

The capacity of visual short-term memory is not a fixed number of objects

GREG DAVIS

University of Cambridge, Cambridge, England

and

AMANDA HOLMES

Roehampton University, London, England

Luck and Vogel (1997) have reported several striking results in support of the view that visual short-term memory (VSTM) has a fixed capacity of four objects, irrespective of how many relevant features those objects comprise. However, more recent studies have challenged this account, indicating only a weak effect of the number of objects once other factors are more evenly equated across conditions. Here, we employed a symmetry manipulation to verify object segmentation in our displays, to demonstrate that when spatial and masking factors are held constant, the number of objects per se has no effect on VSTM. Instead, VSTM capacity may reflect the number of object “parts” or feature conjunctions in a given display.

Visual short-term memory (VSTM) refers to the ability of observers to store a few visual items over brief periods of time, independently of verbal processing. This distinction between VSTM and verbal short-term memory was recognized in Baddeley and Hitch’s (1974) original model of working memory (embodied in the “visuospatial scratchpad” and “phonological loop” components, respectively), and subsequent evidence has provided support for segregation of these two systems in the human brain. In addition, VSTM is differentiated from “iconic” memory along several dimensions. Whereas VSTM has a very limited capacity, is relatively immune to masking and can last for seconds, iconic memory is thought to have a far larger capacity, to be susceptible to perceptual masking, and to last for around 300 msec.

Here, we attempt to resolve the recent controversy over the units of VSTM capacity. Several recent studies have characterized the capacity limit of VSTM as three or four “object” representations. This limit is assumed to hold independently of how many relevant features (e.g., color, shape, and texture) these objects comprise. Perhaps the most compelling of these findings were reported by Luck and Vogel (1997), who employed a change-detection task to examine VSTM for several briefly, simultaneously presented objects. They found that when observers were required to remember only the orientations of four lines, or only their colors, they could accu-

rately report a change in one of the single features for four objects. In a third task, in which observers were required to remember *both* the color and orientation of objects, they could accurately report a change in any of these features for four objects. In this last condition, observers could store eight features pertaining to four objects as accurately as they could four features belonging to four objects. That is, the capacity of VSTM was restricted to a given number of objects (three or four), with a null effect of the overall number of individual features.

Spatial Factors in VSTM

In the majority of Luck and Vogel’s (1997) tasks, whenever two or more features belonged to the same object, they also arose at precisely the same location. In such cases, the role of the number of objects versus the number of locations cannot be distinguished (this potential problem also applies to Olson & Jiang’s [2002] and Wheeler & Treisman’s [2002] studies where an apparent object effect was found). However, in one crucial study, Luck and Vogel examined VSTM for bicolored squares (center, one color; surround, another) and found that the colors of four such squares (eight features at eight locations) could be stored as efficiently as four squares that had just one color each. Such a finding would suggest that eight color features all belonging to the same feature dimension (color) can be stored if they belong to four objects, providing apparent evidence for the view that VSTM capacity is indexed in object units rather than locations or features. However, Wheeler and Treisman (and Olson & Jiang, 2002) have failed to replicate this finding. Employing less salient colors than those in Luck and Vogel’s displays, they found that four bicolored squares (eight features) were stored much less efficiently than were

This work was funded by a joint E.P.S.R.C. (U.K.) and Ministry of Defence (U.K.) grant. We thank Robert L. Greene, Steven Luck, and Jacob Waldman-Khaira for comments on an earlier draft. Correspondence concerning this article should be addressed to G. Davis, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England (e-mail: gjd1000@hermes.cam.ac.uk).

four squares of just one color. Thus, VSTM could not be indexed purely in terms of the number of objects in those displays.

In order to account for this failure to replicate, the authors of these new studies suggested that in Luck and Vogel's (1997) task, observers did not need to code both colors of each bicolored square to perform the task. Rather, they could simply encode the color contrast between the center color and the surround color in each object (one feature). This account seems plausible because color contrast between adjacent regions is the primary source of color coding throughout early stages of vision (e.g., De Valois, 1965). Such interactions can only have a significant effect because the two colored regions comprised by each "object" in these tasks *abut* each other directly.

Lee and Chun (2001) have also examined spatial versus object factors in six-object displays comprising three pairs of overlapping objects (akin to Duncan's [1984] line and box displays) versus six objects appearing at six entirely separate locations. They found equivalent performance in the features-near (overlapping objects) versus features-far (separated objects) conditions and concluded that there was no effect of spatial separation on performance. However, the use of overlapping objects may have adversely affected performance in the features-near condition, thereby masking what would have been an effect of spatial separation. Despite ample evidence that either one of two overlapping objects can be easily attended (e.g., Moore, Yantis, & Vaughan, 1998), when two overlapping objects must *both* be attended, we suggest that this may be a more difficult task. Essentially, such a phenomenon may be similar to that referred to by Rubin (1921) when describing the difficulty of attending both figure and ground of figure-ground displays (which are perceived as an object overlapping in front of a background object). Lee and Chun's results may therefore underestimate the role of spatial factors in VSTM.

Xu (2002a, 2002b), taking a different approach, found a weak number-of-objects effect upon VSTM even when different features belonging to the same object did not appear at exactly the same location. In several studies, Xu employed mushroom-shaped stimuli, each comprising two elements—a colored semicircle and either a colored straight bar or a black bar with a bend in the end furthest from the semicircle element. Figure 1A illustrates a typical stimulus from these conditions. Performance on the mushroom stimuli was then compared with performance on similar displays comprising the same number of elements (i.e., the same number of semicircles and bars) but now spatially separated from each other (see Figure 1B). In the latter case, the number of objects was effectively doubled relative to the former condition, whereas the number of features to be encoded remained the same. Xu (2002b) found slightly better performance in the former connected conditions (Figure 1A) than in the latter unconnected conditions (Figure 1B), attributing this difference to the number of objects in the task.

However, whereas the *global* spatial extent of the displays was equated in Xu's (2002b) connected versus unconnected displays, *local* proximity factors differed between the two types of display. Individual pairs of features were nearer to each other in the connected conditions (e.g., Figure 1A) than in the unconnected conditions (Figure 1B). To illustrate the importance of such local spatial factors, consider any pair of elements in Figures 1A and 1B. When faced with a display that exceeds the capacity of VSTM (as do most natural scenes and many experimental displays), mechanisms of visual attention may prioritize just a few of the objects displayed for storage in VSTM. In such cases, the attended and encoded features will often fall within a smaller surface area in Figure 1A than in Figure 1B, due to the greater distances between features in those displays.

Amodal Completion and VSTM

One recent study has controlled for all spatial factors and still found a number-of-objects effect upon VSTM performance. Walker and Davies (2003) employed amodal shape completion in order to manipulate object segmentation while holding spatial relations constant. When pairs of colored bars in their displays appeared to form two sections of a single, long, partially occluded bar (effectively halving the number of objects in a display), Walker and Davies found that VSTM performance was superior to when the same colored bars were perceived to form separate objects. Note that spatial separation of the features, even in terms of local spatial factors, was identical across conditions and could not, therefore, account for Walker and Davies's findings. Amodal shape completion would, therefore, appear to provide a powerful means of manipulating object segmentation while holding spatial factors constant—a manipulation also employed in the present studies.

One interesting aspect of Walker and Davies's (2003) findings was that they only found a number-of-objects effect for pairs of bars that were of the *same* color as each other. When each bar in a pair of bars was of a different color, they found no number-of-objects effect. This absence of a number-of-objects effect in the latter case may reflect one of two primary factors. First, it may simply be that in these conditions, object segmentation had been impoverished. When bars were of two different colors, this may have caused them to be perceived as two separate objects rather than as a single object, irrespective of condition in Walker and Davies's studies. In such a case, the number of objects perceived would have been identical in the various conditions, and no effects would be expected. Walker and Davies explicitly assume that this was the case (see the Discussion section of Experiment 1; Walker & Davies, 2003). Alternatively, it may be that the two differently colored bars were perceived to form parts of the same object, but that two differently colored bars, even when they appeared to form sections of the same long, partially occluded object, could not be coded as a single unit of VSTM capacity. Each of these candidate

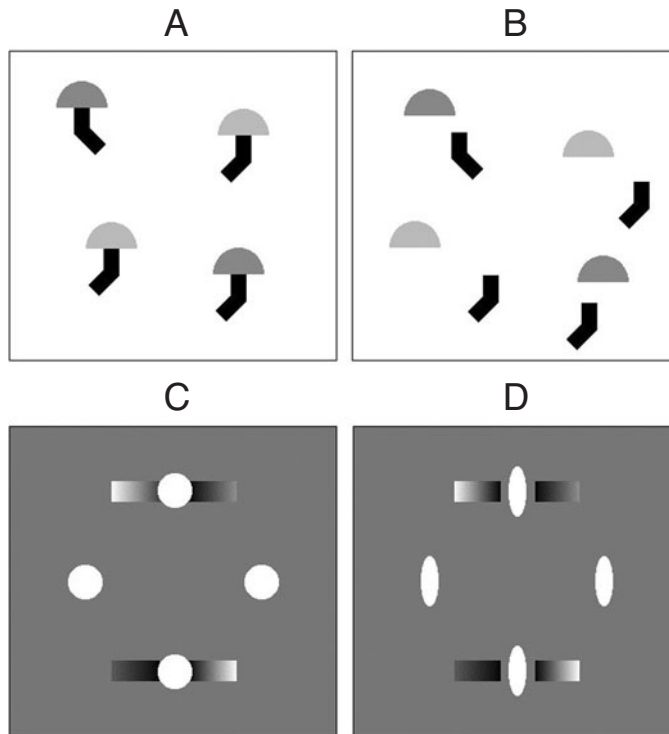


Figure 1. (A and B) Two displays based on verbal description of Xu's (2002a, 2002b) stimuli, in which overall "global" proximity of items is equated in the two cases. That is, an attention "spotlight" of the same extent would suffice to cover either display. However, the two displays differ in terms of "local" proximity. Pairs of features in panel A are nearer to each other than those in panel B. (C) Typical target display from a connected condition in Experiment 1. Note that the four target features (three different colors—red, green, and blue—indicated here by gray levels) appear to form parts of only two objects. (D) Typical target display from an unconnected condition in Experiment 1. Four target features identical to those in panel C now appear to form four separate objects.

explanations has a very different set of implications for our understanding of VSTM.

If the absence of an object segmentation effect for differently colored bars arose simply due to a poverty of object grouping in Walker and Davies's (2003) displays, this would leave open the possibility that the number of objects in a display may affect VSTM performance independently of other variables. In such a case, were object segmentation to be refined in the task, one might yet find evidence of a number-of-objects effect upon VSTM for differently colored bars. Walker and Davies concluded that such was the case and indeed exploited the supposedly weaker object segmentation in those displays as to provide a control condition for displays in which features perceived to belong to the same object were of the same color. However, if the absence of object effects were not due to impoverished object-segmentation cues for differently colored bars, this would point to the following conclusions.

First, because a number-of-objects effect *was* evident for pairs of bars that were of the *same* color, it would ap-

pear that a single, long, partially occluded bar, made up of two separate small bars of the same color, can form a single unit of VSTM capacity. Note that in essence, such units comprise only one feature conjunction—between one orientation (of the bar as a whole) and one color (of both bars). Hence, it may be that a single-feature conjunction can form a unit of VSTM capacity. Such findings are similar to those of Luck and Vogel (1997) and other subsequent studies. Similarly, it may be, as Vogel, Woodman, and Luck (2001) and Xu (2002a) have suggested, that each object "part" (e.g., defined by local minima; see Hoffman & Richards, 1984) may form a single unit of VSTM capacity. Both of these approaches can account for Walker and Davies's (2003) data but suggest that simple combinations of features (object parts or single-feature conjunctions) *can* form units of VSTM.

Second, when two *differently* colored segments appear to form a single object in Walker and Davies's (2003) studies (also in the bicolored squares conditions of Luck & Vogel, 1997; Wheeler & Treisman, 2002), they appear *not* to behave as a single unit of VSTM capacity. If such

a finding is not due to impoverished object segmentation in those displays, it would suggest that whereas single-feature conjunctions (comprising just one color and one orientation feature) or object parts *can* form units of VSTM capacity, more complex object representations *cannot*.

In six new VSTM experiments (and two other control studies) reported here, we employed amodal completion to examine potential number-of-objects effects on VSTM for objects comprising more information than a single-feature conjunction or one part, holding spatial factors constant. To anticipate our conclusions, we found no hint of an effect of the number of objects stored upon VSTM under these conditions. While this finding is a null result, we *do* find significant effects of the same object segmentation upon VSTM when features can form symmetrical objects. This finding provides independent verification of the quality of object segmentation in our task for the purposes of VSTM.

The difference between showing a small but significant number-of-objects effect (as other authors have found—e.g., Xu, 2002a), versus no hint of one, is important. If, once spatial factors are controlled for, there were to be both a small effect of the number of objects *and* of the number of features, these results might be incorporated into a hierarchical multiple-resources model (see Olson & Jiang, 2002) in which the number of objects was still a major determinant of VSTM capacity. In such a scheme, VSTM capacity would be jointly determined by the number of objects at many different levels of representation. At one level of representation, an object may refer to a single feature, at another level, a simple feature conjunction, and at higher levels, a more sophisticated representation. Such an explanation depends on a flexible concept of the term *object* and seems intuitively plausible. However, a finding of absolutely *no* effect of the number of objects would be incompatible with a hierarchical view of this type; in the hierarchical view, a manipulation of object representations on any level of the hierarchy should have a small effect on VSTM.

In the present studies, amodal shape completion was employed to manipulate object segmentation independently of spatial and masking factors. To illustrate, consider a typical display from a connected trial from Experiment 1 in Figure 1C. The two color features at the top of the display (indicated by different gray levels—actual features could be red, green, or blue) form *physically* separate fragments that are separated by a white circle. Nonetheless, because these regions both abut the circle and are aligned with each other, they are perceived to form a single, partially occluded bar, such that both colored regions belong to the same bar. In contrast, two fragments from a typical unconnected display in Figure 1D (identical to the fragments at the top of Figure 1C) are separated by a taller, thinner ellipse. Because the fragments do not abut the ellipse, they are not amodally complete and appear as separate objects. Note that spatial factors are precisely equated (locally and globally) across

these two display types. In addition, because the two color features do not directly abut each other, observers cannot code-color contrast at the border between two colors as a strategy for halving the number of features they must store in VSTM.

Another characteristic of the color features in these studies was that the color information changed gradually to black as it neared the white ellipse or circle (see Figures 1C and 1D). This was to ensure that each fragment had a common color (black) where it abutted the white circle in the connected conditions (Figure 1C), maximizing grouping of those elements to form single, partially occluded objects. In contrast to our approach, Walker and Davies's (2003) colored bars were of one uniform color, including the regions where they abutted the occluding surface. Potentially, in their study (as the authors themselves assumed to be the case) the absence of grouping by common color in some conditions may have weakened perceived object segmentation. We sought to avoid this potential problem by employing graded colors.

In addition to color features, we employed shape features (Figures A and 2B) and combinations of color and shape features (Figures 2C and 2D) in the present experiments, to provide a thorough investigation of potential number-of-objects effects under these new conditions. The basic logic of the experiments described here was as follows. In the connected conditions (Figure 1C), the four colored regions in each display form parts of only *two* objects, whereas in the unconnected conditions the same features formed *four* separate objects (Figure 1D). Accordingly, connected trials always comprised half the number of objects of corresponding unconnected trials. We were thus able to manipulate the number of objects independently of the number of features, the amount of masking, and spatial factors.

EXPERIMENTS 1–3 Four Target Features

Method

Observers. Twenty-four observers, 8 per experiment, were recruited by advertisement and paid £7 each for participating. Twelve were male, and 12 were female, ranging in age from 20 to 32 years (median = 27).

Displays. The figures are all drawn to the same scale. Each white circle in the connected conditions was 26 mm wide. The left circle was separated from the right by 120 mm and the top circle from the bottom circle by 111 mm. A target object in those conditions measured 81–90 mm tip to tip, dependent upon the particular pair of features the object comprised. The display was viewed from a distance of around 50 cm, such that each full object measured a maximum of 9° tip to tip. Stimuli were presented on a Macintosh G3 computer with a 17-in. screen, using PsyScope software.

Procedure. All of the VSTM studies employed the same basic trial structure, typical of previous VSTM studies discussed above. At the beginning of each trial, circles or ellipses (depending upon condition) were presented alone for 500 msec, followed by the first target display (e.g., Figures 1C and 1D) for 200 msec. Next, the circles or ellipses were presented for 900 msec, followed by a second target display (200 msec) and finally the circles or ellipses until response. The observers simply had to indicate as quickly and accu-

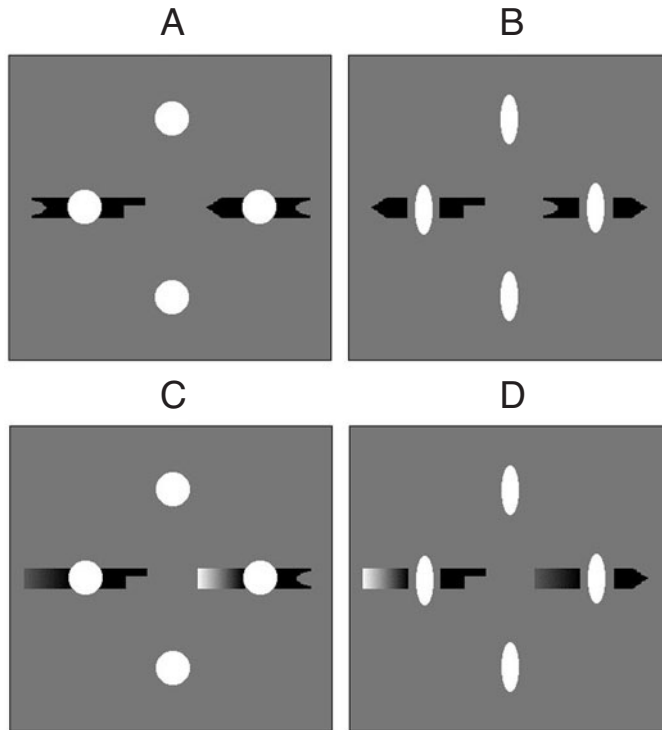


Figure 2. (A and B) Typical target displays from Experiments 2 and 3. Connected and unconnected displays, respectively, from Experiment 2, in which target features were all shape elements. (C and D) Connected and unconnected displays, respectively, from Experiment 3, in which half the target features were shape elements and half were color elements.

rately as possible (with primary emphasis on accuracy) whether or not any element had changed between the first and second target displays. On half the trials, the first and second target displays were identical, and on the other half, one of the features changed. Change and no-change trials were randomly intermixed. Each observer (8 observers per study) performed 12 blocks of 60 trials. Connected and unconnected trials were blocked, with half the observers run on connected trials in even-numbered blocks, unconnected trials on odd-numbered blocks, and vice versa for the other observers.

Figures 1C and 1D illustrate target displays for Experiment 1. Only color features were presented, and two pairs of these appeared in every target display, either at the top and bottom positions (e.g., as in Figure 1C) or the left and right positions (as in Figure 1D), irrespective of whether the trial was connected or unconnected. There were three possible target colors: red, green, and blue (measured using a Minolta CS-100A light spot meter; approximate CIE coordinates, $Y, x, y = 31.8, .616, .355; 57.8, .275, .497; 27.8, .167, 1.57$, respectively). Within one pair of features (appearing on either side of an ellipse) or within a single bar in the connected conditions, the two colors were always different, so as to preclude any possible effects of symmetry. Any such effects could, in principle, have given rise to selective advantages for the connected condition. Indeed, in a later experiment we see that they do, in the case of shape features.

Figures 2A and 2B illustrate target displays from Experiment 2, which were identical to Experiment 1 except that the red, green, and blue color features were now replaced with three possible shape features (all of which are present in the figures). Again, no symmetry of features within each pair of neighboring features was permitted. Figures 2C and 2D illustrate target displays from Experiment 3, which was identical to Experiments 1 and 2, except that both color

and shape features were used. Within each pair of features in the unconnected conditions or within each completed object in the connected conditions, there was a color feature on one side (the left in Figures 2C and 2D) and a shape feature on the other side. Half the observers saw color features presented for the first six blocks on the left of each pair of features, shape features on the right, and then color features on the right, shape features on the left for the remaining six blocks. For the remaining observers, this order of presentation was reversed.

In Experiments 1–3, we thus measured the effects of binding pairs of features into single-object representations by means of amodal shape completion. When such binding arose (connected conditions), the four features in each display belonged to only two objects, whereas when no binding arose (as independently verified in a later experiment), the four features belonged to four separate objects. We were thus able to manipulate the number of objects to be remembered while holding constant the number of features and spatial/masking factors. Three conditions were tested in the three separate experiments: Experiment 1—color–color conjunctions (binding pairs of color features), Experiment 2—shape–shape conjunctions (binding pairs of shape features), and Experiment 3—shape–color conjunctions (binding shape features to color features).

Results

Figure 3A graphs percentage errors for connected and unconnected conditions in Experiments 1–3 (all results are grouped into one figure for purposes of economy and easy comparison of results from the different experiments). Experiment 1 (color features only) results are in-

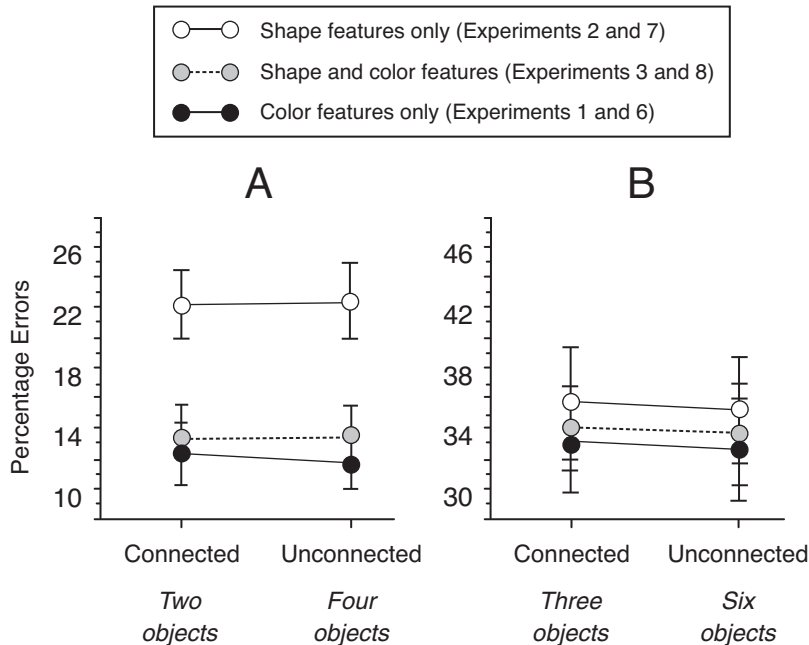


Figure 3. Percentage error rates in Experiments 1–3 and 6–8. Error bars represent *SEMs*.

indicated by black symbols and solid lines, Experiment 2 (shape features only) by empty symbols and solid lines, and Experiment 3 (color and shape features) by gray symbols and dashed lines. Note that in the unconnected conditions, the four target features appeared on four separate objects, whereas in the connected conditions, they appeared to form parts of only two objects. It is clear from the graphs that the manipulation did not affect VSTM performance. This was confirmed in three separate two-way, within-subjects analyses of variance (ANOVAs; one on each experiment) with factors of change (change vs. no change) and object-segmentation (connected vs. unconnected). These ANOVAs yielded no significant main effects of connected versus unconnected conditions (all F s < 0.54, n.s.), the largest F ratio being in Experiment 1 (0.53) where the numerical trend favored *unconnected* conditions (where there were more perceptual objects). In Experiment 2 and Experiment 3, there was no interaction between this factor and change versus no change (both F s < 0.43, n.s.), nor in Experiment 1 [$F(1,7) = 3$, n.s.], the larger F ratio in this last case again reflecting numerical, although nonsignificant tendencies favoring *unconnected* conditions. Similar analyses performed on RT (a secondary measure because accuracy was stressed rather than speed) yielded no significant main effects of connected versus unconnected conditions (all F s < 0.37, n.s.) or interactions (all F s < 0.43, n.s.).

Our first three studies suggested that when spatial factors are held exactly constant across conditions, there is no hint of an effect favoring fewer objects in VSTM (at least four vs. two objects). However, several important potential criticisms of these studies must be addressed before we can draw any firm conclusions from them.

First, it may be argued, by proponents of a four-object limit on VSTM (e.g., Cowan, 2001; Luck & Vogel, 1997) that no effect would be expected in these studies because four objects is the maximum number we presented, and this does not exceed the proposed capacity of VSTM. Performance should therefore be at ceiling in both our connected and unconnected conditions. This seems very unlikely because our observers were nowhere near ceiling in terms of accuracy. On average, performance levels indicated that observers could reliably store around three of the four items in the display. However, to test this possibility thoroughly, we also conducted three further studies (Experiments 6–9) in which three pairs of features were presented in each display (six potential objects in the unconnected condition, three objects in the connected condition).

Before we attempted further comparisons, however, we needed to verify the following two aspects of our paradigm. First, we had to ensure that the manipulation of the number of objects by the presence or absence of perceptual completion had been effective. This was explored in Experiment 5, in which we replicated the conditions of Experiment 2 (shape features only), but with one crucial difference. Whereas in Experiment 2, symmetry within a pair of features had been explicitly prevented, in Experiment 5, it was now permitted. We expected that given previous findings in the literature (e.g., Baylis & Driver, 2001), symmetry should exclusively benefit features belonging to the same object in connected conditions relative to unconnected conditions where features always belonged to separate objects.

Second, we needed to demonstrate that masking factors were equated in connected versus unconnected conditions in Experiments 1–3. The only physical difference

between the connected and unconnected displays was the presence of circles in the connected displays and taller, thinner ellipses in the unconnected displays. It may have been that the circles had greater masking effects than did the ellipses, hiding what otherwise may have been an advantage for the connected displays due to them comprising fewer objects. We therefore conducted three further studies to test the contributions of these two potential concerns.

EXPERIMENT 4 Symmetry Effects

In Experiment 4, we replicated Experiment 2 but permitted symmetry between pairs of features around a white circle or ellipse. As in Experiment 2, there were three shape features: a convex triangular end, a concave circular end, or a horizontal bar end (see Figures 2A and 2B for examples). The first two of these types of shape could now be symmetrical around a circle in the connected trials or around an ellipse in the unconnected trials. These new symmetrical stimuli constituted one third of the stimuli in the new study, with the remaining stimuli being composed of counterbalanced unpredictable combinations of shapes (including the horizontal-bar end features). If manipulation of object segmentation in our displays was effective, this should now yield performance benefits in the connected conditions relative to the unconnected conditions. For trials in which at least one symmetrical pair of features arose, performance should be enhanced in connected conditions relative to unconnected conditions. This prediction held because symmetry has been found to enhance, primarily or exclusively, processing of features from the same object (Baylis & Driver, 2001). Whereas in connected conditions, symmetrical features belonged to the same object (and could benefit from symmetry coding), identical pairs of features in unconnected conditions belonged to separate objects and therefore should not benefit to the same degree.

Method

Observers. Eight new observers were recruited by advertisement and paid £7 each for participating. Four were male, and 4 were female, ranging in age from 19 to 30 years (median = 23).

Displays and Procedure. This was an exact replication of Experiment 2, with the exception that one third of the pairs of features were now symmetrical around the white ellipse or circle that separated them. Only two of the three shape features (as described above) appeared in symmetrical pairs. The observers were informed of the presence of symmetry and advised to employ it in the task.

Results

Trials comprising at least one symmetrical pair of features made up 90% of all trials. Within these trials, the predicted advantage for connected trials (which should benefit from symmetry) over unconnected trials (which should not benefit from symmetry) held—17% versus 20% errors, respectively. This was confirmed in a two-

factor within-subjects ANOVA, with factors of object (connected vs. unconnected) and change (change vs. no change). This yielded a significant main effect of object [$F(1,7) = 13.23, p < .01$], a main effect of change [$F(1,7) = 11.19, p = .0123$], and no interaction between these two factors [$F(1,7) = 0.75, n.s.$].

These new findings provided evidence that pairs of features around a single circle were processed as a single object to a degree where they could benefit from symmetry processing. In contrast, features in unconnected conditions were processed as two objects to a sufficient degree that they did not benefit from symmetry. These findings, therefore, verify that object segmentation was different in the connected versus unconnected conditions.

One final objection to our findings from Experiments 1–3 might be that the physical difference between displays in the connected and unconnected conditions had yielded differential effects of masking in the two conditions. Specifically, it may have been that circles in the connected conditions had greater masking effects than the ellipses in our unconnected conditions, mitigating any advantages conferred by object segmentation. To test this directly, we replicated the conditions of Experiment 3 (color and orientation features), but now removed a small part of the black region of each target feature nearest the white circles or ellipses (see Figures 4A and 4B). This ensured that the target features no longer abutted the white shapes, irrespective of whether the white shapes were circles (as in the connected conditions of the previous studies) or ellipses (as in the unconnected conditions of the previous studies). Therefore, there was no perceptual completion, no manipulation of object segmentation, and in all conditions each feature should now appear to form a separate object from the other target features.

However, while there was no object-segmentation manipulation in this new study, the circles and ellipses should have the same masking effects they had in previous studies. Thus, we could now provide a measure of the relative masking effects of the circles versus ellipses in previous experiments, in the absence of potential object-segmentation effects. If circles in the connected conditions of Experiments 1–3 had elicited greater masking effects than ellipses in the unconnected conditions, this should be evident as an advantage for the displays comprising ellipses in the new study.

EXPERIMENT 5 Masking Properties of Circles Versus Ellipses

Method

Observers. Eight observers were recruited by advertisement and paid £7 each for participating. Two were male, and 6 were female, ranging in age from 19 to 32 years (median = 24).

Displays and Procedure. All aspects of procedure, displays, and design were as for Experiment 3, with the exception of removal of small black regions from each target feature (as described above), removing any manipulation of object segmentation.

Results

A within-subjects two-way ANOVA performed on error scores, with factors of change versus no change and ellipses versus circles found no main effect of ellipses versus circles [$F(1,7) = 0.18$, n.s.], indicating that ellipses and circles had equivalent masking effects. On average, error scores were 17.2% for ellipse trials and 16.6% for circle trials, indicating no hint of an advantage for ellipse trials. On the contrary, any numerical trend was in the *reverse* direction. We conclude that potentially greater masking effects of circles than ellipses cannot explain the absence of an object-segmentation effect in Experiments 1–3.

In Experiments 4 and 5, we therefore presented evidence of the efficacy of object segmentation and the absence of differential masking effects in the connected versus unconnected conditions in Experiments 1–3. It would appear, therefore, that once spatial and other factors are entirely controlled for, there is no effect of object segmentation on VSTM capacity. However, we only tested a maximum of four potential objects in Experiments 1–3, and we therefore sought to extend our conclusions to larger displays in Experiments 6–8.

EXPERIMENTS 6–8

Six Target Features

Experiments 6–8 precisely replicated the conditions of Experiments 1–3 with the following exceptions. First, six target features were presented in each display rather than four: Three pairs of target features appeared at three of the four possible positions in the display, with the other location remaining blank (see Figure 4C for an example of a connected display from Experiment 7). In Experiment 8, the target features in the top and bottom positions of each display were taken from the same set as those in Experiments 1–3, but the left and right positions were taken from a different set. The new colors were CIE coordinates $Y, x, y = 70.9, .489, .412$; $75.3, .333, .253$; $129, .423, .483$, each gradually faded to black as in previous studies; examples of the new shape stimuli can be viewed in the two leftmost columns of Figure 4D. The advantage of presenting features from two independent sets was as follows. On many occasions, two or more features in a display were identical. Arguably in such situations, these features could be perceptually grouped and treated as a single object. The use of two sets of

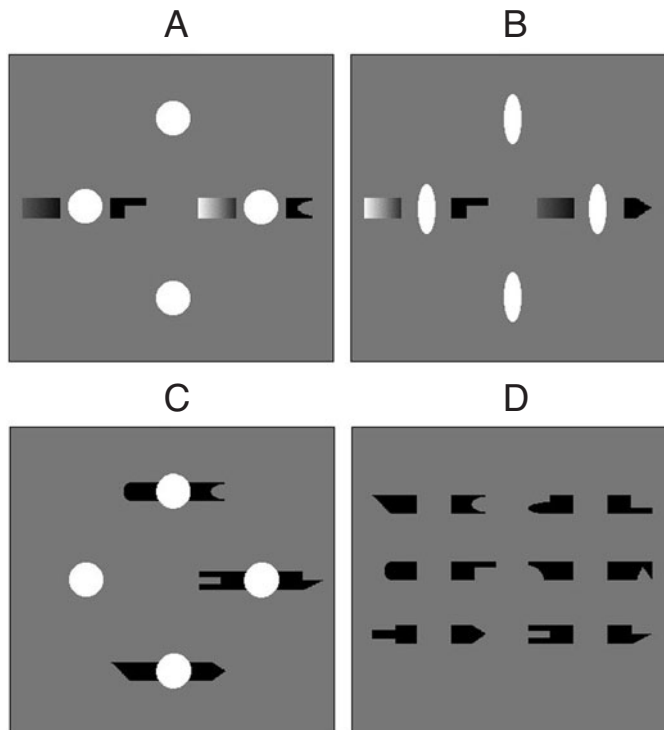


Figure 4. (A and B) Typical target displays from circles and ellipses conditions, respectively, in Experiment 5. (C) Connected display from Experiment 7: Note that the six shape features in the displays appear to form parts of only three objects. In corresponding unconnected conditions, they appear to form six separate objects. (D) Four sets of three possible shape features employed in Experiment 7.

color and shape features, however, ensured that even when such identical features were counted as a single object, six feature displays could comprise a minimum of four objects overall, with the vast majority of trials having five or more objects. Even with this conservative estimate of the number of perceptual objects in a display, therefore, we should expect unconnected displays to comprise five or six objects on the vast majority of trials, in excess of the usual four-object limit supposed by previous authors (e.g., Cowan, 2001; Luck & Vogel, 1997). In contrast, the connected displays should comprise at most three perceptual objects, falling within this proposed limit.

Experiment 6, in which we investigated color features alone, carried this principle further. Now, the top and bottom positions in each display were drawn from two separate sets of three colors each, with the left item in each pair being drawn from the original colors employed in Experiments 1 and 3 and the right feature from a second set of three colors (CIE coordinates $Y, x, y = 70.9, .489, .412; 75.3, .333, .253; 129, .423, .483$). This same division also arose in the left and right positions in the display, where the left feature in each pair could be drawn from one set of colors (CIE coordinates $Y, x, y = 104, .276, .346; 37.6, .277, .173; 106, .349, .398$) and the right feature from another set (CIE coordinates $Y, x, y = 97.8, .350, .530; 109, .310, .333; 100, .363, .337$). Experiment 7, in which we investigated shape features, employed this same strategy, with four sets of three shape features in place of the four sets of three colors described in Experiment 6. These are illustrated in four columns in Figure 4D.

Method

Observers. Twenty-four observers, 8 per experiment, were recruited by advertisement and paid £7 each for participating. Eleven were male, and 13 were female, ranging in age from 19 to 30 years (median = 24).

Displays. The displays and apparatus were the same as in Experiments 1–3, with the exceptions described in the procedure section.

Results

Figure 3B graphs errors for Experiments 6–8 in the same format as Figure 3A does for Experiments 1–3. On average, observers were able to reliably store and retrieve on two of the six items. It is clear even from the graph that there was no effect of the number of objects in a display (six objects in the unconnected conditions, three objects in the connected conditions). Three two-way ANOVAs were performed on the data from each study, yielding no main effects of change versus no change (all $F_s < 0.17$, n.s.). That is, there were no effects of the number of objects (i.e., connected three-object displays versus unconnected six-object displays), with any numerical trend favoring the six-object displays. There was a small but nearly significant interaction [$F(1,7) = 4.36, p = .075$] between this factor in Experiment 8, but this need not affect interpretation of the number-of-objects result. On the contrary, the

small changes in response bias between the connected and unconnected conditions merely underlined the efficacy of the object-segmentation manipulation forming the key difference between the conditions.

DISCUSSION

In the present experiments, we manipulated the number of objects in a display by means of amodal completion. Other than in one example, where symmetry biased performance for connected over unconnected displays, there were no number-of-objects effects in displays comprising four target features (Experiments 1–3) or six target features (Experiments 6–8). We conclude that in our displays, the number of objects to be stored did not affect VSTM performance. More generally, our findings suggest that object representations comprising more than one part or feature conjunction are not the units of VSTM capacity. As discussed in the introduction, our finding of zero effect of the number of objects is also incompatible with an alternative (multiple resources) approach to VSTM where objects at multiple levels of representation jointly determine the capacity of VSTM. Such a possibility has been proposed (but also rejected) by Olson and Jiang (2002).

As discussed in the introduction, one previous study did find significant effects of the number of objects upon VSTM even in experiments that shared crucial features with those described here. Walker and Davies (2003) also employed shape completion to manipulate the number of objects while holding spatial factors constant. They found that binding two identical elements (of the same color and orientation) together into a single occluded object by amodal completion enhanced VSTM performance relative to when the same features were perceived as separate objects. Interestingly, however, Walker and Davies only found a number-of-objects effect when the two elements were of the same color, forming an object that consisted of just one part and just one feature conjunction. In the introduction, we suggested that the absence of a number-of-objects effect in the latter condition could be accounted for in either of the following ways.

First, it may be that the use of different colors had impoverished grouping between the two elements in each object, such that the two elements were not perceived to form parts of the same object. Second, it may be that even when two differently colored features are perceived to form part of the same object, such objects do not form single units of VSTM capacity and cannot enhance performance. In our studies, we attempted to distinguish between these two possibilities by minimizing grouping problems associated with the use of differently colored elements. The colored elements (e.g., Figures 1C and 1D) were all shaded so that each element was the same color (black) where it abutted the white circle in the connected displays. This characteristic should have acted to

maximize perception of elements in the connected conditions as forming large occluded objects. Nonetheless, we found no effects of the number of objects upon VSTM. We conclude that the absence of such an effect in our studies, where neighboring features were always of different colors, is unlikely to have reflected impoverished grouping. Rather, it must have reflected an inability of two differently colored items to form a single unit of VSTM capacity, even when they form parts of the same object.

In Experiments 2 and 5, we also found a parallel to Walker and Davies's (2003) findings, but this time involving shape features. For pairs of identically shaped features (in the symmetrical stimuli of Experiment 5), there was a number-of-objects effect in which connected conditions (fewer objects) yielded superior performance to unconnected conditions (more objects). In contrast, when features within a pair of features were differently shaped, as in Experiment 2, no such number-of-objects effect was found. If these findings are to be accommodated within a number-of-parts account (such as that proposed by Vogel, Woodman, & Luck, 2001), symmetry must be assumed effectively to halve the number of parts comprised by an object.

This conclusion is derived as follows. In order to explain the equivalent performance for connected and unconnected conditions in Experiment 2 on a part-based account, one would need to assume that each amodally completed object in the connected conditions comprised at least two separate parts. Otherwise, if each object in the connected conditions comprised fewer parts than corresponding pairs of objects in the unconnected conditions, superior performance would be predicted on a number-of-parts account. If the objects in Experiment 2 comprised more than one part, those in Experiment 5 (constructed from the same elements as in Experiment 2) should have done so, too. Why, then, was a number-of-objects effect found only in Experiment 5, not in Experiment 2? This discrepancy cannot be explained in terms of the number of parts alone, because this should have been identical in the two experiments. However, if symmetry were to effectively halve the number of part "slots" of VSTM capacity that an object required, this could account for our finding of an effect in Experiment 5, but not in Experiment 2. Accordingly, symmetry in Experiment 5 may have effectively permitted symmetrical objects in the connected conditions (consisting of two identical shape features) to be stored because they comprised just one part, rather than two. However, such efficient coding must not have been possible in the unconnected conditions where symmetrical features pertained to separate objects.

Our results can therefore be accommodated by models in which the number of object parts forms a primary determinant of VSTM capacity. However, we also note that the lack of any significant effect of the number of objects on VSTM is entirely consistent with our earlier, identi-

cal conclusions regarding visual attention. We suggested then (Davis, 2001; Davis, Welch, Holmes, & Shepherd, 2001) that the capacity of visual attention is defined by the number of attended features and the amount of within- and between-object binding that may arise between those features. Such a model may also find application in accounting for VSTM performance, and future research will be required to distinguish this approach from part-based models.

Finally, we should consider two potential criticisms to which the present studies may be vulnerable. First, we have not included any verbal suppression task in the present studies. A few previous studies have included a verbal task in order to ensure that no aspect of performance is governed by verbal short-term memory rather than visual short-term memory. Such a task was not included here because we considered that the displays are presented so rapidly in our paradigm as to render any verbal strategy too slow and inefficient to be of use. Other studies (such as Luck & Vogel, 1997) employing similar stimulus durations have found no effects of verbal load on their tasks, in support of this assumption. Second, it might be argued that there is little incentive in our experiments for observers to attend to objects as a whole, when changes always affect individual features rather than whole objects. To some extent, it must be admitted, this criticism can apply (equally) to all the present and previous studies of object-based effects on VSTM for features. However, it would seem very likely that if attending to whole objects rather than features provided any noticeable benefit, observers would have employed this strategy.

REFERENCES

- BADDELEY, A. D., & HITCH, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.
- BAYLIS, G. C., & DRIVER, J. (2001). Perception of symmetry and repetition within and across visual shapes: Part-descriptions and object-based attention. *Visual Cognition*, *8*, 163-196.
- COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, *24*, 87-185.
- DAVIS, G. (2001). Between-object links and visual attention. *Visual Cognition*, *8*, 411-430.
- DAVIS, G., WELCH, V. L., HOLMES, A., & SHEPHERD, A. J. (2001). Can attention select only a fixed number of objects at a time? *Perception*, *30*, 1227-1248.
- DE VALOIS, R. L. (1965). Analysis and coding of color vision in the primate visual system. *Cold Spring Harbor Symposium on Quantitative Biology*, *30*, 567-579.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501-517.
- HOFFMAN, D. D., & RICHARDS, W. A. (1984). Parts of recognition. *Cognition*, *18*, 65-96.
- 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.

- MOORE, C. M., YANTIS, S., & VAUGHAN, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, 9, 104-110.
- OLSON, I. R., & JIANG, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, 64, 1055-1067.
- RUBIN, E. (1921). *Visuell wahrgenommene Figuren* [Visually perceived figures]. Copenhagen: Gyldendal.
- 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., & DAVIES, S. J. (2003). Perceptual completion and object-based representations in short-term visual memory. *Memory & Cognition*, 31, 746-760.
- WHEELER, M. E., & TREISMAN, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131, 48-64.
- XU, Y. (2002a). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, 64, 1260-1280.
- XU, Y. (2002b). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 458-468.

(Manuscript received November 17, 2003;
revision accepted for publication May 7, 2004.)