Journal of Experimental Psychology:
Human Perception and Performance
2002, Vol. 28, No. 2, 458–468

0096-1523/02/$5.00 DOI: 10.1037//0096-1523.28.2.458

Limitations of Object-Based Feature Encoding in Visual Short-Term Memory

Yaoda Xu
Massachusetts Institute of Technology

The present study investigated object-based feature encoding in visual short-term memory for 2 features within the same dimension that occur on different parts of an object. Using the change-detection paradigm, this experiment studied objects with 2 colors and objects with 2 orientations. Participants found it easier to monitor 1 rather than both features of such objects, even when decision noise was properly controlled for. However, no object-based benefit was observed for encoding the 2 features of each object that were of the same dimension. When similar stimuli were used but the 2 features of each object were from different dimensions (color and orientation), an object-based benefit was observed. These results thus impose a major constraint on object-based feature encoding theories by showing that only features from different dimensions can benefit from object-based encoding.

How are features of objects perceived and retained? In an early study by Allport (1971), colored numerals inside colored shapes were presented briefly for recall of one or more features. It was found that report of a form feature (e.g., the shapes) was not affected by whether a color feature (colors of the shapes or numerals) was also reported, but was negatively affected by the report of another form feature (e.g., the numerals). A second study by Wing and Allport (1972) confirmed these findings and found that the report of both spatial frequency and orientation did not affect either report, but the report of two orientation features (grating orientation and an orientation of a transverse break in the lines of the grating) interfered with each report significantly. These authors argued that perceptual analysis occurs in systems of analyzers, each dealing with a specific feature dimension such as color or orientation. As such, two features from different dimensions can be encoded in parallel without mutual interference, but two features from the same dimension will have to share the same analyzer and cannot be encoded without interference (see also Treisman, 1969).

Duncan (1984) questioned the above conclusion and argued that feature grouping by objects plays a more important role in perception. He asked participants to report two of four features from four independent dimensions (size, tilt, texture, and the location of a gap), either located on the same object or on different objects. He found that two features located on the same object could be reported more readily than two features located on different objects. In a later study in which two letters were presented, Duncan (1993) asked participants to report (a) the size and shape of one letter, (b) the same attribute (size or shape) of two letters, (c) the shape of one letter and size of the other, or (d) both attributes of both letters. He found that report of (a) was much better than that of (b), (c), or (d) and that the differences among (b), (c), and (d) were not significant. In a second experiment, Duncan (1993) presented two objects, each containing orientation, length, and frequency features, and asked participants to report two features, one from each object. The two features could either be from different dimensions (e.g., orientation of one object and length of the other) or from the same dimension (e.g., orientations of both objects). He found that accuracy of report did not differ according to whether the two features were from the same or different dimensions. Duncan (1993) thus concluded that (a) features from the same objects are better encoded than features from different objects if these features are from different feature dimensions, and (b) when features are from different objects, the encoding of features from the same versus different dimensions does not differ.

Although one may argue that the features of shape, size, orientation, and spatial frequency used by Duncan (1984, 1993) are all contour or boundary features and may not be completely independent of each other, Duncan and Nimmo-Smith (1996) were able to obtain the same results on more distinct features such as length and color. More recently, Duncan's (1993) results were confirmed by Lee, Koch, and Braun (1999). Lee et al. conducted a series of experiments in which two stimuli were presented in the same display, one in the center position and one in the peripheral position, and participants were asked to perform two discrimination tasks, one on each of the stimuli. The two discrimination tasks involved either two of the three feature dimensions (form, color, and motion) or the same feature dimension (e.g., both tasks involved form discrimination). It was found that interference was indistinguishable for discrimination tasks involving the same versus different dimensions.

Yaoda Xu, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

The present research was 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. I am grateful to Mary C. Potter for her valuable advice, comments, and encouragement throughout the project. I would also like to thank John Duncan and two anonymous reviewers for helpful comments on earlier versions of this article. This research was supported by National Institute of Mental Health Grant MH 47432 to Mary C. Potter and by a McDonnell–Pew Investigator Initiated Grant in Cognitive Neuroscience to Yaoda Xu.

Correspondence concerning this article should be addressed to Yaoda Xu, who is now at Vision Sciences Laboratory, Psychology Department, Harvard University, 33 Kirkland Street, Room 744, Cambridge, Massachusetts 02138. E-mail: yaoda@wjh.harvard.edu

Table 1 shows a summary of results obtained by previous studies: (a) When features are distributed on separate objects, the encoding of features from the same dimension versus from different dimensions does not differ (Table 1, cells C and D), as revealed by Duncan (1984, 1993) and Lee et al. (1999); and (b) when features from different dimensions belong to the same object, they can be encoded in parallel without mutual interference (Table 1, cell A), as shown by Allport (1971), Duncan (1984), Wing and Allport (1972), and many others, including Irwin and Andrews (1996), Luck and Vogel (1997), and myself (Xu, 2000a—although in that study, using the change-detection paradigm, I did find a small but significant drop in performance for encoding two features vs. one feature).[1]

One issue, however, is left unresolved in Table 1: What about features of the same object that are from the same feature dimension (Table 1, cell B), for example, two color features or two orientation features of the same object? Both the colored numerals inside colored shapes, used by Allport (1971), and the patch of oriented lines with a transverse break in the lines of the grating, used by Wing and Allport (1972), could be considered as single objects with two features on the same dimension. The results found by Allport and his colleague thus seem to fill in the unresolved cell in Table 1 by showing that features from the same dimension cannot be encoded without mutual interference, even when they are from the same object.

The definition of objecthood, however, was never entirely clear in the above experiments. Duncan (1984) viewed the colored numerals inside colored shapes as two objects, with one inside the other; and similarly, he viewed the break in the patch of oriented lines as a white band lying across the grating patch. Therefore, the stimuli used by Allport (1971) and Wing and Allport (1972) could be interpreted as two features on two different objects. As Marr (1982) once asked,

> 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? . . . [A]ll 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)

Moreover, for a stimulus such as a colored oriented bar, the two features are in exactly the same spatial location. This, however, is impossible to achieve when two features of an object belong to the same feature dimension, such as two color features of an object. The two color features will necessarily have to occupy different spatial locations (no matter how close) to be perceived correctly. Therefore, it is unclear whether a fair comparison can ever be made between the encoding of two features from the same dimension and the encoding of two features from different dimensions when in both cases both features belong to the same object. Any differences obtained could either be caused by differences in the feature dimensions or by differences in the spatial locations that these features occupy.

Using the change-detection paradigm of Phillips (1974; see also Pashler, 1988), I (Xu, 2000a) studied the encoding of color and orientation features located on the same versus different parts of an object. I found that the normal object-based benefit for encoding two features of the same object was modulated by the spatial relationship between the features. In one of my experiments (Xu, 2000a), there were three display conditions: In Condition A, colored oriented bars were presented, and the relevant features were the colors and orientations of the bars; in Condition B, colored circles, each with a black stripe running across it, were presented, and the relevant features were the orientations of the black stripes and the colors of the circles; and in Condition C, spatially separated black oriented bars and colored circles were presented. Either the relevant color, the relevant orientation, or both relevant features were monitored. On change trials, only one feature of one object was changed. For each display type, the difference in change-detection performance for monitoring one versus two features was taken to indicate the degree of integration of the two features. I found that features were best integrated and retained in visual short-term memory (VSTM) when color and orientation belonged to the same part of an object (Condition A), such that performance dropped only slightly (although significantly) for monitoring two features versus one feature; when these two features belonged to different parts of an object (Condition B), the drop in performance for monitoring two features versus one feature was much greater; the greatest drop in performance occurred when features were distributed in spatially separated objects in Condition C.

I (Xu, 2000a) thus extended the object-based feature-encoding effect by showing that even when color and orientation features were not at the exact same location but were on different parts of the same object (Condition B), a significant encoding advantage in VSTM was still observed compared with when these features were on spatially separated objects (Condition C). This result thus enables me to pursue the question raised earlier, namely, can a similar object-based benefit be observed for two features of the same dimension when these features are located on different parts of an object? Duncan's (1984) model appears to predict an object-based advantage regardless of whether features are from the same or different dimensions, as long as they are from the same object. In contrast, something similar to the analyzer model proposed by Allport (1971; Wing & Allport, 1972) would predict that features from the same dimension cannot be encoded without mutual interference, whether these features are from the same object or from different objects.

Using the change-detection paradigm, Luck and Vogel (1997; see also Vogel, Woodman, & Luck, 2001) found that both color

[1] One may argue that the present study conflated paradigms that measure the accuracy of perception with paradigms that measure the accuracy of memory. Most of the object-based attention results I have described so far are taken from tasks in which participants made an immediate discriminative response when a target was presented, whereas the present study examined VSTM. However, paradigms used in most of the object-based attention studies described so far are highly similar to the change-detection paradigm used in the present study except that the present study used (a) recognition instead of recall and (b) response after a delay of about 1 s following display presentation instead of immediate response following the display presentation and a mask. Recognition and recall are just different ways of collecting responses; if anything, recognition is usually more accurate and is subject to less interference than is recall. Phillips (1974) showed that using (a) immediate response following the display presentation and a mask versus (b) response after a delay of about 1 s following display presentation without masks did not alter change-detection performance in any substantial way. As such, whether recognition or recall takes place immediately after the display and a mask or 1,000 ms later with no mask, performance should be comparable. Nevertheless, further studies are needed to verify that results obtained in the present study in the context of VSTM are replicable in a paradigm involving immediate recall.

Table 1
The Interactions of Objecthood and Feature Dimensions to Be Reported or Monitored

Objecthood	Features from different dimensions	Features from the same dimension
Same object	A Performance does not decrease as more features per object are to be reported. Allport (1971), Duncan (1984), and Luck & Vogel (1997)	B Same as A? Allport (1971) and Wing & Allport (1972): No Luck & Vogel (1997): Yes
Different objects	C Performance decreases as more objects are to be reported. Duncan (1984, 1993) and Lee et al. (1999)	D Same as C Duncan (1984, 1993) and Lee et al. (1999)

and orientation features of a colored oriented bar can be retained as accurately as just one feature encoded alone, replicating the findings of Allport (1971; Wing & Allport, 1972) and Duncan (1984, 1993). Moreover, Luck and Vogel (1997) were able to extend this result to objects consisting of a large square of one color and a smaller inner square of a different color. They found that change-detection performance was as accurate with these bicolored objects as it was with single-colored objects. That is, there was an object-based encoding advantage for these bicolored objects.

This last result was challenged by Xu and Potter (2000; see also Xu, 2000b) and by Wheeler and Treisman (2000; see also Wheeler & Treisman, 1999), who failed to replicate the results of Luck and Vogel (1997) and found that performance for the same bicolored objects was much worse than performance for objects containing only one color: It was the total number of color features, not the number of objects, that determined performance. Xu and Potter (2000) compared change detection for six single-colored squares with that for six bicolored squares and found that performance was significantly worse for the latter. Wheeler and Treisman (2000) compared performance among (a) three bicolored squares, (b) three single-colored squares, and (c) six single-colored squares and found that (a) was significantly worse than (b) and that (a) and (c) were indistinguishable from each other. In both Xu and Potter (2000) and Wheeler and Treisman (2000), however, displays containing the bicolored objects were visually more complex than displays containing the single-colored objects. Without controlling for these differences in the displays, a direct comparison of performance across displays is not appropriate. A color feature may be less well encoded when it is next to another color feature (as in displays containing bicolored squares) than when it stands alone (as in displays containing single-colored squares). As such, any advantage gained from object-based encoding for the bicolored objects might have been erased by the visual complexity of these displays. The paradigm I used previously (Xu, 2000a), however, did not make direct comparisons among the different display types. Instead, interactions between monitoring condition and display type were used as a more reliable way of assessing object-based encoding.

The goal of the present study was therefore to investigate whether any object-based encoding advantage in VSTM could be found for two features of the same dimension, using the paradigm from my previous study (Xu, 2000a). In Experiment 1, objects with two colors were studied, and in Experiment 2, objects with two orientations were studied. In Experiment 3, object-based encoding was compared between features of the same dimension (two colors) and features of different dimensions (a color and an orientation). Experiments 4 and 5 used an improved design to investigate whether the performance drop observed in Experiments 1 and 2 for monitoring one versus two features was caused by decision noise. Experiment 5 also provided further evidence supporting the conclusions reached from Experiment 3.

Experiment 1

In this experiment, object-based encoding was studied in objects with two colors. To control for visual complexity, mushroom-like objects were used in which the shape of the mushroom cap was distinct from that of the mushroom stem (Figure 1). This made it possible for participants to attend to only the colors of the mushroom caps or the colors of the mushroom stems.[2] The cap color and the stem color made up the two color features of each mushroom-like object and were always different from each other. There were two display types: either the mushroom caps and stems were attached (conjunction display; Figure 1A) or they were detached (disjunction display, Figure 1B). Participants were asked to monitor either the color of the mushroom caps, the color of the mushroom stems, or the color of both mushroom parts. If object-based feature encoding exists for objects with two colors, then a significant interaction between display type and monitoring condition should be observed, such that the amount of performance drop for monitoring two features versus one feature is much smaller for the attached mushroom parts than for the detached mushroom parts.

In each trial of the experiment, an initial sample display was presented and was briefly followed by a 1-s delay. A test display was then shown. The test display was identical to the first display except that in 50% of the trials, one feature of one object had changed. Participants were instructed to detect this change.

Method

Participants. Twelve participants (5 women and 7 men) were recruited from the Massachusetts Institute of Technology campus. They were all between 17 and 40 years of age, had normal color vision, and were paid for their participation.

Materials and design. The stimuli used are shown in Figure 1. When viewed from the normal viewing distance of 50 cm, the sizes of the cap and the stem were 1.46° × 0.52° and 0.94° × 0.42°, respectively. Six colors (red, blue, green, orange, violet, and white) were used. All items were

[2] Attending to one of the colored parts of the bicolored objects used by Luck and Vogel (1997), Xu and Potter (2000), and Wheeler and Treisman (2000) proved to be difficult for participants because the two color features shared the same shape feature.

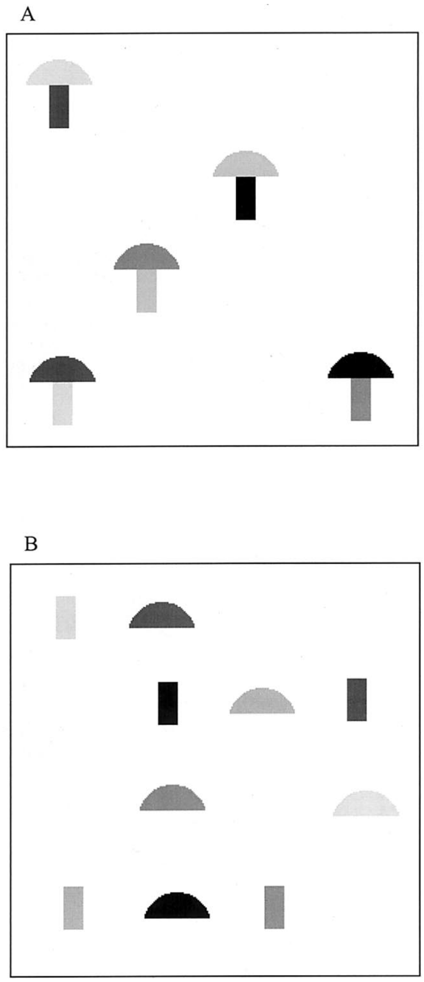

Figure 1. Examples of stimuli used in Experiment 1. A: Conjunction display with attached mushroom parts. B: Disjunction display with detached mushroom parts.

displayed on a black background. 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. In the conjunction displays, five attached pairs of caps and stems were shown (Figure 1A). In the disjunction displays, five detached pairs of caps and stems were shown (Figure 1B). There were three feature-monitoring conditions: (a) only the cap colors were monitored for a possible cap color change, (b) only the stem colors were monitored for a possible stem color change, or (c) both the cap and the stem colors were monitored for a possible color change in either the cap or the stem. Trials were blocked by display type and by feature-monitoring condition, the order of which was balanced across participants. Each block contained 32 trials, with 16 change trials and 16 no-change trials. For each display type in which only one color feature (cap or stem) was monitored, there were a total of 64 trials, with 32 change trials and 32 no-change trials evenly distributed into two blocks; and for each display type in which both color features were monitored, there were a total of 128 trials, with 32 cap-color-change trials, 32 stem-color-change trials, and 64 no-change trials evenly distributed into four blocks. Ten practice trials preceded the experimental trials in each condition.

The computer randomly chose five of the six prespecified colors as the five cap colors. This procedure was repeated to generate the five stem colors. When a cap color changed, the changed color feature was the sixth color not previously assigned to any caps. The same procedure was used for stem color assignment.[3] In both the sample and test displays, when caps and stems were attached, the cap and the stem of a given mushroom always had different colors.

There were four implicit quadrants in the display, each divided into 2 × 2 arrays. The objects were distributed over the four quadrants as follows: For the conjunction displays, two random positions were selected from one randomly chosen quadrant, and one random position was selected from each of the three remaining quadrants; for the disjunction displays, three random positions were selected from each of two randomly chosen quadrants, and two random positions were selected from each of the two remaining quadrants. This sampling procedure ensured that object locations were evenly distributed for any given display and that both the conjunction and disjunction displays occupied similar envelopes.

Apparatus. The displays were generated by an iMac with a 350-MHz Power PC G3 processor and a 15-in. (38.1-cm) monitor. The script for the experiment was generated by MacProbe Macintosh programming software.

Procedure. Participants were seated in a dimly lit, quiet room about 50 cm from the screen. They initiated each trial by pressing the space key on the computer keyboard. Each trial began with a fixation dot at the center of the screen for 505 ms, followed by the sample display for 400 ms. The sample display was then replaced by the blank, black background. After 1,000 ms the test display appeared. The test display remained on the screen until the participant made a keypress. Participants were asked to press the left "control" key with their left index finger if they detected a change and to press the "enter" key on the number keypad with their right index finger if they did not detect any change. The keys were labeled with the words *Different* or *Same*, respectively. As soon as the participant responded, 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 263 ms. The next trial started about 1 s after the feedback for the previous trial disappeared. Within a block, trials occurred one after another. There were breaks between the blocks during which participants could rest for as long as they wanted. The whole experiment lasted about 40 min.

Results

Xu and Potter (2000) provided evidence that feature information is represented in VSTM in a graded manner. Therefore, only models that assume graded information representation in a detection task would be appropriate for characterizing VSTM change-detection performance. Among these models is the signal-detection theory, which generated the measure d′ and inspired the measure A′ (Grier, 1971; Pollack & Norman, 1964). In the present analyses, A′ was used instead of d′ because, in general, A′ is more accurate than d′ (Donaldson, 1993) and A′ does not have the indeterminacy of d′ when a participant makes no false "yes" responses. A′ was calculated for each participant in each condition following the formula developed by Grier (1971):

$$A' = 0.5 + \left[\frac{(H - g)(1 + H - g)}{4H(1 - g)}\right],$$

[3] Because each stimulus array contained five of six possible colors, one may argue that participants could have used an "exclusion" strategy to detect a color change. That is, they may have simply determined which color was absent in the sample display and then looked for this color in the test display. The same color-sampling procedure, however, was used for both the conjunction and the disjunction displays, so even if this strategy was used, the results should not have been distorted. Moreover, the same results were obtained when the use of this strategy was not possible in Experiment 2.

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

$$A' = 0.5 - \left[\frac{(g - H)(1 + g - H)}{4g(1 - H)}\right].$$

When participants had to monitor for both color features and detect a change in either cap color or stem color, separate false-alarm rates for cap and stem color changes could not be recorded; as a result, A′ could not be computed separately for cap- and stem-color change detections. Performance was therefore averaged over cap- and stem-color change detections before A′ was computed for each participant in each condition. The final means of A′ are plotted in Figure 2.

Overall performance for disjunction displays was slightly higher than that for conjunction displays; however, this difference did not reach significance, $F(1, 11) = 2.15$, $p > .10$. Monitoring one color feature was shown to be significantly easier than monitoring both color features, $F(1, 11) = 36.55$, $p < .001$. The most important question, however, was whether there was any significant interaction between display type and monitoring condition. The analysis indicated that this interaction was not significant, $F < 1$.

Discussion

The result of the present experiment showed the following: (a) It is not the case that for a bicolored object, both color features are always encoded together regardless of the task demand; on the contrary, participants were able to encode just one of the two color features, as revealed by a significantly better performance for encoding one versus two color features in displays containing the attached mushroom objects. (b) No object-based encoding advantage for the bicolored objects was found, as indicated by the absence of any significant interaction between monitoring condition (attending to one vs. both color features) and display type (conjunction vs. disjunction displays). Thus, unlike Luck and Vogel (1997) and Vogel et al. (2001), this experiment failed to show any object-based benefit for two color features of the same object.

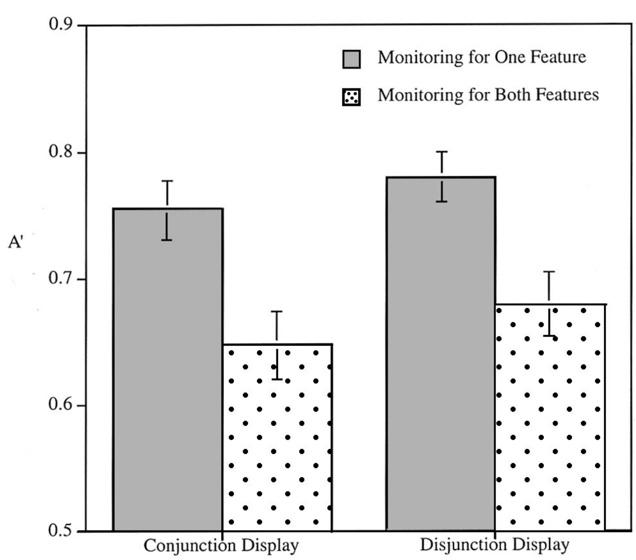

Figure 2. Results of Experiment 1 for monitoring color–color conjunction and disjunction displays, averaged over cap- and stem-color change detections. Error bars represent standard errors of the means.

Experiment 2

In this experiment, object-based encoding advantage for objects with two orientations was studied. Each object contained a red bar and a green bar, and the intersection of the two bars was colored yellow to form a transparent perception (Figure 3A). Consequently, neither colored bar was occluded behind the other bar, thus ensuring a clear perception of the orientation of each bar. To avoid the well-learned configurations of "+" and "×," and thus a more sensitive detection of "+" and "×" in the display, only three orientations of the bar were chosen: 45°, 105°, and 165°.

As in Experiment 1, there were two display types: either the red and green bars were attached (conjunction display, Figure 3A) or they were detached (disjunction display, Figure 3B). Participants were asked to monitor for either the orientation of the red bars, the orientation of the green bars, or the orientations of both bars. If object-based encoding exists for objects with two orientations, a significant interaction between display type and monitoring condition should have been observed.

Method

Participants. Twelve participants (4 women and 8 men) were recruited from the same participant pool as in Experiment 1. They were paid for their participation.

Materials and design. The stimuli used are shown in Figure 3. The length and width of the bars were 1.56° × 0.42°, respectively. Three orientations (45°, 105°, 165°) were used. All items were displayed on a black background. 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. In the conjunction displays, five attached pairs of red and green bars were shown (Figure 3A). In the disjunction displays, five detached pairs of red and green bars were shown (Figure 3B). As in Experiment 1, there were three feature-monitoring conditions: Either (a) only the red bars were monitored for a possible red-bar orientation change; (b) only the green bars were monitored for a possible green-bar orientation change; or (c) both the red and green bars were monitored for a possible orientation change in either a red or a green bar.

The computer randomly assigned the orientations to the red and green bars with the constraint that each orientation appeared at least once in the red bars and once in the green bars in the sample display. In both the sample and test conjunction displays, the attached red and green bars of a given conjunction object always had different orientations. When an orientation change occurred to one of the bars in a conjunction object, the new orientation was the third orientation not previously assigned to either the red or green bar of that object. All other aspects of the design and procedure were identical to those of Experiment 1.

Results

As in Experiment 1, when participants had to monitor for both orientation features and detect a change in either one, A′ scores could not be computed separately for red- and green-bar change detections. Performance was therefore averaged over red- and green-bar orientation change detections before A′ was computed

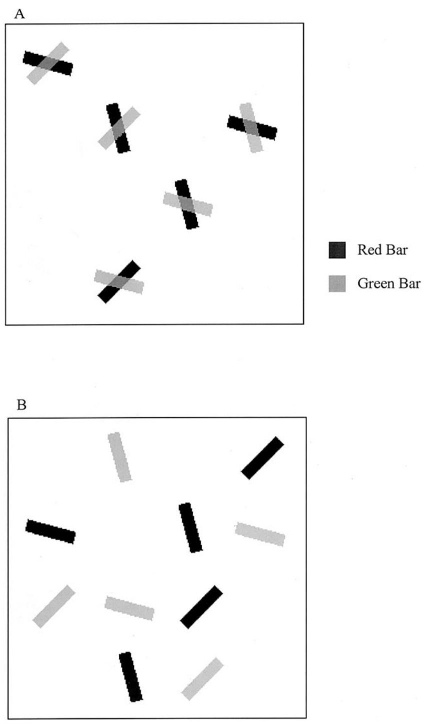

Figure 3. Examples of stimuli used in Experiment 2. A: Conjunction display. B: Disjunction display.

for each participant in each condition. The final means of A′ are plotted in Figure 4.

Overall, performances on the two display types differed from each other significantly, $F(1, 11) = 5.55$, $p < .05$, such that performance for disjunction displays was better than that for conjunction displays. Monitoring for one orientation feature was significantly easier than monitoring for both orientation features, $F(1, 11) = 26.47$, $p < .001$. More important to the present experiment was that the interaction between display types and monitoring condition was far from significant, $F < 1$.

Discussion

The results of the present experiment mirrored those of Experiment 1: First, participants were able to encode just one of the two orientation features in the conjunction display when the task required so; and second, no object-based feature encoding for objects with two orientations was found, as indicated by the absence of any significant interaction between monitoring condition (attending to one vs. both orientation features) and display type (conjunction vs. disjunction displays).

Together, the results of Experiments 1 and 2 show that object-based feature encoding does not exist for two features of an object when these features belong to the same feature dimension. One may argue, however, that the two parts of an object used in Experiments 1 and 2 were quite detachable, unlike the parts I used in my earlier experiment (Xu, 2000a), in which object-based feature encoding for color and orientation features from different parts of an object was found. In that study, one object part was always embedded in another, and the two object parts were therefore more closely bound than they were in the present experiments. This may explain the failure of observing any object-based feature encoding effect in the present study. To examine this possibility, in Experiment 3, the mushroom-like objects used in Experiment 1 were modified to represent color and orientation features on different parts of the object. Performance on these objects was then compared with performance on the bicolored objects used in Experiment 1.

Experiment 3

Experiment 3 contained two parts. In Part A, the mushroom-like objects were used to test object-based encoding of color and orientation features from different parts of an object. To achieve this, the relevant color feature was carried by the color of the mushroom cap and the relevant orientation feature was carried by the bending of the mushroom stem, as shown in Figure 5A. All the mushroom stems had the same light-greenish color (not one of the cap colors), and each stem was in one of three orientations relative to horizontal: 45°, 90°, or 135°.

In Experiment 1, when cap color was the only feature monitored, cap-color change-detection performance as measured by A′ was 0.75 ($SD = 0.03$) for conjunction displays, 0.76 ($SD = 0.03$) for disjunction displays,[4] and the differences between the two were not significant, $F < 1$. Similarly, a pilot study with 6 participants showed that when only stem orientation was monitored, change-detection performance did not differ according to whether the mushroom caps and stems were attached or detached ($F < 1$). In other words, when only one feature was monitored (either the cap color or the stem orientation), change-detection performance did not differ between the conjunction and disjunction displays. Therefore, instead of comparing performance when one feature versus two features were monitored, in Experiment 3 participants always monitored both features, and performance on the conjunction and disjunction displays was directly compared.

Part B of the experiment was a partial replication of Experiment 1 with the bicolored mushroom-like objects. The results of Experiment 1 showed that when only one feature was monitored (either the cap or the stem color), change-detection performance did not differ between conjunction and disjunction displays ($F < 1$). Therefore, as in Part A, participants always monitored both relevant features, and the conjunction and disjunction displays were compared directly to test the existence of any object-based feature encoding. All participants were tested in both Part A and Part B.

Method

Participants. Twenty-four participants (13 women and 11 men) were recruited from the same participant pool as in the preceding experiments. They were paid for their participation.

Materials and design. In Part A, as shown in Figure 5, color–orientation mushroom-like objects were used. The mushroom stems were a light-greenish color and were in one of three orientations (45°, 90°, or

[4] When only one of the two color features was monitored, participants monitored cap and stem color changes in separate blocks. Consequently, separate hit and guessing rates for monitoring cap and stem color changes could be measured, which enabled the calculation of separate A′ scores for cap- and stem-color change performance. This was not possible when both color features were monitored because only one guessing rate could be recorded.

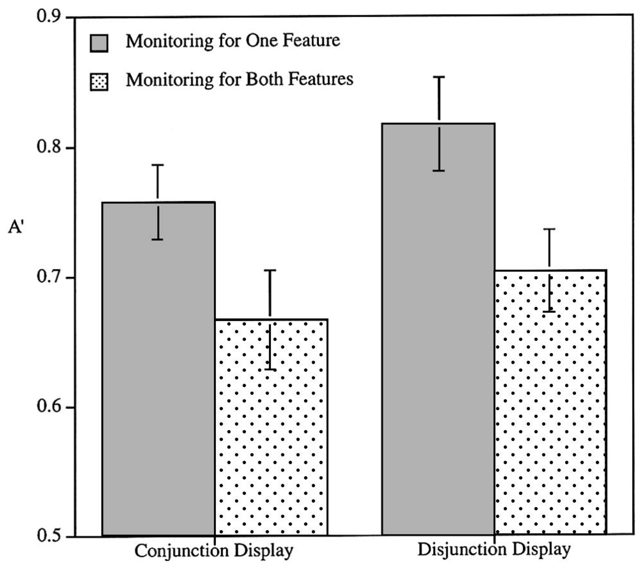

Figure 4. Results of Experiment 2 for monitoring orientation–orientation conjunction and disjunction displays, averaged over red- and green-bar-orientation change detections. Error bars represent standard errors of the means.

135°). The cap colors used were identical to those in Experiment 1. Part B of this experiment used the same stimuli as in Experiment 1 (see Figure 1).

There were two display types in Part A: Conjunction displays, which contained five pairs of attached mushroom caps and stems (Figure 5A), and disjunction displays, which contained five pairs of detached mushroom caps and stems (Figure 5B). Participants monitored both the cap colors and the stem orientations for a possible color change in one of the caps or a possible orientation change in one of the stems. The two display types used in Experiment 1 were used in Part B, and participants monitored the colors of both caps and stems for a possible color change in either feature.

For the mushroom-like objects shown in Figure 5, cap color assignment was identical to that described in Experiment 1. Orientation was assigned randomly to each object in the sample display with the constraint that the same orientation was used no more than twice in a given display. When a stem changed orientation, the changed value was chosen randomly from the two remaining values.

For each part of the experiment, there were a total of 128 trials for each display type, with 32 change trials for each of the two features and 64 no-change trials. The change and no-change trials were distributed evenly in four blocks of 32 trials each. The four blocks of trials of a given display type were always presented together. The presentation order of Parts A and B and the presentation order of the conjunction and disjunction displays within each part were counterbalanced across participants. Ten practice trials preceded the experimental trials in each condition. All other aspects of the design and procedure were identical to those of Experiment 1.

Results

As in Experiment 1, A′ was calculated for each participant in each condition after performance was averaged over cap and stem change detections. The results are plotted in Figure 6. For color–orientation mushroom objects (Part A), the difference between the conjunction and disjunction displays was highly significant, $F(1, 23) = 9.61$, $p < .01$. For bicolored mushroom objects (Part B), the difference between the conjunction and disjunction displays was not significant, $F < 1$. Direct comparison of Part A and Part B showed that the interaction between display type (conjunction vs. disjunction) and object condition (color–orientation vs. color–color objects) was significant, $F(1, 23) = 4.39$, $p < .05$.

Discussion

The results of Part A revealed a marked object-based encoding effect for color–orientation mushroom-like objects, such that when the relevant features were on different parts of the same object, they were much better retained in VSTM than when these features were on separated objects. On the other hand, the results of Part B replicated those of Experiment 1 and showed no object-based encoding advantage for the two color features of the bicolored mushroom-like objects.

In summary, with similar stimuli in Part A and Part B and the same group of participants in each part, Experiment 3 revealed a strong object-based encoding advantage when two features of an object were from different feature dimensions—thus replicating my previous findings (Xu, 2000a)—but again failed to reveal any object-based encoding benefit when the two features of an object came from the same feature dimension.

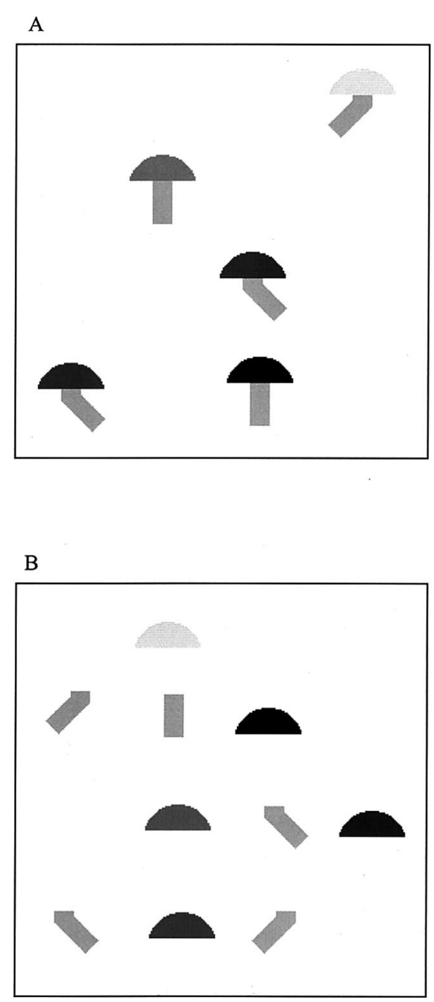

Figure 5. Examples of stimuli used in Experiment 3. A: Conjunction display with attached mushroom parts. B: Disjunction display with detached mushroom parts.

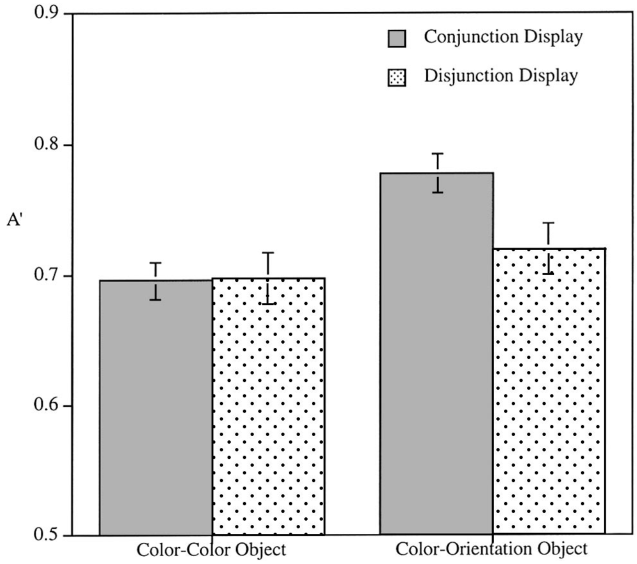

Figure 6. Results of Experiment 3 for monitoring color–color conjunction and disjunction displays, averaged over cap- and stem-color change detection for the color–color mushroom objects, and for monitoring color–orientation conjunction and disjunction displays, averaged over cap-color change detection and stem-orientation change detection for the color–orientation mushroom objects. Error bars represent standard errors of the means.

In both Experiments 1 and 2, performance dropped significantly for monitoring one versus two features. One may argue that when both features were monitored, a false alarm on either feature would lead to an incorrect response. As such, even if both features were monitored exactly as accurately as they were when only one feature was monitored, the actual decision would have more errors when both features were monitored (e.g., Duncan, 1980).

Given that the main comparisons in Experiments 1–3 were made between conjunction and disjunction displays, and because any drop in performance caused by decision noise would have been equal for both conjunction and disjunction displays, the main results of the present study—absence (presence) of object-based effect for features of the same (different) dimension(s)—should not have been different even if decision noise were properly controlled for. Nevertheless, I thought it would still be interesting to see whether performance would drop at all when decision noise was controlled for, especially for objects containing a color and an orientation feature. Moreover, given the results of Experiment 3, for color–color conjunction objects and color–orientation conjunction objects, one would expect a significant interaction between monitoring condition (monitoring one vs. both features) and object type, such that the drop in performance for monitoring two features versus one feature would be smaller for the color–orientation conjunction displays than for the color–color conjunction displays.

Experiment 4

This experiment studied the effect of decision noise on the orientation–orientation conjunction displays used in Experiment 2.

Method

Participants. Six participants (3 women and 3 men) were recruited from the same participant pool as in the preceding experiments. They were paid for their participation.

Materials and design. The same orientation–orientation conjunction display used in Experiment 2 was used in Experiment 4. All aspects of the design and procedure were identical to those of Experiment 2 except that in the testing phase a probe question ("Green Bar Orientation?" or "Red Bar Orientation?") appeared at the bottom of the screen with the test display to instruct participants to make a change-detection judgment on the probed feature only. When only one feature was monitored (either the red- or green-bar orientations), the probe was the same for all the trials; when both red- and green-bar orientations were monitored, the probe instructed participants to make a change-detection judgment on the green-bar orientation in half of the trials and on the red-bar orientation in the other half of the trials. This manipulation not only decreased decision noise at the testing phase but also enabled separate false-alarm rates to be measured for green- and red-bar change detections, which made it possible to calculate separate A′ scores for green- and red-bar change detections when both features were monitored. Trials were blocked by monitoring condition, and the order of presentation was counterbalanced across participants. The experiment lasted about 20 min.

Results and Discussion

As in preceding experiments, A′ was calculated for each participant in each condition. There was an overall significant drop in performance for monitoring two features versus one feature, $F(1, 5) = 11.31$, $p < .05$. No other effect or interaction reached significance. The results are plotted in Figure 7.

The data showed that with decision noise properly controlled for, there was still a significant drop in performance for monitoring two features versus one feature, indicating that the drop in performance observed in Experiment 2 was not caused purely by decision noise but that there was some genuine cost in maintaining both features of an orientation–orientation object compared with maintaining just one feature of such an object.

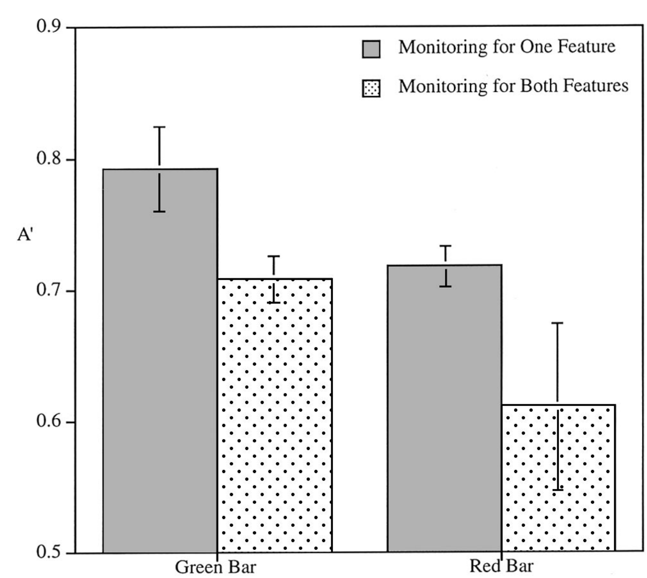

Figure 7. Results of Experiment 4 for monitoring orientation–orientation conjunction displays. Error bars represent standard errors of the means.

Experiment 5

The color–color conjunction display and the color–orientation conjunction display used in Experiment 3 were used again in Experiment 5. Participants either monitored one of the two features or both features of the display.

Method

Participants. Twelve participants (5 women and 7 men) were recruited from the same participant pool as in the preceding experiments. They were paid for their participation.

Materials and design. The design of this experiment was a combination of the designs used in Experiments 3 and 4. The same color–color conjunction displays and color–orientation conjunction displays, as well as the same color and orientation assignments for each object in the display as described in Experiment 3 were used here. As in Experiment 4, either one or both features of each display were monitored. A probe question appeared at the bottom of the test display to instruct participants which one of the two features they should make a change-detection judgment on. For color–color conjunction displays, the probe question was either "Cap Color?" or "Stem Color?"; for color–orientation conjunction displays, the probe question was either "Cap Color?" or "Stem Orientation?" As in Experiment 4, when only one feature was monitored, the probe was the same for all the trials; when both features were monitored, the probe instructed participants to make a change-detection judgment on one of the features in half of the trials and on the other feature in the other half of the trials.

As in Experiment 4, when a single feature was monitored, there were 64 trials for each feature of each display type, with 32 change trials and 32 no-change trials distributed evenly in two 32-trial blocks. When both features were monitored, there were a total of 128 trials for each display type, with 32 change trials and 32 no-change trials for each of the two features, distributed evenly in four 32-trial blocks. The two or four blocks for a given monitoring condition and a given display type were always presented together. The order of presentation was counterbalanced across participants. All other aspects of the design and procedure were identical to those of Experiment 4. The experiment lasted about 40 min.

Results and Discussion

As in preceding experiments, A′ was calculated for each participant in each condition. There was an significant drop in performance for monitoring two features versus one feature: $F(1, 11) = 74.54$, $p < .001$, for the overall effect; $F(1, 11) = 47.41$, $p < .001$, for color–color conjunction displays; and $F(1, 11) = 34.05$, $p < .001$, for color–orientation conjunction displays. The two object types (color–color objects vs. color–orientation objects) differed from each other significantly, $F(1, 11) = 54.14$, $p < .001$, such than overall performance was higher for the color–orientation objects than it was for the color–color objects. There was also a significant interaction between object type and monitoring condition, $F(1, 11) = 6.64$, $p < .05$. The results are plotted in Figure 8.

As in Experiment 4, the results showed that with decision noise properly controlled for, there was still a significant drop in performance for monitoring two features versus one feature for both object types. Moreover, the significant interaction between object type and monitoring condition indicates that features of an object were better retained when they were from different feature dimensions than when they were from the same dimension, thus confirming the results obtained in Experiment 3.

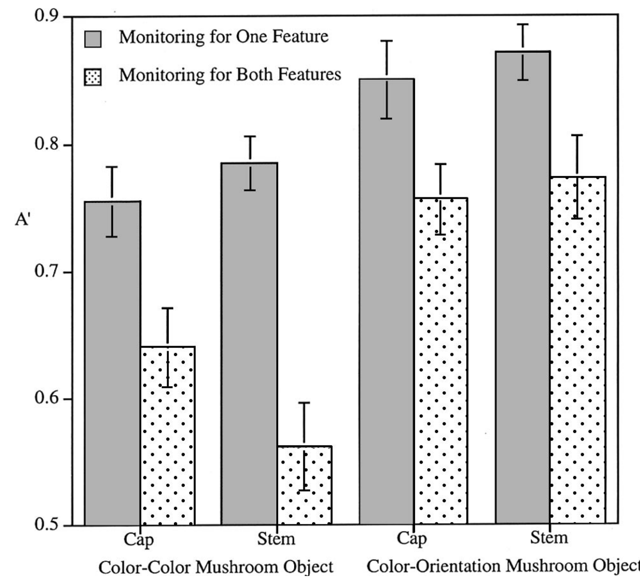

Figure 8. Results of Experiment 5 for monitoring color–color and color–orientation conjunction displays. Error bars represent standard errors of the means.

General Discussion

In the present study, the question of whether object-based feature encoding exists when two features of an object come from the same feature dimensions was investigated. In Experiment 1, bicolored mushroom-like objects were used; for each mushroom-like object, the cap carried one color feature and the stem carried the other. In Experiment 2, objects with two orientations were used; each object consisted of a red bar of one orientation and a green bar of a different orientation. In both Experiments 1 and 2, the two features of the same dimension either appeared on different parts of the same object (conjunction display) or on different objects (disjunction display). Participants were instructed to monitor just one of the two features or to monitor both features. The size of the interaction between display type and monitoring condition was used to measure the degree of object-based encoding for these features. If object-based encoding advantage exists for features of the same dimension when they are from different parts of the same object, there should be a significant interaction between display type and monitoring condition such that the amount of performance drop for monitoring two features rather than one should be much smaller for the conjunction displays than for the disjunction displays.

In both Experiments 1 and 2, for conjunction displays, performance was better when only one feature was monitored compared with when both features were monitored, indicating that it was not the case that both features of the same feature dimension were always encoded together regardless of the task demand. However, no significant interaction between display type and monitoring condition was found in either Experiment 1 or Experiment 2. In fact, in both experiments, F was less than 1 for this interaction. Therefore, unlike my previous study (Xu, 2000a), in which object-based encoding advantage was found for color and orientation features located on different parts of the same object, the present study failed to find any such evidence for two features of an object that are from the same feature dimension, whether it be color or orientation.

To test the possibility that the two-part mushroom-like objects used in Experiment 1 were less integral than the objects I used previously (Xu, 2000a), in Experiment 3 I used the mushroom-like objects to test object-based feature encoding for color and orientation features from different parts of an object. This time, the mushroom cap carried the color feature, and the bend of the mushroom stem carried the orientation feature. The bicolored mushroom-like objects from Experiment 1 were also used in an effort to replicate the findings of that experiment. The same group of participants and similar stimuli were used in both Part A and Part B. Although significant object-based encoding advantage was found for color and orientation features located on different parts of the same object, thus replicating the results of my previous study (Xu, 2000a), no such effect was found for the two color features located on different parts of the same object, thus confirming the findings of Experiment 1.

In Experiment 4, with decision noise properly controlled for, there was still a significant drop in performance for monitoring two features versus one feature of an orientation–orientation object, indicating that the drop in performance observed in Experiment 2 was not caused purely by decision noise but that there was some genuine cost in maintaining both features of an orientation–orientation object compared with maintaining just one feature of such an object. Similarly, in Experiment 5, when decision noise was properly controlled for, there was still a significant drop in performance for monitoring two features versus one feature for both color–color objects and color–orientation objects. Moreover, there was a significant interaction between object type and monitoring condition indicating that features of an object were better retained when they were from different feature dimensions than when they were from the same dimension, thus confirming the results of Experiment 3.

The results of the present experiments therefore indicate that object-based feature encoding exists for two features of an object only when they are from different feature dimensions.[5] Going back to Table 1, the answer to the discrepancy in cell B should now be that there is no object-based performance benefit when two features on the same dimension are to be reported.

In a study employing a somewhat similar logic with spatial frequency and contrast gratings, Magnussen and Greenlee (1997) measured discrimination thresholds for spatial frequency and contrast tested separately and tested together (dual discrimination), and with two contrast or two frequency components tested together. The components either overlapped, formed a compound grating, or were presented as two gratings side by side. Under dual-discrimination conditions, participants were instructed to monitor both components, although only one of the components was changed at a time. The authors found that when dual-discrimination involved two features from different feature dimensions (frequency and contrast), the discrimination threshold increased by a factor of 1.7, as predicted by a model of stimulus uncertainty for orthogonal dimensions; however, when dual discrimination involved two features from the same feature dimension (e.g., two contrasts), the discrimination threshold increased by a factor of 3–6, regardless of whether the two features were presented in a compound grating or as side-by-side gratings. This latter result is consistent with the present findings.[6]

In visual search, Wolfe and colleagues found that color × color and orientation × orientation conjunction searches were extremely inefficient (Bilsky & Wolfe, 1995; Wolfe et al., 1990). Findings from the present study may explain why this was the case: Two colors or two orientations of an object cannot be integrated as readily as can a color and an orientation feature.[7]

Further research is needed to understand why object-based feature encoding only occurs for features of an object when they come from different feature dimensions. Features of the same dimension may be registered and retained in a similar manner, for example, in the same feature map (Treisman, 1988; Treisman & Gelade, 1980). The closer two features of the same dimension are located in the map, the greater the amount of interference there may be in encoding these features. As a result, any advantage gained by object-based encoding is erased by interference between these two features. This account is reminiscent of the analyzer theory proposed by Allport (1971; Wing & Allport, 1972). It is also possible that only a fixed number of features can be retained on the same feature dimension regardless of the spatial locations of these features—whether they are adjacent to each other as features of the same object or far apart from each other as features of different objects. As a result, no object-based benefit can be observed for these features. At the neuronal level, it is possible that the specific wiring of the neurons or the specific mechanism used for binding (e.g., synchronized neural firing; Vogel et al., 2001; see also Gray, König, Engel, & Singer, 1989; Hummel & Holyoak, 1997) allows only features of different dimensions to be integrated.

[5] In the present study, a concurrent verbal task was not included in the change-detection paradigm. As such, one may argue that the present results may be distorted by the use of verbal memory. It has been shown, however, that change-detection performance was largely unaffected by the presence of a concurrent verbal load (Vogel et al., 2001, Experiments 1 and 2). Moreover, given that the same number of features was present in both the conjunction and disjunction displays in Experiments 1, 2, 3, and 5, even if verbal encoding might have been possible, it could not explain the key finding of the present study that object-based encoding exists for two features of an object only when they come from different feature dimensions.

[6] This result of Magnussen and Greenlee (1997) may not be comparable to the present results because these authors failed to obtain any object-based benefit even for two features of different dimensions. Discrimination thresholds were about the same whether the two features were presented in a compound grating (as parts of the same object) or as side-by-side gratings (as different objects). This discrepancy in results might have resulted from the threshold discrimination method used by Magnussen and Greenlee. This procedure may have tapped into processes other than simple VSTM object encoding.

[7] There are, however, differences between results obtained from visual search studies and those from VSTM studies. Wolfe, Friedman-Hill, and Bilsky (1994) found that search for color × color conjunction became more efficient when one color surrounded the other color compared with when two colors were presented side by side. Wolfe et al. (1994) concluded that “surroundedness contributes to the distinction between parts and wholes” (p. 545) and determines efficiency in visual search. This result, however, was not replicated in VSTM change-detection tasks. Wheeler and Treisman (2000) found an absence of object-based encoding in VSTM regardless of whether two colors were presented side by side or with one surrounding the other. These findings suggest that strategies available in visual search may not be applicable to VSTM change-detection tasks.

In summary, the present results impose a significant constraint on object-based encoding theory by showing that although object-based encoding exists for two features of an object when they come from different feature dimensions, such an object-based benefit is absent for two features of an object when they come from the same feature dimension.

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.

Bilsky, A. B., & Wolfe, J. M. (1995). Part-whole information is useful in visual search for size × size but not orientation × orientation conjunctions. *Perception & Psychophysics, 57,* 749–760.

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.

Duncan, J. (1993). Similarity between concurrent visual discriminations: Dimensions and objects. *Perception & Psychophysics, 54,* 425–430.

Duncan, J., & Nimmo-Smith, I. (1996). Objects and attributes in divided attention: Surface and boundary systems. *Perception & Psychophysics, 58,* 1076–1084.

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

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

Hummel, J. E., & Holyoak, K. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427–466.

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.

Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Perception & Psychophysics, 61,* 1241–1255.

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

Magnussen, S., & Greenlee, M. W. (1997). Competition and sharing of processing resources in visual discrimination. *Journal Experimental Psychology: Human Perception and Performance, 23,* 1603–1616.

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

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.

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

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. (1969). Strategies and models of selective attention. *Psychological Review, 76,* 282–299.

Treisman, A. (1988). Features and objects: The 14th Barlett Memorial Lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40*(A), 201–237.

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

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

Wheeler, M. E., & Treisman, A. M. (1999, November). *Aspects of time, space and binding in visual working memory for simple objects.* Paper presented at the Seventh Annual Workshop on Object Perception and Memory, Los Angeles, CA.

Wheeler, M. E., & Treisman, A. M. (2000). *Binding in short-term visual memory.* Manuscript submitted for publication.

Wing, A., & Allport, D. A. (1972). Multidimensional encoding of visual form. *Perception & Psychophysics, 12,* 474–476.

Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics, 55,* 537–550.

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color × color and orientation × orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 879–892.

Xu, Y. (2000a). *Integrating color and shape in visual short-term memory for objects with parts.* Manuscript submitted for publication.

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

Xu, Y., & Potter, M. C. (2000). *Capacity and form of representation in visual short-term memory.* Manuscript submitted for publication.

Received August 14, 2000
Revision received July 30, 2001
Accepted August 16, 2001 ■