Storage of Features, Conjunctions, and Objects in Visual Working Memory

Edward K. Vogel, Geoffrey F. Woodman, and Steven J. Luck
University of Iowa

Working memory can be divided into separate subsystems for verbal and visual information. Although the verbal system has been well characterized, the storage capacity of visual working memory has not yet been established for simple features or for conjunctions of features. The authors demonstrate that it is possible to retain information about only 3-4 colors or orientations in visual working memory at one time. Observers are also able to retain both the color and the orientation of 3-4 objects, indicating that visual working memory stores integrated objects rather than individual features. Indeed, objects defined by a conjunction of four features can be retained in working memory just as well as single-feature objects, allowing many individual features to be retained when distributed across a small number of objects. Thus, the capacity of visual working memory must be understood in terms of integrated objects rather than individual features.

Almost all past and present large-scale models of cognition contain some type of limited-capacity memory system for the short-term retention and manipulation of information (e.g., ACT-R, Anderson, 1993; EPIC, Meyer & Kieras, 1997). This type of memory is thought to underlie our ability to perform a wide range of basic tasks such as assigning semantic identities and syntactic roles to the interdependent words of a complex sentence, understanding the relationships among the ideas within a complex argument, and perceiving the trajectory of a moving object when the visual input is disrupted by blinks and saccades.

Research on short-term memory has a long history, extending back into the early experimental psychology movement of the nineteenth century (Calkins, 1898; Jacobs, 1887). At that time, it was generally termed *immediate* or *primary* memory, and much of the research was aimed at establishing links between the amount of information that could be held in immediate memory (i.e., memory span) and intelligence. In particular, the finding that cognitively impaired children exhibit smaller memory spans than normal children prompted the view that memory capacity could provide a window to intelligence (e.g., Abelson, 1911; Binet & Simon, 1909; Bolton, 1931). Over time, the focus of research shifted away from intelligence testing, yet the broad emphasis on estimating the capacity of immediate memory still remained (for a review, see Blankenship, 1938).

This variety of memory was resurrected during the cognitive revolution and was labeled *short-term memory,* a term that em-

phasized the transient nature of this memory representation and clearly demarcated it from another proposed form of memory, *long-term memory.* Atkinson and Shiffrin (1968) summarized the consensus view of memory in the late 1960s with a theory of memory (which was later called the *modal model* by Murdock, 1974) in which short-term memory served both as a store of immediate information and also as a gateway to long-term storage. However, models of this type were eventually abandoned in the face of several types of incompatible results. For example, the concept of short-term memory implies a fairly rapid, passive decay of information in the absence of rehearsal (as in the classic studies of Brown, 1958; and Peterson & Peterson, 1959), but several studies have indicated that the impaired performance observed at longer delays primarily reflects interference rather than passive decay (e.g., Keppel & Underwood, 1962; Wickens, Born, & Allen, 1963). In addition, neuropsychological studies have demonstrated that brain-damaged patients may have short-term memory deficits without long-term memory deficits, suggesting that information does not need to pass through short-term memory to be stored in long-term memory (Shallice & Warrington, 1970). Similarly, Craik, Gardiner, and Watkins (1970) found that the probability of recalling a word from long-term memory was unrelated to the amount of time that word was stored in short-term memory, which indicates that the storage of information in short-term memory is not sufficient for the creation of a long-term memory. Because of such findings, research on short-term memory gradually declined over the course of the 1970s, leading Crowder to publish a paper in 1982 entitled "The Demise of Short-Term Memory" (Crowder, 1982).

Working Memory

Although the term *short-term memory* fell out of favor, it was still clear that tasks such as solving arithmetic problems and comprehending sentences require some sort of temporary memory system to store the information that is currently being operated on. To add the numbers 2,037 and 7,982 together, for example, one must hold individual digits in an active form as they are being

summed. Research on immediate memory storage thus continued under the rubric of *working memory*, a term that emphasizes the functionality of this type of memory and is agnostic regarding the duration of the representation and its relationship with long-term memory (see Baddeley & Hitch, 1974).

Most of the early models of short-term memory implicitly or explicitly assumed that it was a unitary memory store, irrespective of modality, or admitted that modality-specific stores might exist but did not consider them in any detail (e.g., Atkinson & Shiffrin, 1968). In the 1970s, however, Baddeley and his colleagues proposed a multicomponent model of working memory (WM) in which memory for visual and verbal information are stored separately in different subsystems and are controlled by a central executive (Baddeley, 1986; Baddeley & Hitch, 1974). Early support for this model came from studies in which little or no interference was found when verbal memory tasks were performed concurrently with visual memory tasks, whereas substantial interference was found for two concurrent verbal tasks or two concurrent visualspatial tasks (Baddeley & Hitch, 1974; Baddeley & Lieberman, 1980). Since then, there have been numerous reports supporting the independence of the WM systems for verbal and visualspatial information (for a review see Logie, 1995), and this multicomponent model is currently the most influential and widely accepted model of WM. In fact, recent neuroimaging studies have even begun to subdivide these systems neuroanatomically (Awh et al., 1996; Jonides et al., 1993; Smith et al., 1995). However, the vast majority of the research on WM has focused on the verbal storage subsystem, and most of the existing work on visual WM has focused on memory for spatial information rather than memory for objects. The purpose of the present study was to add to our knowledge of the characteristics of visual WM by assessing its storage capacity for objects defined by simple, suprathreshold features (e.g., color and orientation) and by combinations of these features.

The Storage Capacity of Working Memory

Perhaps the most notable characteristic of WM is that its storage capacity is highly limited. That is, unlike long-term memory, which may store vast amounts of information, only a few pieces of information can be simultaneously held in WM. The precise number of items that can be actively held in WM has been investigated for as long as WM has been studied (Blankenship, 1938). In his well-known review, for example, Miller (1956) noted that the capacity of WM is frequently estimated to be approximately seven items. However, more recent studies have demonstrated that the capacity of verbal WM depends on the phonological length of the words being retained, with longer words consuming more memory capacity than shorter words (Baddeley, Thomson, & Buchanan, 1975; Schweickert & Boruff, 1986). These results indicate that verbal information is held in WM in the form of a phonological code, and it is therefore more appropriate to characterize verbal WM capacity in terms of time rather than the number of items. In this particular case, examining the capacity of verbal WM has not simply provided a benchmark of performance limitations but has helped to elaborate the nature of the representations in memory.

The capacity of visual WM is not nearly as well understood as the capacity of verbal WM. The original measurement of visual WM capacity is often credited to Sperling (1960), who found that when participants were required to report all of the alphanumeric characters in a briefly presented display, they could correctly report only four to five items. However, it is not clear that this estimate of WM capacity provides a pure estimate of the storage capacity of the *visual* subcomponent of WM. For example, the to-be-remembered items used in this study were letters and numbers and were probably coded verbally as well as visually, and performance may have reflected contributions from both verbal and visual WM storage. Moreover, this task required the participants to respond by writing down the characters that they recalled from the display, which necessitated a translation of the visual information into a verbal code. Similarly, an earlier study by Brener (1940) examined the capacity of WM for colors, but the design of this study appeared to encourage the use of verbal WM. Specifically, colored slides were presented serially to participants, and at the end of the list the participants verbally reported the color names in the order they were presented. This study found that about seven colors could be immediately recalled, which is notably close to the prototypical verbal memory span of seven items. More recently, a study that replicated Brener's (1940) design found that the storage capacity for colors in this paradigm is contingent on the pronunciation times for the names of the colors used in the sequence (Schweickert & Boruff, 1986), indicating that memory span in this task is primarily determined by verbal rather than visual WM capacity. Thus, it is crucial to rule out contributions from verbal coding when measuring visual WM capacity.

Another difficulty in using the results from Sperling's (1960) whole-report procedure as an estimate of visual WM capacity is that alphanumeric characters are complex combinations of multiple line segments and may therefore consume more of the available capacity than simpler stimuli. That is, just as verbal WM capacity depends on the number of syllables to be retained, visual WM capacity may depend on the number of elementary features in each object, and visual WM capacity might therefore be much greater for objects that consist of a single feature. Surprisingly, however, no one has systematically assessed the capacity of visual WM for objects defined by simple, suprathreshold features and for objects defined by combinations of these features. The purpose of the present study was therefore to explore visual WM capacity for such objects. More generally, we assume that a systematic exploration of visual WM capacity will yield insights into the sources of limitations in the storage of visual information and the nature of the representation of information in visual WM, just as studies of verbal WM capacity have yielded important insights into the nature of the verbal WM system.

A few previous studies of visual WM capacity have examined the short-term storage of simple features, but these studies have been quite different from studies of verbal WM capacity in that they required participants to perform near-threshold discriminations and typically used a set size of only one item (e.g., Magnussen, Greenlee, & Thomas, 1996; Matin & Drivas, 1979). Just as near-threshold and suprathreshold sensory discriminations may rely on different mechanisms (Graham, 1989), the memory system used to store a single item with great precision may differ from the system used to provide a coarse or categorical storage of multiple

items. Indeed, the one study that examined memory for multiple single-feature items was explicitly designed to minimize categorical representations (Palmer, 1990), and memory performance in this study was enhanced for some participants who managed to code the stimuli categorically. In addition, memory for near-threshold features can be severely impaired by the presentation of a mask even after a delay of several seconds (Magnussen & Greenlee, 1992; Magnussen, Greenlee, Asplund, & Dyrnes, 1991), whereas masks are relatively ineffective for highly discriminable stimuli at such long delays (Breitmeyer, 1984; Vogel, Woodman, Eads, & Luck, 1998). Thus, the storage of near-threshold visual features may be accomplished by a lower-level memory system than the storage of suprathreshold, easily categorized stimuli.

Approaches to Estimating Visual WM Capacity

Historically, the capacity of WM has been assessed by presenting observers with a sequence of items and then requiring the observers to report each item in the correct order. Capacity is then quantified as the number of items that could be correctly reported (Guilford & Dallenbach, 1925; Peatman & Locke, 1934). A visual version of this whole-report procedure was also used to measure visual WM capacity in some of Sperling's (1960) experiments, but as discussed above, this paradigm probably included contributions from verbal working memory. Phillips (1974) developed a different paradigm for assessing visual WM performance that does not rely on verbal report and is therefore less likely to be contaminated by verbal WM (see also Klemmer, 1963). In this procedure, a partially filled grid of dots (called the *sample array*) was briefly presented, and after a delay period in which the screen was blank, a second array (the *test array*) was presented that was either identical to the sample array or differed in the position of one dot. Participants responded by indicating whether the sample and test arrays were the same or different. Visual WM was differentiated from iconic memory by systematically varying both the number of dots in each array as well as the delay period between the two arrays, under the assumption that iconic memory would have a virtually infinite capacity but a short duration. In addition, the test array was sometimes presented at a slightly different retinal location to test the position dependency of the memory. When the sample and test arrays were presented with a short delay (<250 ms) and at the same retinal positions, participants were highly accurate even when the arrays contained a large number of dots. However, when the delay was longer or the arrays were presented at different retinal locations, performance declined systematically as a function of the number of dots in the array. These results suggested a distinction between a short-lived, high-capacity, retinotopically organized iconic memory representation and a longer-lasting, limited-capacity, abstract visual WM representation.

Although this paradigm was a major improvement in terms of isolating visual WM, it is difficult to obtain an unambiguous estimate of visual WM capacity from these results because the individual dots were presented as a part of a complex pattern (although for an interesting developmental approach using this procedure, see Wilson, Scott, & Power, 1987). That is, the visual system might have stored the pattern of dots as a gestalt figure rather than as a set of individual items, and accurate performance for an array of N dots would not necessarily indicate that visual WM has a capacity of N items. However, this general type of paradigm could provide an ideal method for examining visual WM capacity because it minimizes the involvement of verbal WM. The present study therefore used a variant of this task to measure visual WM capacity.

The Units of WM Capacity

To quantify the capacity of visual WM, it is necessary to choose the appropriate units of measurement. In the case of verbal WM, for example, capacity may best be measured in terms of the number of seconds required to pronounce the to-be-remembered material rather than the number of words or even the number of syllables (see, e.g., Schweickert & Boruff, 1986). A compelling demonstration of this was provided by Ellis and Hennelly (1980), who showed that an individual who is bilingual in Welsh and English may appear to have a smaller digit span when the digits are presented in Welsh than when they are presented in English, presumably because of the longer duration of Welsh vowels.

In the visual domain, we propose that WM capacity should be quantified in terms of the number of objects that must be stored. This hypothesis is based on object-based theories of attention, which propose that attention is directed toward integrated object representations rather than independent features. In support of an object-oriented attention mechanism, several studies have shown that if one feature of an object is attended, then all features of that object are available for report at no additional cost (e.g., Duncan, 1984; Egly, Driver, & Rafal, 1994; Kahneman, Treisman, & Gibbs, 1992; Vecera & Farah, 1994). If the object-based representations implied by these studies form the contents of WM, then the number of objects that can be simultaneously stored in visual WM should be just as large for multifeature objects as for single-feature objects.

Suggestive evidence for the storage of integrated objects in visual WM was recently provided in a transsaccadic integration paradigm by Irwin and Andrews (1996). In this study, participants briefly viewed an array of six colored letters and then made an eye movement to a peripheral object, at which time the array of letters was extinguished. Shortly after the eye movement, an asterisk was placed near a location that had previously contained a letter, and the participants were required to report the color and identity of the letter that had been present at that location moments before. Participants frequently reported both the color and the identity correctly, and the observed accuracy level indicated that the participants had retained information about three to four objects, which is consistent with the number of objects retained in previous studies in which only letter identities were retained (Irwin, 1992). Thus, participants appeared to be able to retain the colors and the identities of a set of letters approximately as well as they could retain just the identities. However, although these results are suggestive, this study does not provide strong evidence for integrated WM representations. First, because letters were used as stimuli, it is possible that the observers used a combination of verbal and visual working memory, storing the letters verbally and the colors visually (see also Treisman, 1977). This situation could have doubled the apparent capacity of WM without the use of integrated representations. Second, even if both the letter identities

and the colors were stored visually, it is possible that there are separate visual WM stores for color and form information (especially for highly overlearned forms such as alphanumeric characters). This situation would again double the apparent WM capacity without the use of integrated object representations. Third, the storage of identity alone and the combined storage of color and identity were examined in different experiments, making it difficult to compare the absolute levels of performance between these conditions. A few other experiments have also addressed this issue over the years—with varying conclusions—but these experiments are subject to similar and additional criticisms (for an extensive discussion of these experiments, see Walker & Cuthbert, 1998). Thus, it is not yet known whether integrated object representations are stored in visual WM.

Overview of the Present Study

The goal of the present study was to provide a first step toward understanding the nature of visual WM storage by providing an estimate of the storage capacity of visual WM for simple, suprathreshold features and for objects defined by conjunctions of simple features. To measure visual WM storage capacity, we used a variant of the sequential comparison paradigm developed by Phillips (1974), as illustrated in Figure 1A. On each trial, participants viewed a briefly presented sample array consisting of 1–12 objects (e.g., colored squares). After a short delay, a test array was presented that was either identical to the sample array or differed in terms of a single feature of a single item in the array (e.g., one square might have changed color). Participants then indicated whether the test array was identical to the sample array. The delay between the sample and test arrays was sufficiently long to rule out significant contributions from iconic memory (Loftus, Duncan, & Gehrig, 1992; Phillips, 1974).

To measure visual WM capacity, we varied the number of items in each array (the set size), assuming that accuracy would be nearly perfect for set sizes that were within the capacity of visual WM and would decline systematically as the set size exceeded visual WM capacity. For example, if visual WM capacity were five items, then one might expect perfect performance for Set Sizes 1–5 and a gradual decline in accuracy as the set size increased beyond five items. Finding the breakpoint in the set size function is an imperfect means of assessing memory capacity, however, because capacity may vary across participants and across trials within subjects, leading to a breakpoint at the minimum capacity level rather than the typical capacity level. For example, if the average capacity were five items, but participants sometimes retained three items and sometimes retained seven items, accuracy would begin to deviate from ceiling at three items, leading to a substantial underestimation of the typical capacity. Therefore, we also used a more quantitative approach that was developed by Pashler (1988). In essence, this approach assumes that if a participant can hold k items in memory out of an array of S items, then the item that changed should be one of the items being held in memory on $k \div S$ trials, leading to correct performance on $k \div S$ of the trials on which an item changed. To adjust for the effects of guessing, this approach also takes into account the false alarm rate (i.e., the proportion of trials without a change on which subjects report that an item changed). Specifically, $k = [S \times (H - F)] \div (1 - F)$,
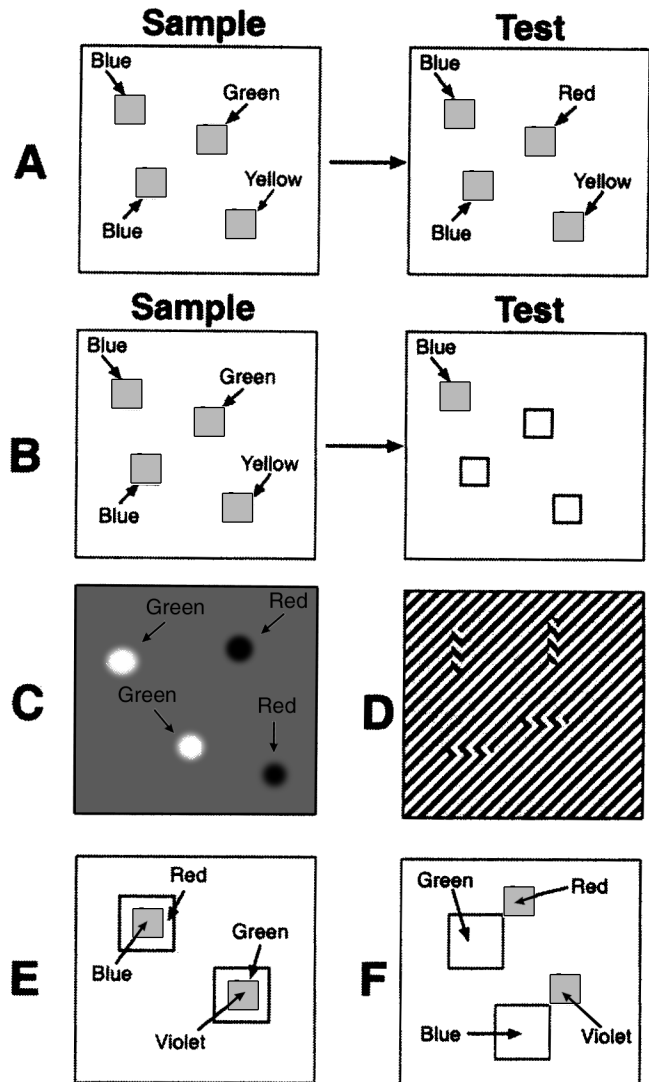

Figure 1. Example stimuli from several of the experiments. A: Stimuli used in Experiments 1 and 2. B: Stimuli used in Experiment 6. Note that the uncued test array items are empty placeholders. C: Gaussian color blobs used in Experiment 13. D: Orientations defined by texture in Experiment 13. E: Dual-color object stimuli used in Experiment 15. F: Displaced dual-color objects used in Experiment 16. Note: The stimuli are not drawn to scale.

where k is the memory capacity, S is the set size of the array, H is the observed hit rate (hits $\div$ [hits + misses]), and F is the observed false alarm rate (false alarms $\div$ [false alarms + correct rejections]). As discussed by Pashler (1988), this means of estimating memory capacity is not assumption-free, but it does provide a good first approximation.

The capacity of visual WM for simple features is addressed in Experiments 1–10. To provide a pure estimate of the storage capacity of visual WM, it is necessary to rule out several extraneous factors that might influence performance in this paradigm.

Perhaps the most obvious extraneous factor is verbal WM. In the display shown in Figure 1A, for example, a participant might be able to improve performance by using verbal WM to store the items in the right visual field and using visual WM to store the items in the left visual field, leading to an overestimation of visual WM capacity. This possibility was examined in Experiments 1–3. A second potential problem is that decreases in accuracy might occur at larger set sizes for purely statistical reasons, even if there is no limit to visual WM capacity. Specifically, if the encoding of each individual item is imperfect, then there is the possibility of a false alarm for each item that must be compared; the larger the set size, the more items must be compared and the greater is the possibility of a false alarm (even if the individual items are stored independently and without mutual interference). This factor, which would lead to an underestimation of visual WM capacity, is explored in Experiments 4–6. A third potential problem arises from limitations that arise prior to WM storage, such as perceptual encoding and the transfer of perceptual representations into WM. It is unlikely that the identification of simple, highly discriminable features is subject to significant capacity limitations, but Jolicoeur and Dell' Acqua (1998) have demonstrated that the process of transferring perceptual representations into WM is limited in capacity, and it is therefore important to ensure that errors in the present experiment reflect limitations in storage rather than limitations in these earlier processes. This issue is explored in Experiments 7 and 8. The next two experiments explore the generality of the estimate of visual WM capacity obtained from Experiments 1–8. Specifically, Experiment 9 assesses visual WM capacity at longer delays, and Experiment 10 assesses whether the capacity is similar for other simple features. The capacity of visual WM for objects defined by conjunctions of features is addressed in Experiments 11–16.

Portions of the results of Experiments 1, 2, 4, 7, 11, 14, and 15 were described previously in Luck and Vogel (1997). In the present study, we present a more complete description of the procedure and results of those experiments. Moreover, we present the results from 10 new experiments that serve to extend and further strengthen the conclusions of the previously reported study.

General Method

The following general procedure was used for all of the experiments described in this study, unless otherwise noted.

Participants

A different group of 10 college students participated in each experiment. These participants reported no history of neurological problems and reported having normal color vision and normal or corrected-to-normal visual acuity.

Stimuli

All stimulus arrays were presented within a 9.80° × 7.30° region on a video monitor with a gray background (8.20 cd/m^2), viewed at a distance of 70 cm. Each item was randomly placed within this region with the constraint that the items in a given array were separated from each other by at least 2° (center-to-center). The x, y, and luminance values for the colored squares were measured with a Tektronix model J17 colorimeter using the 1931 CIE color coordinate system. The color of each object was randomly selected (with replacement) from a set of seven colors: white (92.46 cd/m^2), red (x = .642, y = .327; 22.62 cd/m^2), blue (x = .152, y = .067; 9.66 cd/m^2), green (x = .318, y = .569; 64.99 cd/m^2), black (<0.01 cd/m^2), yellow (x = .478, y = .452; 65.23 cd/m^2), and purple (x = .304, y = .149; 7.04 cd/m^2).

In the experiments that used colored squares, each square subtended 0.65° × 0.65° of visual angle; in the experiments that used oriented bars, each bar subtended 0.03° × 1.15° of visual angle and were presented at an orientation of 0°, 45°, 90°, or 135°. When a feature changed between the sample and test arrays, the new value was selected at random from all of the other possible feature values. Note that a color or orientation could appear more than once in an array, and the color or orientation of a changed item was sometimes the same as the color or orientation of another item in the array. Consequently, correct performance required the encoding of the colors of the individual objects (i.e., token representations) rather than a mere list of the colors that were present in the array (i.e., type representations). However, informal testing indicated that the presence or absence of repetitions does not have a large effect on performance in this task.

Procedure

Each trial consisted of a 100-ms sample array followed by a 900-ms blank delay period and then a 2,000-ms test array. The sample and test arrays were identical except that one feature (e.g., color) of one item in the test array was different from that of the corresponding item in the sample array on 50% of the trials. Participants responded by pressing one of two buttons on each trial to indicate whether the sample and test arrays were the same or different. Accuracy was stressed rather than speed, and the responses were not timed. Each participant was tested for a single session of 30–45 min, including one short practice block for each experimental condition.

To rule out contributions from verbal WM, in all of the experiments we used a verbal load procedure in which two randomly selected white digits (1.70° × 0.85°) were presented side by side in the fovea for 500 ms at the beginning of the trial. After a blank interval of 1 s, the sequence of sample and test arrays was presented. Participants were instructed to remember the two numbers and were encouraged to subvocally rehearse them throughout the trial. The participants were required to report aloud the two digits at the end of the trial, after they had made their button-press response to the test array. Accuracy for this task was near 100% and was independent of the set size of the visual WM task in all experiments; the few trials with errors on the verbal WM task were excluded from analyses of the visual WM task.

Experiments 1–3: Visual WM Capacity Without Verbal Contamination

Experiments 1 and 2

In Experiments 1 and 2, we sought to provide an initial estimate of the capacity of visual WM for simple features. More specifically, we measured WM performance across several set sizes to determine the number of colored squares that could be held simultaneously in visual WM. Experiment 1 examined set sizes of 1, 2, and 3 squares, and Experiment 2 tested set sizes of 4, 8, and 12 squares (with a separate set of participants).

In addition to providing an initial estimate of visual WM capacity, we also designed this experiment to assess the presence of contamination from verbal WM. There are several general approaches that have previously been used to minimize the contri-

bution of verbal WM during visual memory tasks. The first is to use briefly presented stimuli, which serves to discourage the relatively slow process of phonological coding that is necessary for verbal WM (Frick, 1988). Another approach is to use stimuli that are difficult to name, such as irregular shapes (Cermak, 1971). A third and particularly effective procedure for discouraging the use of verbal WM is to use a concurrent verbal task in addition to the visual memory task. For example, giving participants a verbal load while they perform a visual task has been demonstrated to greatly interfere with the participation of verbal WM in the visual task (for a review see Baddeley, 1986). Verbal loads can consist of shadowing procedures, articulatory suppression (e.g., having the participant repeat a syllable aloud throughout the course of each trial), as well as having participants remember a verbal item such as a letter throughout each trial.

In the present study, we used two methods to discourage the use of verbal WM. First, the sample stimuli were presented too briefly to be easily encoded verbally. For example, it would be difficult to generate the names of 8 colors with a 100-ms stimulus presentation, especially given that the location of each color must be encoded because colors were allowed to repeat within an array. Second, we directly assessed the contribution of verbal WM by comparing performance on the visual WM task both with and without the concurrent verbal load described in the General Methods. Previous studies that have used this verbal load procedure

have found that it substantially disrupts performance on verbal WM tasks (Dixon & Shedden, 1993).

Experiments 1 and 2 were completely identical to each other except that different ranges of set sizes were tested and that separate groups of participants were used. The verbal load procedure was used in half of the trial blocks, and set size varied within trial blocks. The order of blocks was counterbalanced across participants. Participants performed 40 trials per combination of condition (load vs. no-load) and set size.

The results of Experiments 1 and 2 are summarized in Figure 2A, and analysis of variance (ANOVA) summaries for all experiments are provided in Table 1. As expected, accuracy was nearly perfect for the smallest set sizes and then fell monotonically as set size increased. Moreover, performance at all set sizes was largely unaffected by the presence of a concurrent verbal load. For both experiments, these results yielded significant effects of set size, but neither the factor of load nor the interaction of load and set size was significant in either experiment (see Table 1).

From these results, it appears that all participants could hold at least 2 items in visual WM on virtually every trial. Performance declined slightly for Set Sizes 3 and 4 but was still quite high at these set sizes, and it is therefore likely that participants were able to store 3–4 items in WM on a large majority of trials. For example, if only 2 items were held in memory, participants should have made errors on approximately 2 to 3 of the change trials at
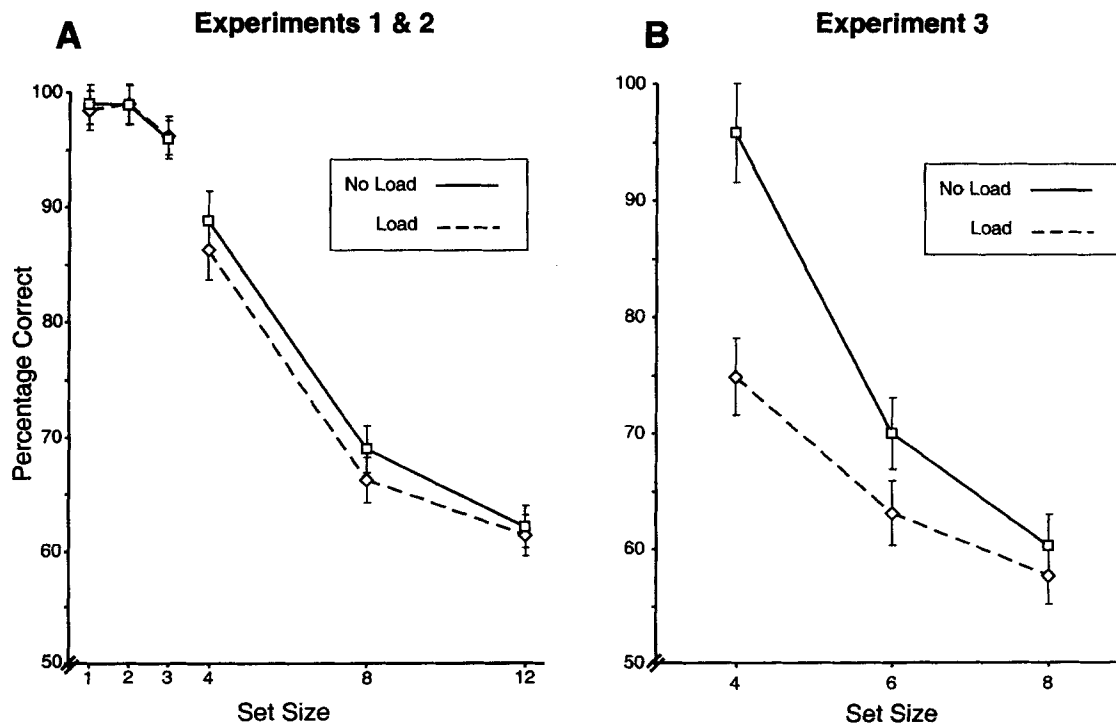

Figure 2. A: Mean accuracy (including change and no-change trials) for the load and no-load conditions of Experiments 1 and 2 as a function of set size. B: Mean accuracy for the load and no-load conditions as a function of set size in the verbal working memory task used in Experiment 3. In this and all subsequent figures, error bars represent 95% confidence intervals as described by Loftus and Masson (1994).

Table 1
Analysis of Variance Results for Experiments 1–16

	Condition		Set size		Interaction	
Experiment	F	p	F	p	F	p
1	1.43	ns	60.48	<.001	0.30	ns
2	0.10	ns	7.13	<.01	0.12	ns
3	13.39	<.01	58.45	<.001	10.35	<.01
4	0.30	ns	96.13	<.001	1.30	ns
5	81.89	<.001	95.74	<.001	13.22	<.001
6	2.79	ns	97.07	<.001	3.68	<.05
7	0.84	ns	53.41	<.001	1.11	ns
8	1.23	ns	0.72	ns	0.38	ns
9	2.24	ns	91.98	<.001	0.15	ns
10	1.64	ns	95.70	<.001	2.07	ns
11	1.93	ns	48.45	<.001	0.64	ns
12	1.72	ns	31.63	<.001	0.78	ns
13	0.71	ns	18.03	<.001	0.31	ns
14	1.64	ns	82.53	<.001	0.43	ns
15	0.73	ns	49.47	<.001	0.21	ns
16	12.45	<.001	79.24	<.001	3.86	<.05

Note. $ns = p > .10$.

Set Size 3 and half of the change trials at Set Size 4. Performance was clearly better than that, so these data indicate that an average of 3–4 items were held in visual WM.[1]

To provide a more formal estimate of the capacity of visual WM, we applied the equation of Pashler (1988), as described above. This equation provides a capacity estimate for the data obtained at a single set size, but it is appropriate only for set sizes that exceed the participant's capacity because it will underestimate the actual capacity if applied to smaller set sizes. We therefore used it to estimate the capacity at Set Sizes 8 and 12 (averaged over the load and no-load conditions). A separate estimate was obtained for each participant, allowing the calculation of 95% confidence intervals. The resulting estimates were 3.23 +/− 0.83 for Set Size 8 and 3.40 +/− 0.81 for Set Size 12.[2] Thus, we estimate that approximately 3–4 colors can be held simultaneously in visual WM, which is similar to previous estimates of visual WM capacity (e.g., Sperling, 1960).

This quantification of WM capacity might seem to suggest that working memory holds a high-fidelity representation of a small number of items rather than partial information about the entire array. However, it is possible that visual WM contains a representation of every item in the array and that the representations simply become lower in fidelity as the number of items increases, yielding poorer performance at larger set sizes. The results of the present experiments cannot satisfactorily distinguish between these two alternatives, and it is therefore more appropriate to conclude from these data that visual WM can store an amount of information that is equivalent to a complete representation of 3–4 items, even though WM might actually contain partial information about many more items.

This capacity estimate was not significantly affected by the presence of a concurrent verbal load, which suggests that performance on the task was limited primarily by the capacity of visual WM, with little or no contribution from verbal WM. This conclusion is further supported by the following experiment, which

demonstrated that this concurrent verbal load task significantly impairs verbal WM performance. However, although these results indicate that verbal WM does not contribute to performance on our visual WM task, all of the following experiments include the verbal load task to ensure that none of the results are contaminated by verbal WM.

Experiment 3

To demonstrate that our verbal load procedure was effective, we conducted Experiment 3 to test the extent to which this procedure disrupts performance on a task that is analogous to our visual WM task but relies heavily on verbal WM. In this task, the sample stimulus consisted of a sequence of letters presented serially at fixation. The test stimulus was a second sequence of letters that was either identical to the first sequence or differed in terms of one letter, and participants indicated whether the two sequences on each trial were identical. To demonstrate the effectiveness of the verbal load used in Experiments 1 and 2, we compared performance on this verbal WM task with and without a concurrent verbal load. To encourage the use of verbal rather than visual WM, we presented the sample letters in uppercase and the test letters in lower case, forcing participants to remember abstract identities rather than low-level visual features. Only consonants were used, which was necessary to decrease the possibility that the participants could increase their apparent capacity by remembering multiletter syllables rather than individual letters. Each letter was presented individually at fixation for 100 ms, with an interstimulus interval of 900 ms. At the end of the first sequence, an asterisk was presented for 100 ms, and the second sequence began 900 ms afterward. At the end of the second sequence, a question mark was presented for 1 s, signaling the participant to respond. The sample and test sequences were identical on half of the trials and differed by one letter on the other half. The letter that changed occurred with equal probability at all positions within the test sequence. Set sizes of 4, 6, and 8 letters were randomly intermixed within blocks. On each trial during the load block, participants performed the same verbal load task described in the earlier experiments. During the no-load block, participants performed only the main verbal WM task. The order of the blocks was counterbalanced, and the number of trials was equal to that used in Experiment 2.

As illustrated in Figure 2B, accuracy on the main task was significantly lower when it was performed concurrently with the verbal load task. This effect was smaller at the larger set sizes, possibly because performance began to approach chance. These effects led to significant main effects of set size and verbal load presence, as well as a significant interaction between these factors.

---

[1] Errors in this visual WM task can be characterized as misses (responding *same* on a change trial), or false alarms (responding *different* on a same trial). In this and all subsequent experiments reported here, misses were approximately three times as common as false alarms across all set sizes.

[2] Cowan, Saults, and Fristoe (1999) have reported an improved formula for estimating capacity that appears to provide a more robust estimate at larger memory array sizes by including contributions from no-change trials. When we applied this new formula to our results, the resulting estimates were 2.82 for Set Size 8 and 2.85 for Set Size 12.

Thus, this experiment demonstrates that our verbal load procedure has a significant impact on verbal WM, and the absence of an effect of verbal load in Experiments 1 and 2 indicates that performance in those experiments was not likely to have been influenced by a verbal coding of the stimuli.

It is possible that the lack of any effect of verbal load in Experiments 1 and 2 was partly due to the small size of the load itself (two digits), rather than the absence of any verbal coding. That is, because the verbal load did not completely consume the entire capacity of verbal WM, it is possible that the residual capacity was used to store the names of some of the colors. Experiment 3 addressed this issue by demonstrating that this load of two items causes a very large decrement in the performance of a verbal WM task. Thus, if verbal working memory contributed substantially to performance of the visual WM task in Experiments 1 and 2, a significant drop in performance should have been observed in the load conditions of these experiments. It should be noted that the verbal load procedure does not simply fill two slots in verbal WM; the procedure also interferes with the process of translating visually presented items into verbal WM. Indeed, previous studies of articulatory suppression in verbal WM tasks have demonstrated that even a one-item verbal load can significantly disrupt verbal coding, and these studies have further shown that increasing this load to several items does not necessarily increase the amount of interference (see Baddeley, 1986, p. 37). In addition, our paradigm also minimizes verbal WM contributions by using brief exposure durations and nonverbal stimuli and by requiring a nonverbal report of the memory items. Moreover, we have also used a much more substantial articulatory suppression procedure in numerous subsequent experiments and have found essentially identical results (Vogel et al., 1998; Woodman, Vogel, & Luck, in press). Thus, it is unlikely that our estimate of visual WM capacity was significantly distorted by residual verbal WM capacity.

Experiments 4–6:
The Contribution of Accumulated Decision Errors

Although the results of Experiments 1–3 suggest that visual WM can accurately retain 3–4 colors, it is possible that the poor performance observed at larger set sizes was not entirely due to a memory limitation per se, but may be at least partially the result of an accumulation of errors at the decision stage. In this paradigm, the memory representation of the sample array must be compared with the test array to determine whether or not the two arrays are identical, and this comparison requires a same–different decision for each item in the array. Each of these decisions is subject to some degree of error, with the amount of error for a single decision being dependent on the confusability of the colors and the fidelity of the representation. For example, if an item in the sample array was a highly saturated red and changed to a slightly less saturated red in the test array, then the possibility of an error for that item would presumably be nonnegligible. Because the task requires participants to make a single behavioral response concerning the entire array, the number of opportunities to make an error increases at larger set sizes, which may lead to poorer performance. Thus, accuracy may decrease as set size increases even in the absence of capacity limitations (for a more extended discussion of this issue, see Palmer, Ames, & Lindsey, 1993; Shaw, 1982). Therefore, the increased error rate observed for larger set sizes in the previous experiments may have been at least partially due to the accumulation of individual-item errors at the decision stage, and this situation would have created a downward bias on our estimate of visual WM capacity.

It seems unlikely that accumulated decision errors were entirely responsible for the set size effects observed in Experiments 1 and 2 because performance at Set Sizes 1 and 2 was at or near ceiling, indicating that the rate of individual-item errors was very small (because of the highly discriminable colors used in these experiments). Nevertheless, this argument is not sufficient to entirely rule out the contribution of accumulated decision errors. In Experiments 4–6, we therefore evaluated the role of accumulated decision errors more directly by using a partial-report procedure that allows the participant to make a single decision on each trial (Averbach & Coriel, 1961; Klemmer, 1963; Palmer, 1990). Specifically, one item was cued in the test array (but not in the sample array), and the cued item was the only item in the array that might have changed in color between the sample and test arrays. The participant did not know which item would be cued until the onset of the test array, and it was therefore necessary to store the entire sample array in WM. However, a comparison was necessary only for the one item that was cued in the test array.

Experiment 4

We compared two conditions in Experiment 4, one with a cue and one without. The no-cue condition was identical to the load condition of Experiment 2. The cue condition was identical to the no-cue condition except that an outline square was placed around one randomly selected item in the test array to indicate which item might have changed. As in the no-cue condition, an item changed color between the sample and test arrays on 50% of trials, but if an item did change, it was always the cued item. If accumulated decision errors contributed to the capacity estimate in Experiment 1, then accuracy at larger set sizes should be substantially higher in the cue condition compared with the no-cue condition. However, if accumulated decision errors did not significantly affect performance, then there should be little or no difference in accuracy between the cue and no-cue conditions.

The cue square was drawn in light brown ($x = 0.571, y = 0.385$, luminance $= 13.38$ cd/m^2), which made it easily discriminable from the colors used for the sample and test arrays. The cue square was $0.97° \times 0.97°$ with a line thickness of $0.03°$ and was centered around one of the $0.65° \times 0.65°$ squares in the test array. The cue was visible for the entire duration of the test array (2,000 ms). Participants performed 40 trials for each combination of set size and condition (cue vs. no cue).

The results from Experiment 4 are illustrated in Figure 3A. As in Experiment 2, accuracy in both the cue and no-cue conditions declined monotonically as the set size increased, yielding a significant effect of set size. Accuracy was slightly better without the cue at Set Size 4 and was slightly better with the cue at Set Size 12; however, neither the factor of cue-presence nor the interaction of cue presence and set size was significant. Thus, the increase in errors at larger set sizes does not appear to reflect an accumulation of individual-item errors at the decision stage and instead appears to purely reflect limitations in WM capacity.
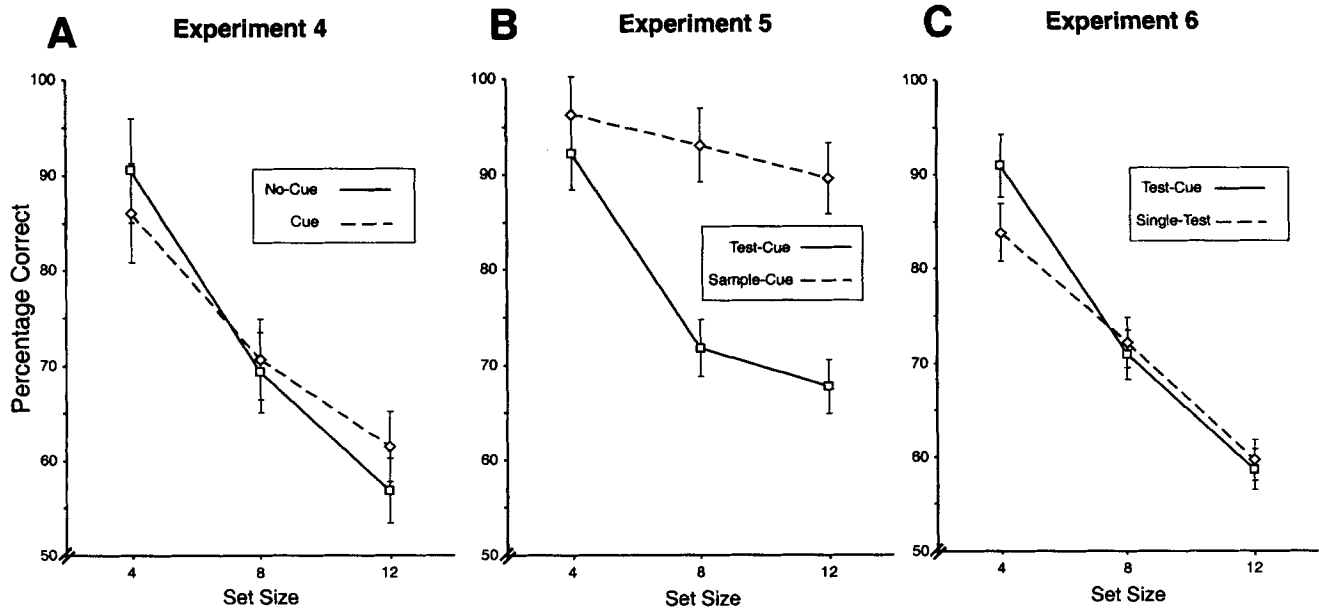
Figure 3. A: Mean accuracy for the cue and no-cue conditions of Experiment 4. B: Mean accuracy for the sample-cue and test-cue conditions of Experiment 5. C: Mean accuracy for the test-cue and single-test conditions of Experiment 6.

Experiment 5

The cue in Experiment 4 did not significantly improve performance, but before accepting this null result, it is necessary to demonstrate that participants are capable of adequately using a cue of this nature. In Experiment 5 we therefore tested the effectiveness of the cue by presenting it during the sample array rather than the test array. The cue again indicated the only item in the array that could possibly differ between the sample and test arrays, but in this experiment the cue made it possible for participants to limit both storage and decision to a single item. That is, if participants are capable of using this cue, they need only retain a single item in memory, which should result in highly accurate performance at all set sizes. However, if the participants are unable to use the cue, then performance for cued sample arrays should be similar to performance for cued test arrays.

Two conditions were tested in this experiment. In the *test-cue* condition, the cue was presented during the test array. This condition was identical to the cue condition of Experiment 4. In the *sample-cue* condition, the cue was presented during the sample array. In both conditions, the cue was presented for the entire duration of the array (100 ms for the sample-cue condition and 2,000 ms for the test-cue condition). The two conditions were otherwise identical.

The results from Experiment 5 are summarized in Figure 3B. Performance in the test-cue condition declined greatly as set size increased, just as in Experiment 2. In contrast, performance declined only slightly at larger set sizes in the sample-cue condition. This difference in set size effects resulted in significant main effects of both set size and cuing condition as well as a significant interaction between set size and cuing condition. Thus, the cue was very effective at reducing set size effects when presented during

the sample array, at which time it could help participants limit which items were stored in working memory as well as limiting the decision to a single item. When the cue was presented during the test array, however, it was ineffective, indicating that limiting the decision to a single item does not improve performance in this task. Thus, the decline in performance observed at larger set sizes in Experiments 1 and 2 appears to be a result of capacity limitations rather than accumulated decision errors.

Experiment 6

The use of a cue in the test array to eliminate accumulated decision errors is based on the assumption that the decision process is not automatic, such that decisions can be limited to a subset of the information in the test array. It is possible, however, that the mere presence of the uncued items in the test array may initiate automatic memory-matching decisions that may impair performance. Experiment 6 tested this possibility by comparing performance in the test-cue condition from Experiments 4 and 5 with performance in a *single-test* condition in which only the cued item was presented in the test array, forcing the response to be made on the basis of a single decision. Previous studies (Klemmer, 1963) as well as our own pilot experiments have indicated that completely removing the uncued items in the array tends to disrupt memory performance, probably because the Gestalt of the array is so greatly altered. Consequently, instead of completely removing the uncued items, we presented placeholders consisting of black outlined squares in the locations of the uncued items in the test array. As illustrated in Figure 1B, the use of placeholders allowed the overall spatial distribution of objects to remain the same between the sample and test arrays and eliminated any color information from the uncued items in the test array that might trigger an

automatic comparison process. As in Experiments 4 and 5, the decline in accuracy at larger set sizes should be greatly reduced in this condition if performance is significantly influenced by accumulated decision errors. If the set size effect is due to real limitations in capacity, however, then a normal set size effect should be observed.

The single-test condition may also rule out an alternative hypothesis regarding the capacity of visual WM. Specifically, it is possible that comparing 4 items in the sample array with 4 items in the test array actually requires that 8 items be stored temporarily in visual WM. Thus, visual WM capacity may actually be twice as large as we originally estimated. If this hypothesis is correct, then performance on the placeholders-cue condition should show a smaller set size effect than that found in our previous experiments because the visual WM load would be the sample set size plus the 1 item in the test array rather than twice the sample set size.

This experiment was identical to Experiment 4 with the exception that the no-cue condition was replaced with the single-test condition. The two conditions were identical, except that the cued item in the single-test condition was cued indirectly by replacing all of the uncued squares with black outline squares. Thus, the test array in the single-test condition consisted of one colored square and several outline squares (one less than the set size), and participants were instructed to report whether the one colored square was the same color as the corresponding square in the sample array. The outline squares subtended 0.65° × 0.65°, just like the colored squares that they replaced, and they were filled with the background gray of the display. The lines that formed the outline squares were 0.03° thick. These black outline squares surrounded all of the colored squares in both the sample and test arrays in both the test-cue and single-test conditions.

The results from Experiment 6 are shown in Figure 3C. Performance in both conditions showed a substantial decline at the larger set sizes, and the magnitude of this effect was similar to that observed without the use of any sort of cue (e.g., in Experiment 2). In addition, performance was virtually identical for the single-test and test-cue conditions at Set Sizes 8 and 12, and performance was actually slightly worse in the single-test condition than in the test-cue condition at Set Size 4. These effects led to a significant main effect of set size and a significant interaction between set size and cuing condition, but the main effect of cuing condition was not significant. Pairwise comparisons revealed that the test-cue and single-test conditions differed significantly at Set Size 4, $F(1, 9) = 8.69$, $p < .05$, but did not differ significantly at Set Sizes 8 and 12. We have no explanation for the small effect at Set Size 4.

Two conclusions can be drawn from this experiment. First, these results provide further evidence that the decline in accuracy at larger set sizes reflects limitations in visual WM capacity rather than accumulated decision errors. Second, these results suggest that participants do not perform this task by storing both the sample array and the test array in visual WM when comparing them, which would make the estimated capacity of visual WM appear to be half its actual capacity. If this had been the case, then participants should have been more accurate in the single-test condition, in which the test array contained only one colored square. It is possible that the outline squares were automatically stored in working memory, even though they were completely task irrelevant, but pilot experiments have shown that accuracy is not improved by eliminating the outline squares such that there is only

one object in the test array. Thus, our original estimate of a capacity of three to four items does not appear to be half of the actual value, and it does not appear to be significantly influenced by accumulated decision errors.

Experiments 7 and 8: The Role of Encoding Limitations

The results of Experiments 1–6 suggest that visual WM storage has a storage capacity of only 3–4 simple colors. However, it is possible that this relatively small estimate of capacity is distorted by limitations in the encoding stages that precede visual WM storage. Specifically, the sample arrays in the above experiments were presented for only 100 ms, and it is possible that this short duration provided an insufficient amount of time to identify the colors and transfer the perceptual representations into stable visual WM representations (see, e.g., Potter, 1976). We refer to these identification and storage processes as *encoding* processes. Experiments 7 and 8 test the possibility that the poor performance that was observed at larger set-sizes in the preceding experiments was the result of impairments in encoding processes rather than limitations in storage capacity.

Experiment 7

Experiment 7 tested the role of encoding limitations by simply increasing the duration of the sample array while holding the retention interval (i.e., the inter-stimulus interval) constant. Two conditions were tested, a *short-duration* condition that was identical to the load condition of Experiment 2 and a *long-duration* condition that was identical except that the duration of the sample array was increased to 500 ms. The delay between the sample and test arrays was 900 ms in both conditions. If the decline in accuracy at larger memory-set sizes observed in the previous experiments was at least partially due to limitations in encoding, then increasing the duration of the sample array by a factor of 5 should cause a substantial increase in performance.[3] However, if the set size effects found in the previous experiments were due solely to storage capacity limitations, then a longer exposure time should have little or no effect on performance.

The results from Experiment 7 are summarized in Figure 4. As in the previous experiments, accuracy declined monotonically as the set size increased, yielding a significant effect of set size.

[3] Although we have increased the physical duration of the sample array by a factor of 5 in this experiment, this is not the same as a fivefold increase in the perceptual duration of the array because of visible and informational persistence (i.e., iconic memory, Coltheart, 1980; Di Lollo & Dixon, 1988). Loftus and colleagues have described an approach to measuring the informational worth of the icon, proposing that it is equivalent to an additional 100 ms of exposure duration (Loftus et al., 1992; Loftus, Johnson, & Shimamura, 1985). In these terms, the comparison of a 100-ms sample duration with a 500-ms duration is actually a comparison of 200 versus 600 ms worth of information. However, this is still a three-fold increase in the period of available information, which should be sufficient to reveal any effects of encoding limitations on performance. Moreover, even if the estimate provided by Loftus and his collaborators underestimates informational persistence by a factor of 2, a fivefold increase in physical stimulus duration would still be equivalent to more than a doubling of the period of available information.
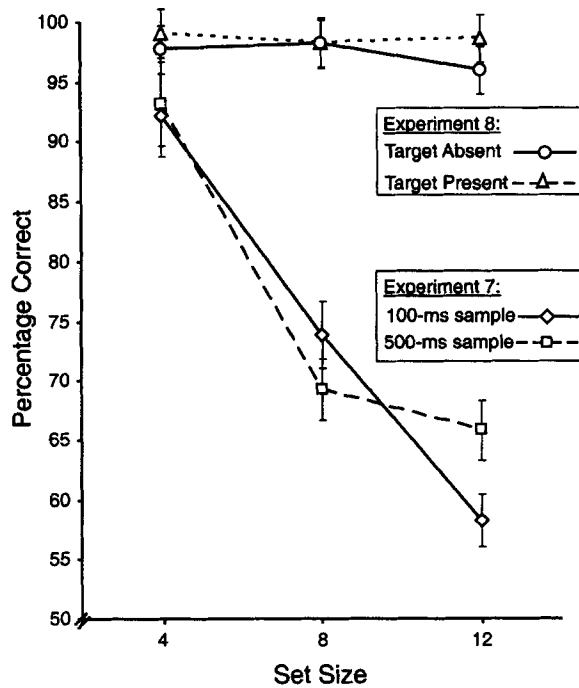
Figure 4. Mean accuracy for the 100-ms sample and 500-ms sample conditions of Experiment 7 and for the target-absent and target-present conditions of Experiment 8.

Although mean accuracy differed slightly between the short-duration and long-duration conditions at larger memory-set sizes, neither the main effect of duration nor the interaction between duration and set size approached significance. Thus, a very large increase in the time available for encoding had no significant effects on performance, suggesting that performance in this task is not limited by encoding and is more likely to be limited by the storage capacity of visual WM. Recent work on the attentional blink (e.g., Raymond, Shapiro, & Arnell, 1992) and short-term consolidation (e.g., Jolicoeur & Dell' Acqua, 1998) has demonstrated that encoding an item into WM may actually take as long as 300–400 ms per item, and it is therefore possible that our duration manipulation was not large enough to completely rule out a possible encoding limitation on our capacity estimate. However, we have shown that colored squares such as those used in this experiment require only about 100 ms per item to be encoded in visual WM (Vogel et al., 1998). Further research is necessary to completely resolve the issue of encoding limitations.

Experiment 8

In Experiment 8, we sought to provide converging evidence for the hypothesis that the poor performance at large set sizes is caused solely by storage limitations and not by insufficient encoding time. Our approach in this experiment was to use a visual search task to demonstrate that the 100-ms duration of the sample array provides sufficient time for participants to identify the colors of the squares. At the beginning of each trial, a single colored square was presented for 500 ms, indicating the target color for

that trial. After a delay of 500 ms, an array of 4, 8, or 12 colored squares was presented for 100 ms, and participants were required to press one of two buttons to indicate whether the color target was present or absent in the array. These search arrays were identical to the sample arrays used in Experiment 2.

Accurate performance on this visual search task requires a discrimination between the target color and the colors of each of the items in the array. This is particularly true for target-absent trials, in which accurate performance requires that every item in the array is accurately compared with the target. Therefore, if a 100-ms duration provides sufficient time to compare each of the colors in the sample array with the target color, then accuracy should be very high at all set sizes for both target-present and target-absent trials. However, if 100 ms is insufficient for discriminating all of the colors at large set sizes, then there should be a decline in accuracy at the larger set sizes.

It should be noted that almost any theory of visual perception would predict that 100 ms should be sufficient to create a representation of 12 large colored squares that would be sufficient to perform this task. For example, Potter (1976) found that participants could accurately classify complex photographs of real-world scenes even when they were presented at rapid rates (8 pictures/s), and the identification of 12 highly discriminable colored squares should be at least this rapid. However, it is better to demonstrate this empirically rather than relying on extrapolations from previous experiments.

The results from the visual search task are summarized in Figure 4. For all three set sizes, accuracy was found to be near 100% on both target-present and target-absent trials. The main effects of set size and target presence did not approach significance, nor did the interaction between these factors. Thus, the encoding requirements of these stimuli were clearly very minimal and cannot explain the substantial reductions in accuracy observed at larger set sizes in the preceding experiments. We therefore conclude that the capacity estimate of three to four items obtained in Experiments 1 and 2 is not likely to have been distorted by verbal WM, by accumulated decision errors, or by encoding limitations.

Experiments 9 and 10:
Generality of the Capacity Estimate

Experiment 9

In Experiment 9, we examined the effects of increasing the retention interval in the same paradigm used in Experiments 1–7. This manipulation served two purposes. First, it provided a test of the generality of the capacity estimate provided by the preceding experiments. Specifically, it is possible that the memory system used to perform this task is subject to rapid decay, which would make it difficult to describe the memory capacity with a single number, but a demonstration of consistent levels of performance over a range of delays would indicate that a single number can reasonably be used to quantify memory capacity (at least for a limited range of stimulus materials). Second, these results provide a means of testing whether the memory system used to perform this task is similar to the memory system that underlies performance in other visual WM tasks. In other visual WM paradigms,

the representation typically remains quite robust over the first 4–5 s (Phillips, 1974; Simons, 1996), and a similar finding in the present experiment would provide one piece of evidence indicating that the same memory system underlies performance across paradigms.

This experiment was identical to the load condition of Experiment 2, except that three retention intervals (900 ms, 2,900 ms, and 4,900 ms) and two set sizes (4 and 8) were tested. Participants performed 36 trials for each combination of set size and retention interval, with both factors manipulated within blocks.

The results from Experiment 9 are shown in Figure 5A. As in the previous experiments, accuracy declined substantially as the set size increased from four to eight items, yielding a significant effect of set size. Although accuracy declined slightly at the longer retention intervals for both set sizes, there was no significant effect of retention interval and no significant interaction between set size and retention interval. This pattern of results suggests that there is little decay of the memory representation in this task over the course of several seconds, which is consistent with previous descriptions of visual WM performance. Thus, our capacity estimate of three to four items generalizes to longer delay intervals (although a decline in memory performance may occur at even longer delays, as found by Phillips, 1974).

Experiment 10

Experiment 10 was designed to further generalize our capacity estimate by examining another simple feature, namely, orientation.

If the previous results with colored squares reflect a general capacity limitation for simple features in visual WM—rather than a specific memory limitation for color—then participants should be able to accurately retain three to four oriented lines in visual WM.

This experiment was similar to the load condition of Experiment 2, with oriented lines rather than colored squares. The lines in a given array were selected from a set of four different orientations. This is a smaller set than the set of seven colors used in Experiments 1–9, but this size was necessary because it was not possible to generate seven orientations that were sufficiently discriminable from each other (which is important to avoid performance decrements because of accumulated decision errors). Because the number of possible feature values may influence task performance, we compared the orientation condition with a color condition in which the colors were selected from a set of four colors. To maximize the comparability of these conditions, we used the same stimuli in both conditions, and participants were asked to attend only to color in some blocks and only to orientation in others. The items within a given array were 0.03° × 1.15° lines drawn in a randomly selected combination of orientation (0°, 45°, 90°, or 135°) and color (blue, green, red, or white). Participants were instructed to detect color changes in one condition and orientation changes in the other condition, and there were never any changes between the sample and test arrays in the irrelevant feature dimension.
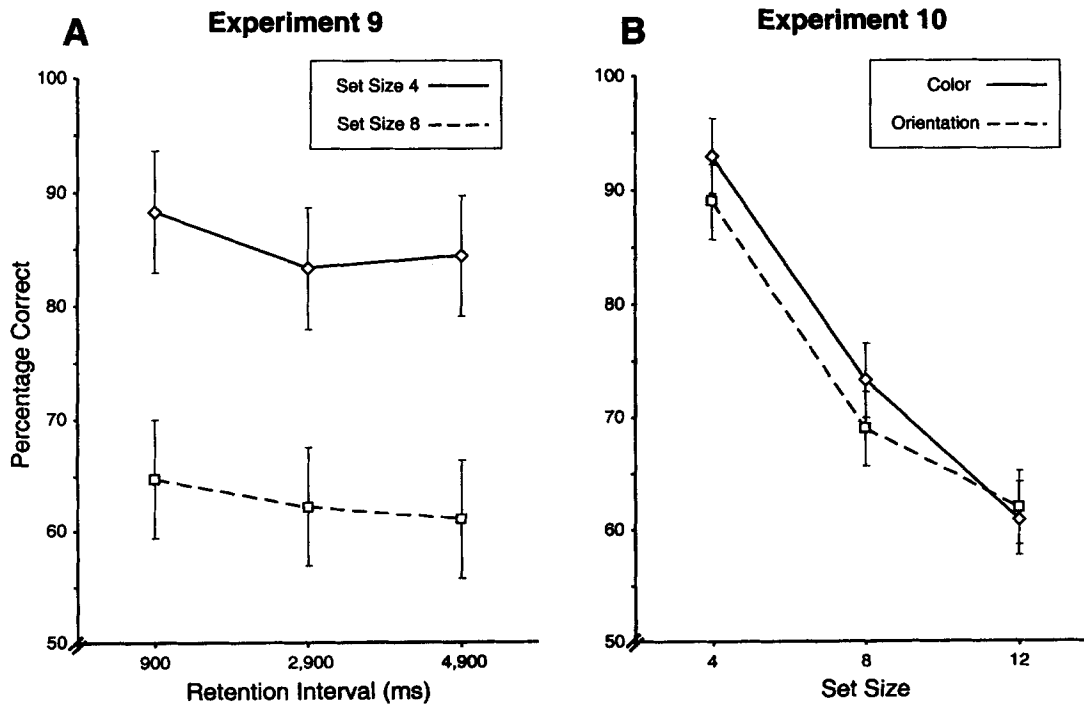

Figure 5. A: Mean accuracy for the Set Size 4 and Set Size 8 conditions of Experiment 9 as a function of retention interval (in milliseconds). B: Mean accuracy for the color and orientation conditions of Experiment 10.

Except as described above, this experiment was identical to the load condition of Experiment 2.

The results from Experiment 10 are summarized in Figure 5B. As in the previous experiments, accuracy in the color condition declined substantially as the set size increased beyond four items. Accuracy in the orientation condition was very similar but was slightly poorer than in the color condition at Set Sizes 4 and 8. However, although there was a significant main effect of set size, there was no significant main effect of feature type or interaction between feature type and set size.

The results of this experiment suggest that like visual WM for color, visual WM for orientation appears to have a capacity of three to four items. However, it is possible that performance on the memory task was negatively influenced by the presence of variability in the irrelevant dimension. For example, even though the colors never changed in the orientation condition, the presence of orthogonal variation in the colors may have impaired performance. However, irrelevant variability was also present during the color condition, and accuracy in that condition was very similar to the accuracy levels observed in the previous experiments. The influence of variability in an irrelevant dimension is also addressed directly in Experiment 12.

The finding of the same capacity estimate for color and orientation suggests that this estimate reflects a general capacity limitation of visual WM for simple features rather than a specific memory limitation for color. However, it is possible that the capacity of visual WM for both orientation and color are special cases and do not reflect all simple features. Although this may seem unparsimonious, further research is necessary to determine whether visual WM capacity is different for features other than color and orientation.

Experiments 11–16: Visual WM for Multifeature Objects

Experiment 11

This experiment examined the relationship between WM capacity and stimulus complexity by assessing whether multifeature objects consume more visual WM capacity than single-feature objects. As in Experiment 10, the sample arrays consisted of colored, oriented bars. The observers were asked to remember only the colors of the bars in the *color* condition, only the orientations of the bars in the *orientation* condition, and both the colors and the orientations of the bars in the *conjunction* condition.[4] Thus, the conjunction condition required the observers to remember twice as many features as the color and orientation conditions at a given set size, and this situation provided an opportunity to distinguish between two possible forms of information storage in visual WM.

One possibility is that each feature is stored as a separate entry in visual WM, such that storing both the colors and the orientations of four objects would require as much capacity as storing just the colors or just the orientations of eight objects. If this were true, then accuracy in the conjunction condition at a given set size would be equal to performance in the feature conditions at twice that set size. A second possibility is that visual working memory stores the features of an object in an integrated object representation and that the capacity of visual WM is limited by the number

of objects rather than the number of features. This account would predict that the observers would be just as accurate in the conjunction condition as in the feature conditions. This prediction is similar to the observation of Duncan (1984) that observers can report two features of an object just as accurately as they can report a single feature.

One potential problem in testing memory performance for multifeature objects is that performance may be limited by perceptual errors in addition to memory capacity. Previous studies of the perception of conjunctively defined objects have shown that subjects may incorrectly miscombine features that belong to different objects, yielding impaired performance for conjunction stimuli even though the individual features are perceived correctly (Prinzmetal, Henderson, & Ivry, 1995; Treisman & Schmidt, 1982). To reduce the risk of these illusory conjunctions in the present experiment, we limited the set sizes to two, four, and six and ensured that the items were spaced at least 2° apart (center-to-center); these factors have previously been shown to reduce the occurrence of illusory conjunctions (Cohen & Ivry, 1989, 1991).

This experiment was identical to Experiment 10 with the three exceptions. First, smaller set sizes were used. Second, in addition to the color and orientation conditions, a conjunction condition was used in which either the color or the orientation of an item could change. Third, the number of trials per condition was reduced slightly to 32.

The results from Experiment 11 are summarized in Figure 6A. As in Experiment 10, performance was slightly less accurate for the orientation condition than for the color condition, and accuracy declined systematically with set size in both of these conditions. Accuracy for the conjunction condition was found to be almost exactly halfway between the accuracies of the color and orientation conditions at all set sizes. An ANOVA yielded a highly significant main effect of set size but no significant effect of condition and no significant interaction between condition and set size. These results indicate that observers can hold both the color and the orientation of an object in visual WM just as easily as they can hold just the color or just the orientation. At Set Size 4, for example, accurate performance in the conjunction condition required the observers to hold eight features in memory, but they were just as accurate in this condition as they were in the color and orientation conditions, in which the storage of only four features was required. These results support the proposal that objects are stored in an integrated manner in visual WM rather than as separate features, such that the capacity of visual WM is limited by the number of objects rather than the number of features.

---

[4] The condition in which the observers were required to remember both the colors and the orientations may not, strictly speaking, require the observers to conjoin the color and orientation information from a given object. However, the results of this experiment indicate that the features were indeed linked together, and this term also provides a link to analogous studies of features and conjunctions in the attention literature.
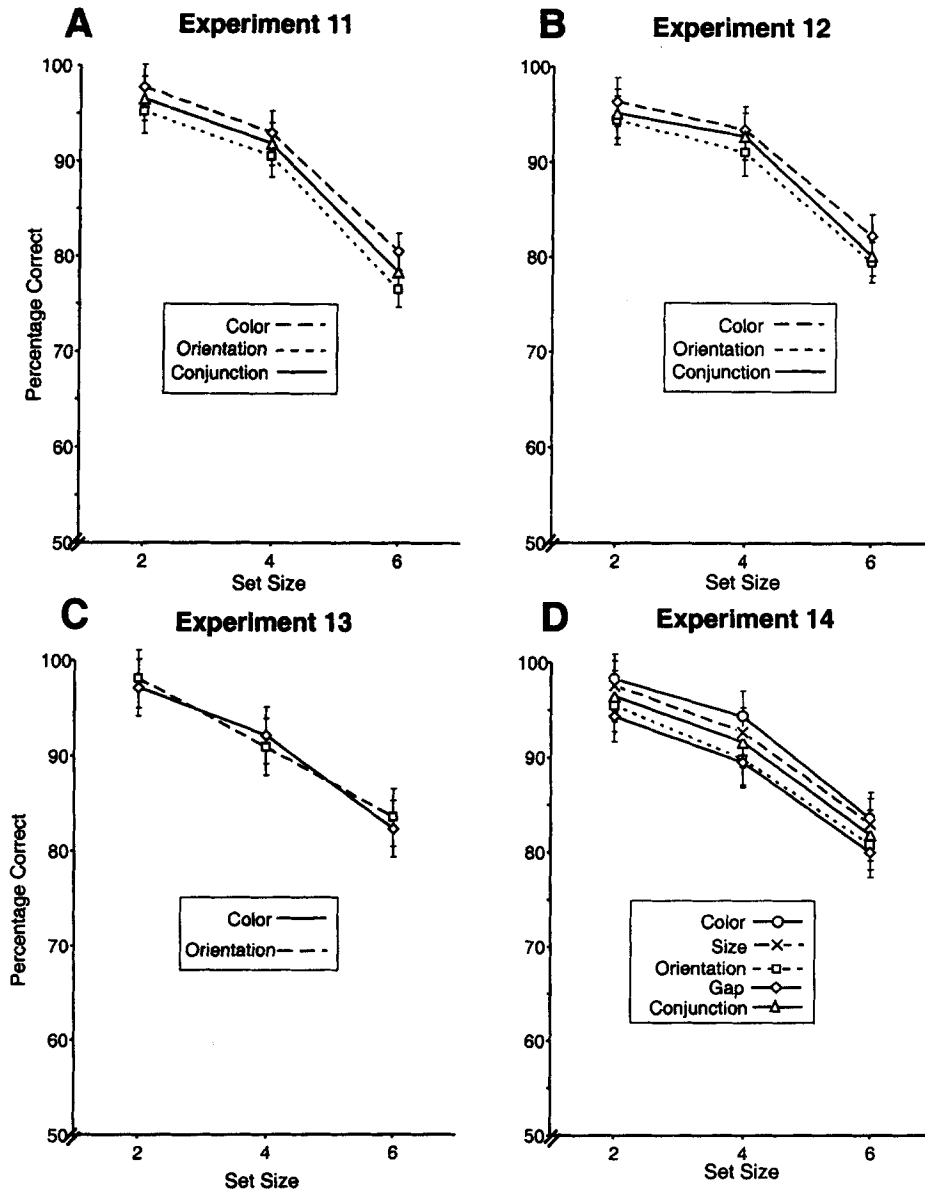
Figure 6. A: Mean accuracy for the color, orientation, and conjunction conditions of Experiment 11. B: Mean accuracy for the color (with all vertical orientations), orientation (with all black colors), and conjunction conditions of Experiment 12. C: Mean accuracy for the Gaussian color blobs and texture-defined orientations conditions used in Experiment 13. D: Mean accuracy for the color, size, orientation, gap-presence, and the conjunction conditions of Experiment 14.

Experiments 12 and 13

An alternative explanation of the results of Experiment 11 is that the observers could not selectively attend to the individual feature dimensions in the color and orientation conditions and therefore coded both the color and the orientation of the objects in these conditions even though this was not required by the task. If the observers could not avoid encoding both features, then the lack of impairment in the conjunction task could be explained without

postulating integrated object representations.[5] We conducted two experiments to test this alternative explanation. In both experiments, we determined whether performance in the color and orientation conditions would be improved by eliminating information in the other dimension. If performance in the color and orientation

---

[5] We thank Hal Pashler for suggesting this alternative explanation.

conditions of Experiment 11 was limited by inadvertent encoding of the irrelevant feature, then eliminating the irrelevant feature should lead to improved performance.

Experiment 12 examined this possibility by making all of the rectangles vertical in the color condition and by making all of the rectangles black in the orientation condition. This experiment was otherwise identical to the color and orientation conditions of Experiment 11. Participants also performed a conjunction condition that was identical to the conjunction condition in Experiment 11. As shown in Figure 6B, performance in the single-feature conditions was again identical to performance in the conjunction condition. Thus, performance is not enhanced when there is no variation in the irrelevant dimension.

It is possible that the observers in Experiment 12 encoded the irrelevant dimension even though it did not vary across objects. In the color condition, for example, the observers may have encoded the fact that a given bar was vertical even though all of the bars were vertical. Experiment 13 was therefore designed to completely eliminate information about the irrelevant dimension. We achieved this design by using colored blobs that lacked any local or global orientation in the color condition and by using texture-defined rectangles that lacked any color information in the orientation condition. As illustrated in Figure 1C, the blobs in the color condition were either red or green and were created by modulating the intensity of either the red or green signals on the video monitor to create a two-dimensional Gaussian luminance profile. For example, a red blob was created by beginning at the perimeter of the blob with the same gray as the background (defined by a constant set of red, green, and blue intensities) and then gradually increasing the red intensity for each concentric circle of the blob until the maximum red value was reached at the circle's center. The red intensity was increased according to a Gaussian function with a standard deviation of 0.34°, and the same green and blue intensities found in the background gray were also added to the red throughout the blob. A green blob was created in the same manner, except that the green intensity was modulated instead of the red intensity. The blobs can be thought of as the result of blurring an infinitesimally small red or green point by filtering out the high spatial frequencies; because a point has no defined orientation, the blob created by symmetrically filtering a point also contains no orientation information.[6]

We used texture differences to create colorless oriented bars in the orientation condition (see Figure 1D). Specifically, the background of the video monitor was set to white (62.34 cd/m²), and 45° black lines were drawn every 0.08° across the screen to create a background pattern. Horizontal and vertical bars were then created by changing the orientation of the lines to −45° within a 0.65° × 200° region. Thus, both the background and foreground regions in this condition contained the same luminance and color information and differed only in second-order texture characteristics.

As shown in Figure 6 (panels A, B, and C), performance was nearly identical (within 3%) for the colored blobs, the oriented textures, and the corresponding color and orientation conditions of Experiments 11 and 12. If performance in the color and orientation conditions of Experiments 11 and 12 was limited because the observers obligatorily encoded the irrelevant fea-

ture, then performance should have been substantially better in the colored blob and oriented texture conditions of Experiment 13 because half as many features were available to be coded in this experiment.[7] Thus, the best explanation of the results of Experiments 11–13 appears to be that the features of an object are stored together in a unified representation.

Experiment 14

To examine the generality of the equivalent performance in the conjunction and feature conditions of Experiment 11, in this experiment we tested conjunctions of four features. The objects in this condition varied in size (0.13° × 1.00° or 0.13° × 2.00°), color (red or green), orientation (horizontal or vertical), and the presence or absence of a black gap (0.26°) in the center of the object. Each feature was tested individually in a condition in which only that feature could change between the sample and test arrays. We also tested a quadruple conjunction condition in which any one of the four features could change between the sample and test arrays. This experiment was otherwise identical to Experiment 11.

As illustrated in Figure 6D, accuracy in the conjunction condition was almost exactly at the mean of the four feature conditions. The statistical analysis further indicated no significant main effect of condition or interaction between condition and set size. Thus, at a set size of four objects, the observers could retain 16 features in the conjunction condition just as well as they could retain 4 features in any of the single-feature conditions. This provides additional evidence for the storage of integrated object representations.

Experiment 15

Although the finding of equivalent performance for feature and conjunction conditions appears to indicate that integrated object representations are stored in visual WM, these results are also consistent with the possibility of completely independent storage systems for each feature dimension. In fact, when Magnussen et al. (1996) obtained similar results with near-

---

[6] The Gaussian intensity modulation used in this experiment did not take into account the nonlinearities of the video monitor, so these stimuli were only approximations of true Gaussian blobs. However, the radial symmetry and gradual fall-off of the function guaranteed that these blobs contained virtually no perceivable orientation information.

[7] It might be objected that there were four possible values of each feature dimension in Experiments 11 and 12, but there were only two values in Experiment 13, which might have distorted the results. Specifically, the greater homogeneity in Experiment 13 may have provided increased opportunities for the observers to group the stimuli together, thus increasing accuracy. For example, if five of six items in an array were red, the five items might be grouped into a large red grouping, potentially decreasing the WM demands of the array. However, this should have produced greater accuracy in Experiment 13, working against the hypothesis of integrated object representations. In addition, Experiment 14 uses colored bars in combination with two possible values for each dimension, and the results were comparable to those in Experiments 11 and 12, in which there were four possible values for each dimension.

threshold features, they concluded that the features were stored independently.

To distinguish between independent storage and integrated object representations, we examined color–color conjunctions in Experiment 15. As illustrated in Figure 1E, each object in the conjunction condition consisted of a small colored square inside a larger colored square. The observers were required to remember both colors of each object in this condition because either a small square or a large square might change color between the sample and test arrays. It is implausible to suppose that separate memory systems exist for the inner and outer squares, and the hypothesis of separate memory systems for different features would therefore predict a large decrement in performance for this condition compared with the single-feature conditions. In contrast, the integrated-object hypothesis predicts that performance should be just as accurate in this color–color conjunction condition as in the single-feature conditions.

The most obvious way to assess performance for the individual features of these squares would be to ask the participants to remember just the inner square colors in one condition and just the outer square colors in a second condition. However, it seemed likely that the observers would be unable to selectively store just the inner square colors or just the outer square colors, and we therefore used simpler stimulus configurations to measure performance for the individual features. In one condition, we presented only the inner squares, and in another condition, we presented only the outer squares (the outer squares were solid rather than having a gap in the center; pilot experiments

indicated that this did not make a difference). It might be objected that these single-feature conditions are less perceptually demanding than the conjunction condition; however, this would tend to increase the relative difficulty of the conjunction condition, which would work against the integrated-object hypothesis.

The inner and outer squares in this experiment were 0.65° and 1.30° wide, respectively. Pilot studies indicated that the colors of the squares had to be very highly discriminable to avoid perceptual limitations, and the set of colors was therefore changed to red ($x = .588$, $y = .370$; 30.80 cd/m^2), green ($x = .321$, $y = .566$; 41.66 cd/m^2), violet ($x = .321$, $y = .154$; 8.44 cd/m^2), and blue ($x = .182$, $y = .144$; 20.11 cd/m^2). The inner and outer colors of a given square were always different from each other, and when the color of a square changed between the sample and test arrays, it changed to a color that was different from the colors of both the inner and outer squares of that object. The color changed for an inner square on 25% of trials, for an outer square on 25% of trials, and for neither on 50% of trials. This experiment was otherwise identical to Experiment 11.

The results of this experiment are shown in Figure 7A. As in Experiments 11 and 14, accuracy for the conjunction condition in this experiment was statistically indistinguishable from accuracy in the feature conditions. Specifically, there was a highly significant effect of set size, but the condition main effect and the condition X set size interaction were not significant. Thus, the observers were able to store twice as many colors in
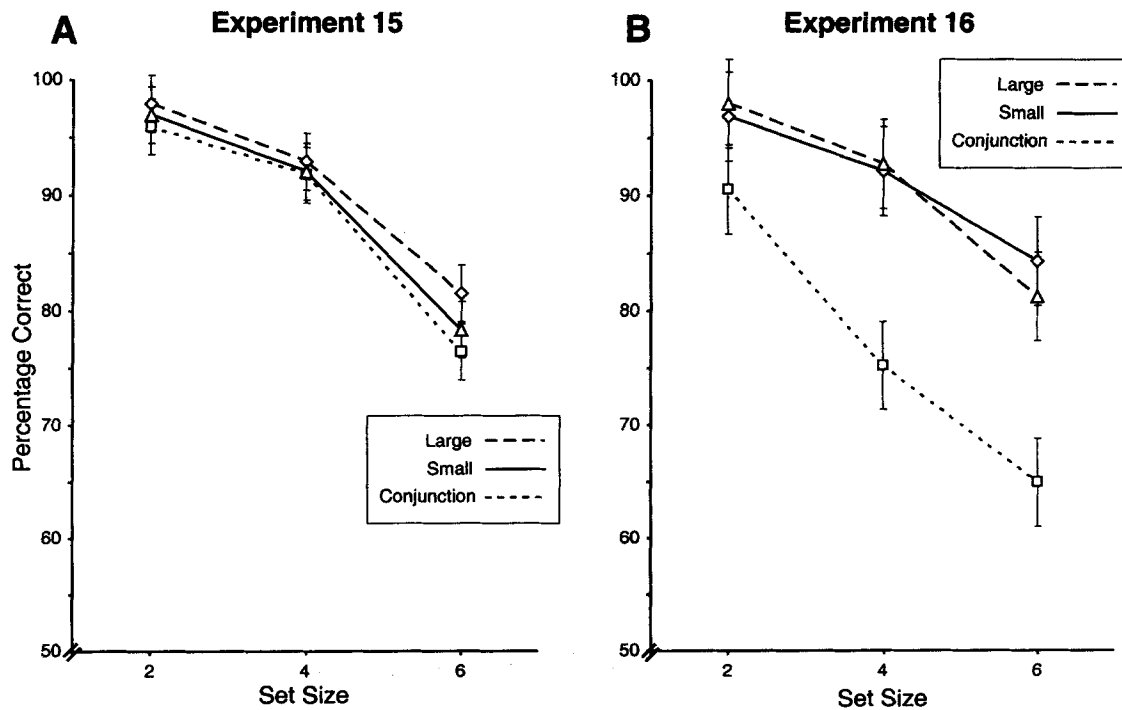

Figure 7. A: Mean accuracy for the large squares, small squares, and conjunction conditions of Experiment 15.
B: Mean accuracy for the large squares, small squares, and the conjunction conditions of Experiment 16.

memory in the conjunction condition than in the feature conditions, and performance was limited by the number of objects rather than the number of features.[8]

These results lead to something of a conundrum: If participants can store a large number of features together in a single object with no additional WM cost, why do they not treat the entire display as a single object, yielding virtually infinite memory capacity? The most likely answer is that the visual system is not able to group features together arbitrarily to form objects. In fact, in developing this experiment, we tested several different stimulus configurations that failed to produce equivalent performance in the feature and conjunction conditions. For example, we created rectangles by abutting two colored squares, but accuracy for these bicolor rectangles was worse than accuracy for the same number of single-colored objects. In other words, accuracy depended on the number of colored squares, not on the number of bicolor rectangular objects. Our subjective impression was that the rectangles could easily be viewed as either a single multicolored rectangle or two single-colored squares, whereas it was difficult to perceive the square-inside-square objects used in the present experiment as anything but a single object. We also tried using rectangles with alternating stripes of two colors, but performance was also much less accurate for these multicolored objects than for single-colored rectangles. In this case, the problem appeared to be that it was difficult to accurately perceive the colors of the thin, adjacent stripes. Thus, it is likely that the excellent memory performance we observed for within-dimension conjunctions in Experiment 14 is limited to cases in which the two features unambiguously form a single object and do not lead to perceptual errors.

Experiment 16

To examine this issue more formally, we conducted an experiment that was identical to Experiment 15 except that the inner and outer squares were shifted diagonally so that they no longer appeared to be parts of the same object (see Figure 1F). Because it is difficult to perceive the two shifted squares as a single object, performance in this experiment was expected to be impaired for the conjunction condition compared with the feature conditions (which were identical to the feature conditions in Experiment 15). Indeed, as shown in Figure 7B, accuracy in the conjunction condition was found to fall more rapidly than accuracy in the feature conditions as set size increased, leading to a significant main effect of condition and a significant Condition × Set Size interaction. We also conducted a statistical comparison of the conjunction condition from Experiment 15 with the conjunction condition of the present experiment, with a between-subjects factor of experiment and a within-subjects factor of set size. This ANOVA yielded a significant main effect of experiment, $F(1, 18) = 20.85, p < .001$, and a significant Experiment × Set Size interaction, $F(2, 36) = 5.82, p < .01$, indicating that shifting the inner squares relative to the outer squares increased the capacity required to store these bicolor objects.

This experiment rules out two possible problems with the preceding conjunction experiments. First, it is possible that participants in Experiment 15 were able to perform the color–color conjunctions accurately by remembering a single difference feature that reflected the difference in color between the inner and outer squares. If this were possible, then it should have also allowed accurate performance in the conjunction condition of Experiment 16 (and in the pilot experiments with bicolored rectangles), and the poor performance observed for the conjunction condition of this experiment makes this interpretation unlikely. Second, the statistical comparison between the conjunction conditions of Experiments 15 and 16 provides positive evidence of improved performance when two features are integrated into a single object. The previous experiments relied on the lack of a difference between feature and conjunction conditions as evidence of integrated object representations, and it is dangerous to rely solely on null results. Thus, the comparison of Experiments 15 and 16 provides a statistically significant interaction as evidence for the integrated-object hypothesis.

Discussion

WM Capacity for Features

It might seem like a trivial matter to measure the storage capacity of visual WM for simple features, but in fact 10 experiments were necessary to ensure a reasonably pure and general estimate. The reason for this is that the performance of almost any interesting task can potentially be influenced by limitations arising in several different types of processes, and even a paradigm that was designed specifically for the purpose of estimating visual WM capacity must be thoroughly tested for inadvertent contributions from other cognitive systems. We therefore attempted to rule out contributions from nonvisual WM systems (Experiments 1–3), from decision processes that presumably operate on the output of WM (Experiments 4–6), and from encoding processes that presumably provide the input to WM (Experiments 7 and 8). These experiments appeared to rule out contributions from the processes that precede and follow visual WM as well as contributions from verbal WM, making it possible to provide a relatively pure estimate of visual WM capacity. We also made some initial attempts at establishing the generality of our estimate of visual WM capacity (Experiments 9 and 10), and these experiments indicated that visual WM capacity does not vary greatly as a function of the retention interval or as a function of the type of feature being stored. Together, these 10 experiments consistently indicated that the average amount of information that can be stored in visual WM is equivalent to three to four objects, even though the objects were very simple and the observers were required to remember only a single feature of each item. It must be noted, however, that we are not claiming that three to four high-fidelity representations are necessarily stored by all observers on all trials. Indeed, it is very likely that amount of information stored in visual WM varies across participants and across trials, and it is also possible that large arrays are stored with a low-fidelity representation of every item in the array rather than a high-fidelity representation of a subset of the items. In addition, just as experienced participants

---

[8] A recent study reported a failure to find equivalent performance between the color–color conjunction condition and the individual feature conditions (Wheeler & Treisman, 1999). To ensure that the present results are reliable, we have replicated Experiment 15 with a new set of 10 participants and found the same pattern of results reported here.

may increase their apparent verbal WM capacity by using a variety of chunking strategies, it is likely that experienced participants may also be able to increase their apparent visual WM capacity.[9] However, the capacity of visual WM appears to be rather consistent at about 3–4 items in naïve participants.

Our capacity estimate of three to four items is fairly close to the value of four to five items that was estimated from the whole-report results of Sperling (1960). It is also very close to the capacity that is typically observed for verbal material when rehearsal is prevented (Baddeley, Lewis, & Vallar, 1984; Murray, 1968). Our specific estimate of visual WM capacity is based on the relatively simple model of task performance that was developed by Pashler (1988), and it is likely that performance is influenced by extraneous factors that are not anticipated by this model (e.g., grouping of identical items). However, even if our capacity estimate is slightly distorted, the present experiments clearly indicate that the number of simple visual items that can be simultaneously held in visual WM is remarkably small.

This highly restricted storage capacity seems contrary to the rich detail that we experience in the immediate visual environment, an environment filled with a multitude of objects that appear stable even as we move our eyes and our attention. If only a few items' worth of information can be retained between saccades and blinks, how can we explain the apparent richness and stability of our visual experience? Recent studies of change blindness have provided an answer to this paradox by showing that our experience of a rich and stable world is a result of the stability of the external environment rather than being a result of a rich and stable visual WM representation. In change blindness experiments, participants are typically presented with a flickering picture of a real world scene (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). One object within the scene is visible for one presentation of the scene, disappears for the next, and then subsequently reappears, and the flickering of the display between presentations eliminates low-level transients at the location of the appearing and disappearing object. The results of change blindness studies have suggested that very little of the visual environment is retained in visual memory from one moment to the next because participants typically require many cycles of a scene to detect the appearing and disappearing object, even if it is quite large. One explanation of these results is that it is unnecessary to maintain representations of all objects within the immediate visual environment because the physical objects themselves can serve as memory, making it unnecessary to retain representations of all objects in memory. The present experiments complement the change blindness studies by providing a precise quantification of visual WM capacity and by showing that the capacity is small even for very simple stimuli.

WM Capacity for Conjunctions

The present results also provide an initial step toward systematically evaluating the influence of complexity on visual WM capacity. Remarkably, the limited forms of complexity that we have examined appear to have no influence on WM capacity. Participants could retain multifeature objects in WM just as easily as single-feature objects, allowing as many as 16 features to be retained when distributed across four objects. These results indicate that the capacity of visual WM is determined by the number

of objects rather than the number of features, which further suggests that features are stored in WM within integrated object representations (for a similar conclusion, see Walker & Cuthbert, 1998). These results could also be explained by the existence of completely independent memory systems for each feature without any binding of features into object representations, but there are at least two factors that suggest that this is not the best explanation of the present results. First, the potential repetition of features within the stimulus arrays in this paradigm made it necessary for participants to code the locations of the features, and the separate features of a multifeature object were therefore bound together at least minimally by virtue of their common spatial location. Second, the finding of equivalent performance for color–color conjunctions relative to simple features in Experiment 15 provides direct evidence against independent feature storage in this paradigm. This conclusion contrasts with the conclusions drawn by Magnussen et al. (1996), who provided evidence for the independent storage of near-threshold features. This result suggests that near-threshold features may be stored at an earlier stage of the visual system that represents information about different features independently but with high fidelity.

These results dovetail with studies of object-oriented attention. For example, Duncan (1984) found that observers could report two features of a given object (e.g., the texture and orientation of a line) just as easily as they could report one feature, but accuracy was significantly impaired when the observers were required to report features of two separate objects (e.g., the orientation of a line and the size of a box). Similarly, Egly et al. (1994) cued attention toward one part of a rectangle and found that attention spread throughout the entire rectangle. Thus, these previous studies indicate that attention is directed to entire objects rather than to independently represented features (at least under certain conditions), and the present study indicates that integrated object representations are also stored in visual WM. Together, these findings suggest that perceptual processes are responsible for creating integrated object representations, which are then stored in visual WM.

In the present study, we examined conjunctions of only two or four features, and it is not known whether similar effects would be observed with larger numbers of features. It seems likely that the relationship between features and WM capacity will be different for two different classes of features, namely, *properties* and *parts* (see Treisman, 1986). The features that we have explored in this study have all been properties such as color that extend throughout the entire object (with the possible exception of the gaps used in Experiment 14), and it is likely that very different results would be obtained if spatially segregated parts were used as features. With multipart objects, each part may be stored in WM as a separate subobject, such that only one 4-part or two 2-part objects could be stored simultaneously in WM. This situation may be difficult to test, however, because it is likely that arrays of multipart objects would be very perceptually demanding, making it difficult to independently assess the effects of set size variations on perception

___

[9] Indeed, we have noticed that the visual WM capacity of laboratory personnel appears to increase over time when they frequently serve as pilot participants.

and WM. It is also possible that familiar multipart objects such as letters, digits, and perhaps faces could be stored in WM as single items, just like the single-feature objects studied here, and it seems likely that this possibility could be examined fairly easily given the ease with which such objects are perceived.

A Potential Neural Storage Mechanism

In this final section, we discuss a neural storage mechanism that could explain the general pattern of results observed in this study. Specifically, several investigators have proposed that synchronized neural firing could be used to bind together the widely separated neurons that code the individual features of a given object during visual object identification (e.g., Gray, König, Engel, & Singer, 1989; Hummel & Biederman, 1992; Niebur, Koch, & Rosin, 1993; von der Malsburg, 1996), and we and others have proposed that a similar mechanism might also be used to bind features in WM (Hummel & Holyoak, 1997; Lisman & Idiart, 1995; Luck & Beach, 1998). According to this storage scheme, one component of the storage of an object in WM is the use of elevated firing rates in the neurons that code the object's features. This is a relatively uncontroversial proposition that has been verified in both monkey single-unit recordings (e.g., Fuster & Jervey, 1981; Miller, Erick-

son, & Desimone, 1996) and human functional neuroimaging studies (e.g., Cohen et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997). However, elevated firing rates alone are not sufficient for accurate coding, especially when multiple, conjunctively defined objects must be coded. As illustrated in Figure 8A, for example, it would be simple to represent a red vertical bar in WM by means of sustained firing in red-selective and vertical-selective neurons. As shown in Figure 8B, however, this simple approach fails when a red vertical bar and a green horizontal bar are represented simultaneously because it is impossible to distinguish between a memory for red horizontal and green vertical and a memory for red vertical and green horizontal. Thus, a second component is needed to bind together the different features of an object in WM.

This binding problem could be partially solved by using neurons that code only a small area of space, such that the neurons coding a given object would be implicitly bound by means of their similar receptive fields. However, this is unlikely to be a complete solution. First, this solution would be location-dependent, whereas visual WM representations appear to be at least partially location-independent (Irwin, 1992; Phillips, 1974). Second, visual WM storage appears to occur primarily in brain areas in which the
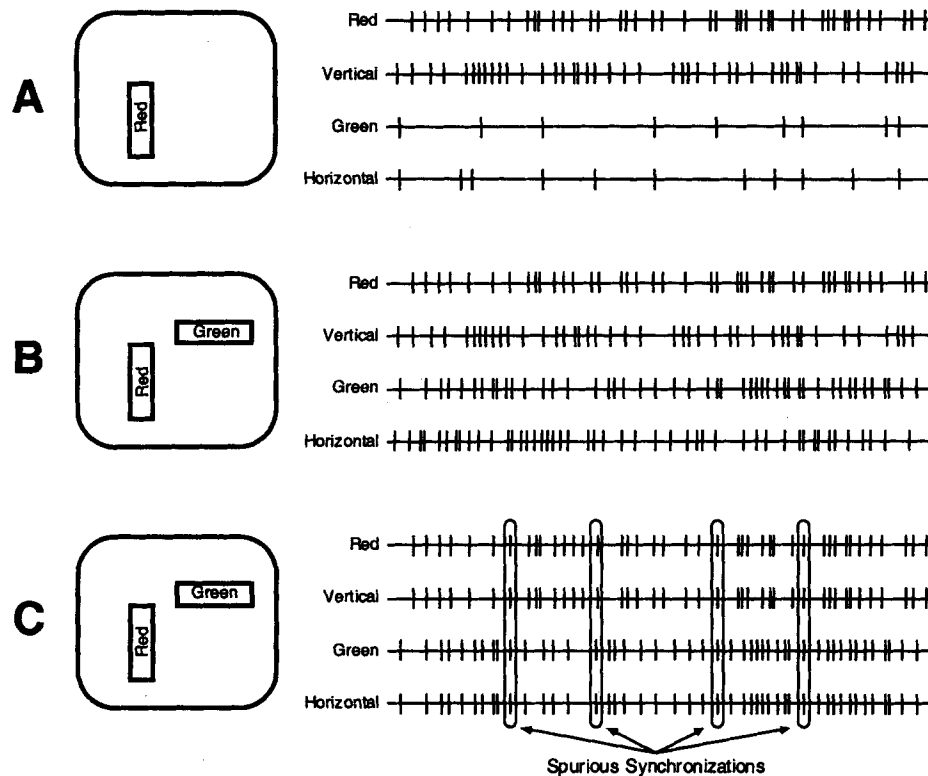

Figure 8. A: Example of the use of synchronized neural firing to code combinations of features in working memory. The responses of four individual neurons are shown, with separate neurons coding red, green, horizontal, and vertical. Each vertical mark represents a neural response (i.e., an action potential), and the horizontal axis represents the passage of time. B: Example of the coding scheme necessary for two objects. C: Example of accidental correlations that may occur when multiple items are stored simultaneously. The spurious synchronizations will occur more frequently as more items are stored.

neurons have very large receptive fields (Fuster & Jervey, 1981; Miller & Desimone, 1994; Miller et al., 1996; Miyashita & Chang, 1988), and these neurons would be ill-suited for such a local representation scheme. The binding problem could also be partially solved by using neurons that code conjunctions of features, thus eliminating the need for a special binding mechanism. This is implausible, however, because the vast number of possible combinations of features that could potentially be held in WM would lead to a combinatorial explosion in the number of neurons that would be required for this type of localist coding (Feldman, 1985). Thus, some type of special binding mechanism is necessary to represent multiple objects in WM without substantial crosstalk between the different representations. Synchronized firing provides this additional binding mechanism.

The general principle behind synchronization-based binding is that the features of an object can be bound together by causing the neurons that represent those features to fire synchronously,[10] forming what Hebb (1949) termed a cell assembly. The neurons within a cell assembly fire synchronously with each other but asynchronously with respect to other cell assemblies. This process is illustrated in Figure 8C, which shows that the neurons coding red and vertical fire synchronously with respect to each other but asynchronously with respect to the neurons coding green and horizontal. In this manner, it is possible to link red and vertical and to link green and horizontal while isolating these two cell assemblies from each other.

As noted by Mozer, Zemel, Behrmann, and Williams (1992), the essence of this type of representational scheme is that each neuron has two output values, one that indicates the extent to which the feature coded by the neuron is present and another that indicates which object is being coded by the neuron. There are, in principle, several ways in which a neuron's output could have two values, and it is therefore important to consider why so many investigators have proposed the use of firing rate to code which features are present and synchronization to indicate which objects are associated with which features. The use of firing rate is uncontroversial and is supported by many studies showing precise interrelationships between neural firing rates and psychophysical performance (see, e.g., Graham, 1989; Newsome, Britten, & Movshon, 1989). Synchronization is a newer concept, but there are both empirical and theoretical reasons to believe that it may be an important parameter in neural information coding. First, several studies have shown that neurons become synchronized under exactly the conditions that would be expected to lead to binding, such as when two bars move together and are perceptually linked by the Gestalt principle of common fate (see reviews by Singer et al., 1997; Singer & Gray, 1995; von der Malsburg, 1996). Second, as discussed by Singer et al. (1997), precisely synchronized inputs to a postsynaptic neuron make the neuron more likely to fire (and to fire rapidly) compared with asynchronous inputs. The particular effectiveness of synchronized signals may allow a synchronized set of neurons to be more easily maintained in WM via positive feedback loops: As long as a set of neurons is synchronized, they will be more effective at eliciting responses in the cells to which they project, which may in turn reinforce the responses of the synchronized cells via recurrent feedback connections. The effectiveness of synchronized signals may also cause the information coded by a synchronized cell assembly to be read out from mem-

ory as a unit rather than as a set of separately accessed features. Thus, synchronization is a plausible candidate for representing which objects are coded by which neurons and, by extension, which features are present in which objects.

When information is represented by means of synchronized firing, multiple objects can be represented in parallel, but interference will occur if too many objects are coded in parallel. This is a consequence of accidental synchronizations, as illustrated in Figure 8C. When multiple objects are coded simultaneously, the neurons coding different objects may occasionally fire at the same time, simply by chance. These accidental synchronizations reduce the isolation of the different object representations, causing them to become degraded (e.g., when vertical-selective and horizontal-selective neurons fire simultaneously, orientation coding becomes ambiguous). The more objects that are represented simultaneously, the greater is the probability of accidental correlations.[11] This model can therefore explain why the storage capacity of WM is limited to a small number of objects. In fact, the known characteristics of binding in the visual system appear to be consistent with the coding of a maximum of four to six concurrent objects (Hummel & Holyoak, 1997), which fits well with our estimate of visual WM capacity.

This hypothetical storage mechanism can also explain why WM storage was not degraded in the present study when each item was defined by multiple features (i.e., in the conjunction conditions). Specifically, the probability of accidental synchronizations depends on the number of concurrently represented objects and not on the number of neurons that represent each object. That is, because all of the neurons that form a cell assembly fire at the same time, increasing the number of neurons in a cell assembly does not increase the probability that the assembly will fire at the same time as another cell assembly. Thus, the synchronization model can explain both the limited number of objects that can be stored concurrently in WM and the apparently unlimited number of features that can be included in each object.

Although synchronization-based binding appears to explain the observed performance for the conjunction tasks, it is less obvious that it can explain the finding that WM capacity is also very small for simple features, which would not seem to require synchronized firing. However, there are several factors that suggest that synchronization may also be important for storing simple features in WM. First, as discussed above, synchronization may be useful for creating more potent responses in postsynaptic neurons and for maintaining a representation over time, and these benefits of synchronization would apply to single-feature objects as well as multifeature objects. Second, in both the natural environment and

---

[10] Many electrophysiological studies of feature binding have examined oscillations in field potentials or multiple-unit responses rather than synchronization of single-unit responses. However, these studies are based on the premise that large-scale oscillations reflect local synchronizations. Thus, we use the term synchronization to refer to findings from studies of both large-scale oscillations and small-scale synchronizations.

[11] The probability of accidental synchronizations could be reduced by reducing the rate at which each cell assembly fires. However, a lower firing rate may increase the amount of time required to detect that a cell assembly is present (i.e., that an object is present in WM), and it may also decrease the ability of a cell assembly to maintain itself over time.

the experimental paradigm used in this study, it is sometimes necessary to represent two objects that are both defined by the same simple feature, and this might be achieved by creating separate cell assemblies in which the feature-coding neurons are linked with location-coding neurons. To use a different parlance, even single-feature objects must be represented as individuated tokens for most tasks, and this situation may be achieved by means of synchronized firing. Third, feature-selective neurons typically have very broad tuning curves, and the accurate representation of a feature is therefore based on the pattern of activation across a set of broadly tuned neurons with different peak sensitivities. For example, an orientation of 15° is not represented by the activation of a set of narrowly tuned cells that respond only to orientations of 15° but is instead represented by the pattern of activity over a large set of neurons that have different but overlapping tuning curves and fire at different rates (see Hinton, McClelland, & Rumelhart, 1986; Luck, Girelli, McDermott, & Ford, 1997; Treisman & Gormican, 1988). For this reason, synchronized firing might be crucial for linking together the multiple neurons that are used to precisely code a simple feature.

In summary, synchronous neural firing provides a plausible neurophysiological mechanism for binding together the features of an object in visual WM, and can explain the two most significant results of the present study: (a) the finding that only 3–4 objects can be stored concurrently in visual WM and (b) the finding that multifeature objects do not require more WM capacity than single-feature objects. A more detailed model would be necessary to go beyond these qualitative explanations, and this simplistic version of the synchronization concept probably cannot account for all of the results presented here (such as the unimpaired performance observed for color–color conjunctions). However, it may provide a more specific and testable account of WM storage than traditional concepts such as resource limitations and interference.

References

Abelson, A. R. (1911). The measurement of mental ability of "backward" children. *British Journal of Psychology, 4,* 268–314.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Averbach, E., & Coriel, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal, 40,* 309–328.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science, 7,* 25–31.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36A,* 233–252.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Clarendon.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. VIII, pp. 47–90). New York: Academic Press.

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 521–539). Hillsdale, NJ: Erlbaum.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). World length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Binet, A., & Simon, T. (1909). L'intelligence des imbeciles. *Annals of Psychology, 15,* 1–147.

Blankenship, A. B. (1938). Memory span: A review of the literature. *Psychological Bulletin, 35,* 1–25.

Bolton, E. B. (1931). The relation of memory to intelligence. *Journal of Experimental Psychology, 14,* 37–67.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26,* 467–482.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

Calkins, M. W. (1898). A study of immediate and delayed recall of the concrete and of the verbal. *Psychological Review, 5,* 451–456.

Cermak, G. W. (1971). Short-term recognition memory for complex free-form figures. *Psychonomic Science, 5,* 209–211.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 650–663.

Cohen, A., & Ivry, R. B. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 891–901.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386,* 604–608.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386,* 608–611.

Cowan, N., Saults, J. S., & Fristoe, N. M. (1999). *Estimation of visual working memory capacity: An improved formula.* Manuscript submitted for publication.

Craik, F. I., Gardiner, J. M., & Watkins, M. J. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning & Verbal Behavior, 9,* 554–560.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50,* 291–323.

Di Lollo, V., & Dixon, P. (1988). Two forms of persistence in visual information processing. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 671–681.

Dixon, P., & Shedden, J. M. (1993). On the nature of the span of apprehension. *Psychological Research, 55,* 29–39.

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

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71,* 43–51.

Feldman, J. (1985). Connectionist models and parallelism in high level vision. *Computer Vision, Graphics, and Image Processing, 31,* 178–200.

Frick, R. W. (1988). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology, 79,* 289–308.

Fuster, J. M., & Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science, 212,* 952–954.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

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

Guilford, J. P., & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology, 36*, 621–628.

Hebb, D. O. (1949). *Organization of behavior: A neuropsychological theory*. New York: Wiley, Inc.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

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

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and Performance XVI* (pp. 125–155). Cambridge, MA: MIT Press.

Jacobs, J. (1887). Experiments on "prehension." *Mind, 12*, 75–79.

Jolicoeur, P., & Dell' Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138–202.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363*, 623–625.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning & Verbal Behavior, 1*, 153–161.

Klemmer, E. T. (1963). Perception of linear dot patterns. *Journal of Experimental Psychology, 65*, 468–473.

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7 +/− 2 short-term memories in oscillatory subcycles. *Science, 267*, 1512–1515.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 530–549.

Loftus, G. R., Johnson, C. A., & Shimamura, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance, 11*, 1–13.

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

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, NJ: Erlbaum.

Luck, S. J., & Beach, N. J. (1998). Visual attention and the binding problem: A neurophysiological perspective. In R. D. Wright (Ed.), *Visual attention* (pp. 455–478). New York: Oxford University Press.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology, 33*, 64–87.

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

Magnussen, S., & Greenlee, M. W. (1992). Retention and disruption of

motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 151–156.

Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research, 31*, 1213–1219.

Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 202–212.

Matin, E., & Drivas, A. (1979). Acuity for orientation measured with a sequential recognition task and signal detection methods. *Perception & Psychophysics, 25*, 161–168.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3–65.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*, 520–522.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16*, 5154–5167.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*, 68–70.

Mozer, M. C., Zemel, R. S., Behrmann, M., & Williams, C. K. (1992). Learning to segment images using dynamic feature binding. *Neural Computation, 4*, 650–665.

Murdock, B. B. J. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679–684.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*, 52–54.

Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research, 33*, 2789–2802.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 332–350.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 108–130.

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

Peatman, J. G., & Locke, H. M. (1934). Studies in the methodology of the digit-span test. *Archives of Psychology, 25*, 35.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.

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

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509–522.

Prinzmetal, W., Henderson, D., & Ivry, R. B. (1995). Loosening the constraints on illusory conjunctions: The role of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1362–1375.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849–860.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 419–425.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22,* 261–273.

Shaw, M. L. (1982). Attending to multiple sources of information. I: The integration of information in decision making. *Cognitive Psychology, 14,* 353–409.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301–305.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: Necessity, signature, and detectability. *Trends in Cognitive Sciences, 1,* 252–261.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18,* 555–586.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7,* 337–356.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (11, Whole No. 498).

Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics, 22,* 1–11.

Treisman, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 35.31–35.70). New York: Wiley.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General, 123,* 146–160.

Vogel, E. K., Woodman, G. F., Eads, A. C., & Luck, S. J. (1998). Masking in visual working memory: Evidence for a limited-capacity encoding mechanism. *Cognitive Neuroscience Society 1998 Annual Meeting Abstract Program,* 84.

von der Malsburg, C. (1996). The binding problem of neural networks. In R. Llinás & P. S. Churchland (Eds.), *The mind-brain continuum* (pp. 131–146). Cambridge, MA: MIT Press.

Walker, P., & Cuthbert, L. (1998). Remembering visual feature conjunctions: Visual memory for shape-colour associations is object-based. *Visual Cognition, 5,* 409–455.

Wheeler, M., & Treisman, A. (1999, April). *Location and binding in visual working memory.* Paper presented at the Cognitive Neuroscience Society, Washington, D.C.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning & Verbal Behavior, 2,* 440–445.

Wilson, J. T., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5,* 249–255.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (in press). Visual search remains efficient when visual working memory is full. *Psychological Science.*