols on black squares.

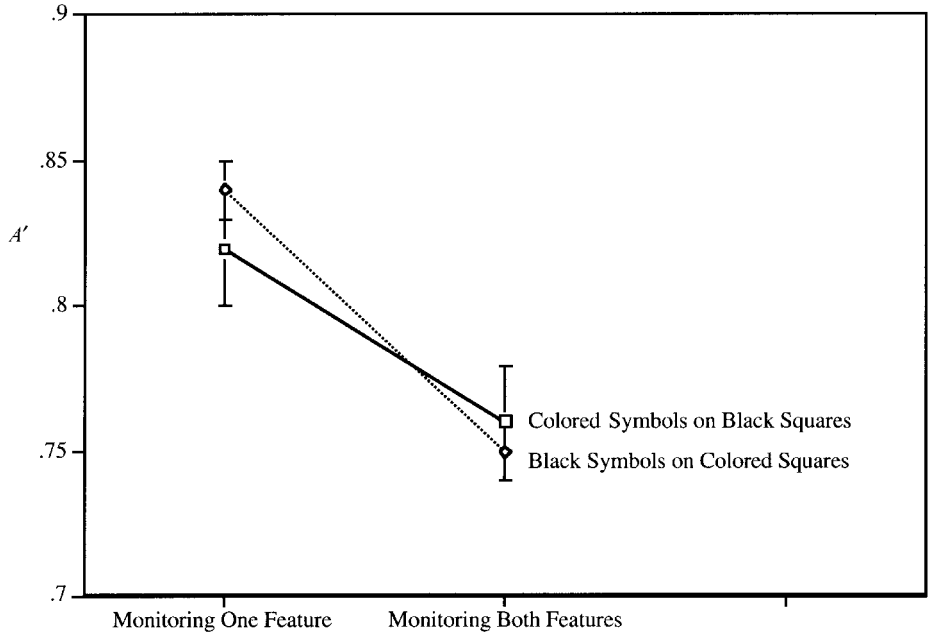


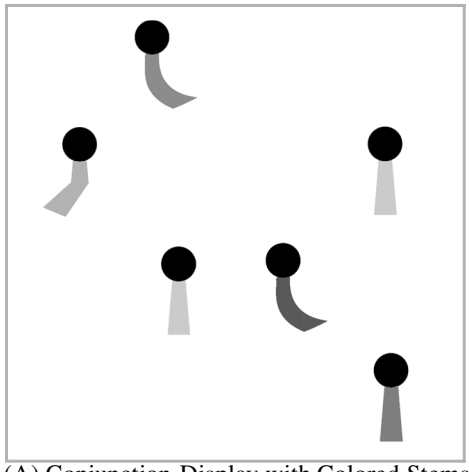
Figure 11. Results of Experiment 4. As in Experiments 2 and 3, the amount of performance drop for monitoring two versus one relevant feature was greater for the black symbols on colored squares when the relevant features were located on different parts of object than for the colored symbols on black squares when the relevant features were located on the same part of an object.

Discussion

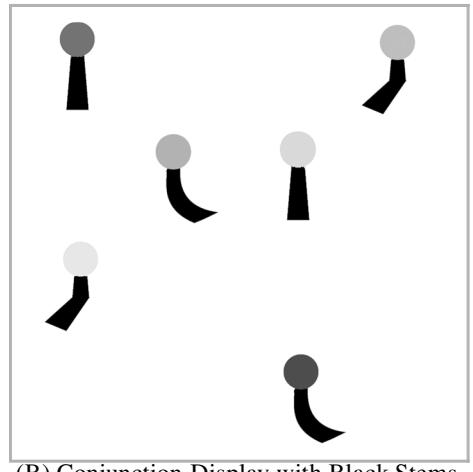
The present results thus replicated the findings of Experiments 2 and 3 with a different set of stimuli and showed that two features from different parts of an object were encoded less well in VSTM than were two features from the same part of an object. In addition, the magnitude of the effect did not seem to be influenced by the subjective impression of how strongly the two-part object could be perceived as a single object.

EXPERIMENT 5

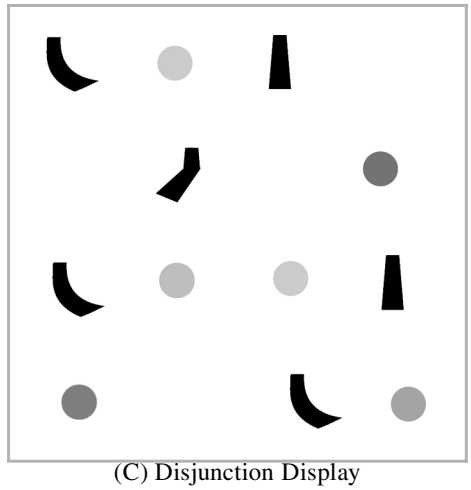
So far, all the effects observed in the present study were obtained from objects whose parts were defined by figure–ground separation. What about objects whose parts are defined by negative minima of curvature? In this final experiment, objects whose parts were defined by negative minima of curvature were used. Each object contained a circle joining a stem at negative minima of curvature (see Figure 12). The relevant color and orientation features could (1) both be located on the stem, so that the features were on the same part of the object (Figure 12A), (2) have the color located on the circle and the orientation located on the stem, with the two parts attached, so that features were on different parts of the same object (Figure 12B), or (3) be the same as in (2), but with the circle and the stem detached from each other, so that the features were on separate objects (Figure 12C). If the same hierarchical object-based feature encoding in VSTM observed with parts defined by figure–ground separa-



(A) Conjunction Display with Colored Stems



(B) Conjunction Display with Black Stems



(C) Disjunction Display

Figure 12. Examples of the stimuli used in Experiment 5. Here, object parts defined by negative minima of curvature were used. In (A), the relevant color and orientation features were located on the same part of an object (the stem); in (B), the relevant features were located on different parts of an object (color on the circle and orientation on the stem); and in (C), the relevant features were located on separate objects (color on the circle and orientation on the stem, with the circle and the stem spatially separated from each other).

tion applies to parts defined by negative minima of curvature, we should expect (1) to be better encoded than (2) and (2) better than (3).

Method

Participants. Eighteen volunteers, 12 females and 6 males, from the Harvard University campus were recruited. They fulfilled the same selection criteria as the participants in the previous four experiments.

Material and Design. The stimuli used are shown in Figure 12. The diameter of each color circle was 1.0°, and the maximum extent of the black shape was 1.3° × 1.6°. The whole display extended 10.3° × 12.0°, with the objects in a given display separated by at least 2.6° (center to center). The circles were presented in one of four colors, as in Experiments 1–3: pink, cyan, white, or yellow. The black shapes were presented in one of three orientations (right, center, or left; the right shape was slightly different from the left shape so that an orientation change could be more easily detected).

There were three display types: (1) *conjunction display with colored stems*, in which the relevant color and orientation features were contained in the same part of an object (Figure 12A), (2) *conjunction display with black stems*, in which the relevant color and orientation features were contained in different parts of the same object (colors carried by the circles and orientations carried by the stems; Figure 12B), and (3) *disjunction display*, in which the relevant color and orientation features were on spatially separated objects, which was the same as (2), but with the circles and the stems detached from each other (Figure 12C). As in Experiments 2–4, there were three feature-monitoring conditions: colors only, orientations only, and both colors and orientations together.

To generate the sample display, the computer randomly assigned the color and the orientation to each object so that a given color would not appear more than twice and a given orientation would not appear more than three times in a given display. When a change occurred, the changed color or orientation value was randomly selected. As in the previous experiments, trials were blocked by display type and monitoring condition. When a single feature was monitored, there were 64 trials, with 32 change trials and 32 no-change trials, distributed evenly into two 32-trial blocks; when both features were monitored, there were 128 trials, with 32 color change trials, 32 orientation change trials, and 64 no-change trials, distributed evenly into four 32-trial blocks. The experiment lasted about 50 min. The displays were generated by an iMac with a 350-MHz Power PC G3 processor and a 15-in. monitor. Other aspects of the design were identical to those of Experiments 2–4.

Results

As in Experiments 2–4, performance was averaged over color- and orientation-change detections before A' was computed for each participant in each condition. The final means of A' are shown in Table 1 and are plotted in Figure 13. The results replicated the findings of Experiments 2–4, which had parts defined by figure–ground separation, and showed that the drop in performance was smallest when color and orientation were located on the same part of the object, greater when these features were on different parts of the same object, and greatest when these features were on spatially separated objects.

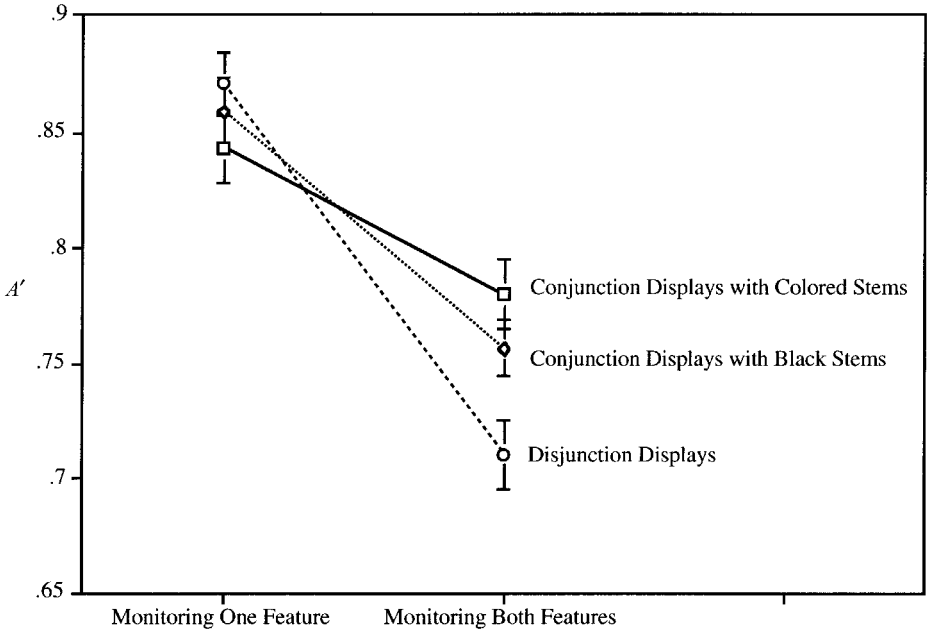
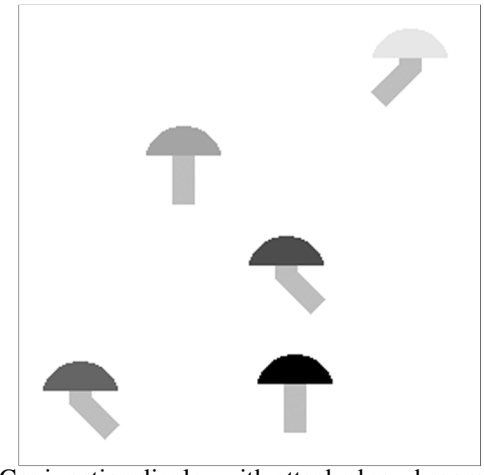
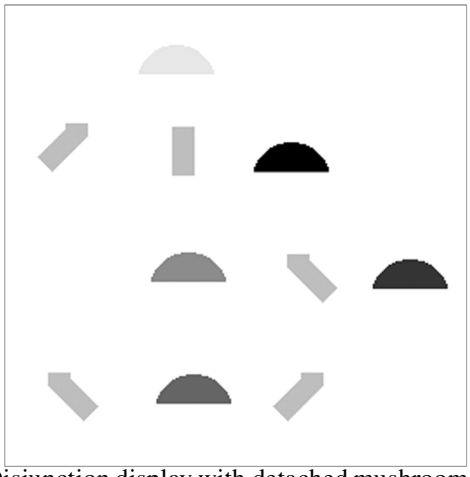


Figure 13. Results of Experiment 5. As in the previous three experiments, the amount of performance drop for monitoring two versus one relevant feature was smallest for the conjunction displays with the colored stems when both relevant features were located on the same part of an object, greater for the conjunction displays with the black stems when the relevant features were located on different parts of object, and greatest for the disjunction objects when the relevant features were located on separate objects. These results suggest that a hierarchical object-based feature encoding in visual short-term memory applies to parts regardless of whether they are defined by figure–ground separation or negative minima of curvature



(A) Conjunction display with attached mushroom parts



(B) Disjunction display with detached mushroom parts

Figure 14. Displays used in Xu (2002) with attached mushroom parts (A) and detached mushroom parts (B). Color and orientation features were found to be better encoded in visual short-term memory when the mushroom parts were attached than when they were detached, thus replicating the object-based encoding benefit for object parts found in the present study.

Overall, there were main effects of display type [$F(2,34) = 3.67$, $p < .05$] and monitoring condition [$F(1,17) = 123.02$, $p < .001$] and a significant interaction between the two [$F(2,34) = 23.06$, $p < .001$]. In pairwise comparisons, the interaction between monitoring condition and conjunction displays with the colored stems versus conjunction displays with the black stems was significant [$F(1,17) = 6.26$, $p < .05$], indicating that it was harder to encode the two relevant features when they were located on different parts of the same object than when they were located on the same part of an object. The interaction between monitoring condition and conjunction displays with the black stems versus disjunction displays was also significant [$F(1,17) = 15.66$, $p < .01$], indicating that there was still an object-based advantage for encoding features on different parts of the same object, as compared with encoding features on different objects.

Discussion

With parts defined by negative minima of curvature instead of figure–ground separation, virtually the same results were obtained in the present experiment as in Experiments 2–4, suggesting that a hierarchical object-based feature encoding in VSTM applies to parts regardless of whether they are defined by figure–ground separation or negative minima of curvature.

With slightly different stimuli, the results of the present experiment were replicated with a different group of participants in a different study. Xu (2002) asked participants to remember the colors of mushroom caps and the orientations of mushroom stems. As in the present experiment, the mushroom parts were either attached (Figure 14A), forming parts of the same objects, or detached (Figure 14B), forming separate objects. Color and orientation features were found to be better encoded in VSTM when the mushroom parts were attached than when they were detached, thus replicating the object-based benefit for object parts found in the present experiment.

Studies by Vecera and colleagues (Vecera, Behrmann, & Filapek, 2001; Vecera, Behrmann, & McGoldrick, 2000) have reached similar conclusions, using the feature report paradigm instead of the change detection paradigm. Using a display similar to that in Figure 15, Vecera et al. (2000; Vecera et al., 2001) found that features from the same part of an object (pronged bar length and arm orientation or straight bar length and gap location) are better reported than features from different parts of an object (e.g., pronged bar length and straight bar gap location).

GENERAL DISCUSSION

Using objects whose color and shape features were contained in the same part of an object (e.g., colored letters), researchers such as Allport (1971), Irwin and Andrews (1996), and Luck and Vogel (1997) found that two features of an object can be encoded in VSTM as accurately as just one feature of an object (see also Duncan, 1984). Motivated by the fact that most objects in our surroundings contain multiple parts, the present study addressed the question of whether the object-based encoding benefit in VSTM can still be found when the to-be-remembered features come from different parts of an object. This was done by using the change detection paradigm (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). In Experiment 1, Saturn-like objects were constructed with the "ring" carrying the relevant orientation feature and the "planet" carrying the relevant color feature. When both relevant features were monitored, the amount of feature information encoded from five of these objects

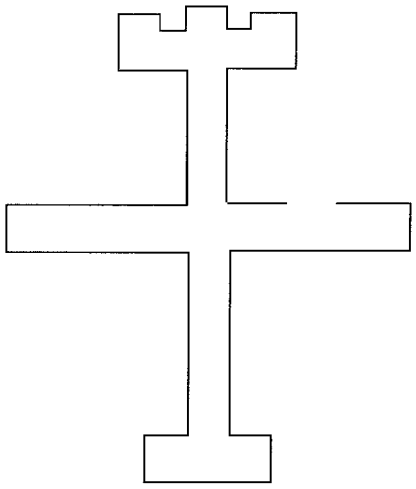


Figure 15. An example of the kind of stimuli used by Vecera, Behrmann, and McGoldrick (2000; see also Vecera, Behrmann & Filapek, 2001). These authors found that features from the same part of an object (pronged bar length and arm orientation or straight bar length and gap location) were better reported than features from different parts of an object (e.g., pronged bar length and straight bar gap location), which is consistent with the findings of the present experiments.

was significantly less than that from 5 colored circles or 5 black oriented bars. In fact, the encoding of displays containing five Saturn-like objects was about the same as that from displays containing the same five color and five orientation features distributed on spatially separated objects or from displays containing 10 colored circles or 10 oriented bars. These results indicated an absence of object-based encoding in VSTM for the Saturn-like objects.

In Experiment 1, however, visual complexity among the different displays was not properly controlled. Displays containing the Saturn-like objects were visually more complex than the other displays. As such, any object-based encoding advantage for the Saturn-like objects might have been erased by the added difficulties in encoding that were due to higher visual complexity. Experiment 2 addressed this concern. Instead of a comparison of the different display types directly, the amount of performance drop for monitoring both features versus just one relevant feature was measured and compared among the different display types (i.e., the interaction between monitoring condition and display type). With this paradigm, it was found that the amount of performance drop for monitoring both relevant features versus one relevant feature of the Saturn-like objects was significantly greater than it was for the colored oriented bars; however, it was still less than for the same two features distributed on spatially separated objects. In other words, features were best encoded in VSTM when they were from the same part of an object. Although less well encoded in comparison, features from different parts of an object were still better encoded than features from spatially separated objects.

One might argue that each of the Saturn-like objects used in Experiments 1 and 2 could be subjectively interpreted as consisting of two objects, with one on top of the other, rather than a single object with two parts, and that was why features of the Saturn-like objects were encoded less well than those of the colored oriented bars. By this account, if we could bias the interpretation of an object so that it would be more likely to be interpreted as a single two-part object than as two objects with one on top of the other, change detection performance should be affected accordingly. In Experiment 3, colored beachball-like objects replaced the Saturn-like objects in Experiment 2. For the colored beachball-like objects, the color of the ball carried the relevant color feature and a black stripe running across the ball carried the relevant orientation feature; most important, the ends of the black oriented bar now followed the contour of the colored circle. According to the generic view principle (Rock, 1983; see also Nakayama & Shimojo, 1992), when the ends of the black oriented bar exactly follow the contour of the colored circle, it is much more likely for the visual system to interpret the ensemble as one object and to interpret the black bar as being part of a textured surface on the object than to interpret the ensemble as two distinct objects with one lying on top of the other, only accidentally in perfect juxtaposition. In addition, given that all the other circle–bar pairs presented in the display were also in perfect alignment, the interpretation of each circle–bar pair as a colored beachball with a black stripe—a single two-part object—should be even stronger. And yet it was found that the two relevant features of the colored beachball-like objects were less well encoded in VSTM than were those of the colored oriented bar; in fact, the magnitude of the effect obtained in Experiment 3 was about the same as that in Experiment 2.

In addition, in Experiment 3, when the features of a colored beachball-like object were rearranged so that the relevant color and orientation features were contained within the same part of the object (the stripe) and the object was turned into a black beachball-like object with a color stripe, monitoring for both relevant features of the object became easier. Note that each of these black beachball-like objects contained the same number of shape and color features as the colored beachball-like objects, thus controlling for visual complexity. And yet, depending on whether the two relevant features were contained in the same part or different parts of an object, the amount of information encoded in VSTM differed significantly.

In Experiment 4, colored math symbols on black squares and black math symbols on colored squares were used. With the same paradigm as that in Experiments 2 and 3, it was found once again that two relevant features (color and shape) were harder to encode when features were from different parts of an object (black symbols on colored squares) than when these features were from the same part of an object (colored symbols on black squares). Although the two-part objects used in Experiment 4 could

be more subjectively interpreted as two independent objects with the symbols lying on top of the squares, the magnitude of the effect observed did not differ from that obtained in Experiments 2 and 3. It thus seems that this difficulty in integrating/binding features across parts (despite differences in subjective interpretation of parts) reflected some genuine limitations of the visual system that are independent of top-down influences.

In Experiment 5, instead of object parts defined by figure–ground separation being used, as in the previous four experiments, object parts defined by negative minima of curvature were used. Each object consisted of a circle attached to a stem. The relevant color and orientation features could be located on (1) the same part of an object (the stem), (2) different parts of an object (color on the circle and orientation on the stem, with the circle and stem attached at negative minima of curvature), or (3) spatially separated objects, similar to (2), but with the circle and the stem detached. With the same change detection paradigm as that in Experiments 2–4, virtually the same results were obtained: Color and orientation were best encoded when they were from the same part of an object, less well encoded when they were from different parts of the same object, and least well encoded when they were from spatially separated objects. These results indicate that a hierarchical object-based feature encoding in VSTM applies to parts regardless of whether they are defined by figure–ground separation or negative minima of curvature.

Experiments 3 and 4 also replicated the results of a series of studies that addressed the unitization effect in long-term memory: Memory is superior when features are perceived to be properties of a unitary stimulus, relative to when features are perceived to be properties of a nonunitary stimulus (Ceraso, 1985; Farr, 1961, as cited in Asch, 1969; Walker & Cuthbert, 1998; Wilton, 1989). In one of these studies, participants were presented with a stack of cards, each of which contained a color shape on a white card or a white shape on a colored card (Wilton, 1989). The participants were asked to examine each card and to try to remember the color associated with each shape, the color being either the color of the shape or the color of the background. After an irrelevant distractor task, the participants were probed with the shape to recall the color associated with that particular shape. Recall of the color was significantly better when the color was associated with the shape than when the color was associated with the background. The present experiments can be viewed as an extension of these previous findings to immediate memory, where recall is probed only 1 sec after the initial stimulus presentation. In addition, in the present study, feature encoding was compared not only when features belonged to the same part versus different parts of an object, but also when features belonged to different objects. Moreover, the present study extended the results both to parts defined by figure–ground separation and to parts defined by negative minima of curvature.

Perception and VSTM

One might argue that the present study conflates paradigms that measure the accuracy of perception with paradigms that measure the accuracy of memory, because most of the object-based attention studies cited in the introduction involved tasks in which participants made an immediate discriminative response when a target was presented, whereas in the present study VSTM was examined.

As was mentioned in the introduction, VSTM is defined as short-term memory for nonverbal, visual information, a buffer that temporarily stores visual information. VSTM is tightly connected with perception, because in order for perception to occur, visual information has to be encoded in some sort of short-term memory buffer before further processing can take place. Therefore, it would be impossible to separate perception from VSTM, from a theoretical point of view. In practice, paradigms used to study perception are not fundamentally different from those used to study VSTM. For example, paradigms used in most of the object-based attention studies described in the introduction were highly similar to the change detection paradigm used in the present study, except that, in change detection, (1) recognition instead of recall and (2) response after a delay of about 1 sec following display presentation, instead of an immediate response following the display and a mask, were used in the present study. Recognition and recall are just different ways of collecting responses; if anything, recognition is usually more accurate and is subjected to less interference than is recall. Phillips (1974) showed that using an immediate response following a display and a mask, rather than a response after a delay of about 1 sec following a display without a mask, does not alter change detection performance in any substantial way. As such, whether recognition or recall was immediately probed after a display and a mask or 1,000 msec later with no mask, performance should be quite comparable.

Phillips and Christie (1977) showed that information in VSTM persisted for at least 10 sec if the viewer's attention was maintained (see also Phillips, 1974). A recent study by Vogel, Woodman, Eads, and Luck (1998) found that after stimuli had been successfully encoded into VSTM, the presentation of a mask did not interfere with change detection performance. These results suggest that once information is encoded into VSTM, with no distraction, information does not decay rapidly. In other words, the maintenance of information in VSTM is quite good over a period of a few seconds. If there was a limitation in the participants' performance, it was, therefore, most likely due to a limitation in the encoding stage of VSTM. Note that this limitation in the encoding stage was not due to the fact that the participants did not have enough time to view and encode the stimuli in the display: Vogel et al. (2001) varied stimulus presentation time (100 vs. 500 msec) and found no difference in performance. Similar hierarchical object-based feature integration/binding has also been reported in visual conjunction searches that required only on-line processing of visual information (Goldsmith, 1998; Xu, in press). This find-

ing provides further evidence that the limitation observed in VSTM tasks is most likely due to a limitation in the encoding stage.

Given the fact that perception and VSTM are tightly connected and the fact that the limitation in VSTM is most likely due to a limitation in the encoding stage, I would like to argue that the present finding—object-based hierarchical feature encoding in VSTM—reflects a more general feature-binding process in perception.

Feature Encoding in VSTM and Treisman's Feature Integration Theory

The border between the black and the colored areas delineated orientation in Experiment 3 and shape in Experiment 4. At the local feature level, this border between the black and the colored areas (and thus, the perception of color and form) was identical for both the black beachball-like objects and the colored beachball-like objects in Experiment 3, as well as for the colored symbols on black squares and the black symbols on colored squares in Experiment 4. Yet, at the object level, figure–ground interpretation of color and shape was quite different for the two display types: In one case, the border was assigned to the same part of the object that also contained the color, and in the other case, the border was assigned to one part of the object and the color was assigned to the other part of the object. In other words, it must be the case that features are not linked together simply via a common spatial location, as Treisman proposed (Treisman, 1988; Treisman & Gelade, 1980). If that were the case, no difference would be observed in change detection performance between the black beachball-like objects and the colored beachball-like objects or between the colored symbols on black squares and the black symbols on colored squares. Rather, features are interpreted first and assigned to appropriate objects or parts of an object before any feature integration/binding in VSTM can take place. As was mentioned earlier, the same feature interpretation and assignment have also been shown to affect efficiencies in visual conjunction search (Goldsmith, 1998; Xu, in press).

Nonetheless, if the relationship between object parts (such as figure–ground) is added to Treisman's model, the present results could be easily accommodated in the general framework of Treisman's feature integration theory.

Conclusion

In summary, with object parts defined by either figure–ground separation or negative minima of curvature, the results of the present study show that features are encoded in VSTM in a hierarchical manner, so that features are best encoded in VSTM when color and shape are from the same part of an object, less well encoded when they are from different parts of an object, and least well encoded when they are from spatially separated objects. To conclude, the present study shows that object-based feature binding is modulated by how features are assigned to the parts of an object, and that an object-based feature binding exists even when the color and shape features to be remembered are from different parts of an object.

REFERENCES

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B'' to below-chance performance. *Psychological Bulletin*, **102**, 439-442.

Allport, D. A. (1971). Parallel encoding within and between elementary stimulus dimensions. *Perception & Psychophysics*, **10**, 104-108.

Asch, S. E. (1969). A reformulation of the problem of associations. *American Psychologist*, **24**, 94-102.

Baylis, G. C., & Driver, J. (1994). Parallel computation of symmetry but not repetition in single visual objects. *Visual Cognition*, **1**, 377-400.

Baylis, G. C., & Driver, J. (1995). Obligatory edge assignment in vision: The role of figure and part segmentation in symmetry detection. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1323-1342.

Ceraso, J. (1985). Unit formation in perception and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 19, pp. 179-210). New York: Academic Press.

Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society*, **31**, 271-274.

Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology*, **12**, 75-96.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.

Goldsmith, M. (1998). What's in a location? Comparing object-based and space-based models of feature integration in visual search. *Journal of Experimental Psychology: General*, **127**, 189-219.

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

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition*, **63**, 29-78.

Hulleman, J., te Winkel, W., & Boselie, F. (2000). Concavities as basic features in visual search: Evidence from search asymmetries. *Perception & Psychophysics*, **62**, 162-174.

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

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125-155). Cambridge, MA: MIT Press, Bradford Books.

Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, **63**, 253-257.

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, **257**, 1357-1363.

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 non-parametric analysis of recognition experiments. *Psychonomic Science*, **1**, 125-126.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing silhouettes: The short-cut rule. *Perception & Psychophysics*, **61**, 636-660.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
Treisman, A. (1988). Features and objects: The Fourteenth Barlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, **40A**, 201-237.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
Vecera, S. P., Behrmann, M., & Filapek, J. C. (2001). Attending to the parts of a single object: Part-based selection limitations. *Perception & Psychophysics*, **63**, 308-321.
Vecera, S. P., Behrmann, M., & McGoldrick, J. (2000). Selective attention to the parts of an object. *Psychonomic Bulletin & Review*, **7**, 301-308.
Vogel, E. K., Woodman, G. F., Eads, A. C., & Luck, S. J. (1998, April). *Masking in visual working memory: Evidence for a limited-capacity encoding mechanism*. Poster presented at the 1998 Cognitive Neuroscience Society Annual Meeting, San Francisco.
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 & Performance*, **27**, 92-114.
Walker, P., & Cuthbert, L. (1998). Remembering visual feature conjunctions: Visual memory for shape–color associations is object-based. *Visual Cognition*, **5**, 409-455.
Wilton, R. N. (1989). The structure of memory: Evidence concerning the recall of surface and background color of shapes. *Quarterly Journal of Experimental Psychology*, **41A**, 579-598.
Wing, A., & Allport, D. A. (1972). Multidimensional encoding of visual form. *Perception & Psychophysics*, **12**, 474-476.
Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, **37**, 25-43.
Xu, Y. (2000). *Object parts in visual short-term memory and visual search*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
Xu, Y. (2002). Limitations in object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 458-468.
Xu, Y. (in press). Object-based feature integration in visual search. *Perception*.
Xu, Y., & Potter, M. C. (2000). *Capacity and form of representation in visual short-term memory*. Manuscript submitted for publication.
Xu, Y., & Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception & Psychophysics*, **67**, 1039-1054.

NOTES

1. Note that although change detection accuracy decreases for monitoring both features of an object, as compared with just one feature of an object, that does not necessarily mean that VSTM cannot encode both features of an object as accurately as just one feature of an object. Decision noise may have contributed to the observed decrement in performance. When both features are being monitored, a false alarm on either feature will add to the overall false alarm rate. Therefore, even if both features are encoded exactly as accurately as they are when only one feature is monitored, the actual decision will have more errors (see, e.g., Duncan, 1980).

2. One might argue that, for the Saturn-like objects when only one feature was monitored, the performance was low because the participants encoded both features of these objects, even though the encoding of only one feature was required. This argument, however, would only suggest that the features of the Saturn-like objects were better integrated than the features of the disjunctive objects, which was the exact implication of the present finding. On the other hand, this argument cannot explain why performance dropped significantly for the Saturn-like objects when both features had to be encoded. Therefore, the lower performance observed for the Saturn-like objects, as compared with that for the disjunctive objects, when only one feature had to be monitored was most likely due to the high visual complexity associated with the Saturn-like objects.

(Manuscript received June 19, 2000;
revision accepted for publication March 22, 2002.)