

What are the units of visual short-term memory, objects or spatial locations?

DAEYEOL LEE

University of Rochester, Rochester, New York

and

MARVIN M. CHUN

Vanderbilt University, Nashville, Tennessee

We investigated whether the capacity of visual short-term memory (VSTM) is defined by number of objects or number of spatial locations. Previous work is consistent with either alternative. To distinguish these factors, we used overlapping stimuli that allowed us to independently manipulate the number of spatial locations while holding constant the number of objects and features to be encoded (Duncan, 1984; Vecera & Farah, 1994). In Experiment 1, the number of spatial locations had no effect on VSTM, suggesting that VSTM is object based. Experiments 2 and 3 ruled out alternative explanations based on perceptual segregation difficulty or decision noise factors. Our results provide additional support to Luck and Vogel's (1997) demonstration that integrated objects form the units of VSTM capacity.

Visual short-term memory (VSTM) allows for temporary storage of visual information that can be distinguished from iconic memory (Sperling, 1960) and verbal short-term memory (Baddeley, 1986; Baddeley & Hitch, 1974; Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974). An important question concerns the capacity of VSTM. A defining characteristic of VSTM is that the number of visual items that it can store is relatively small. As the number of visual items exceeds VSTM capacity, memory for these items deteriorates. But what is a visual item? What is the unit of VSTM capacity? Although this is not well understood, one can consider several likely candidates.

First, visual objects are composed of individual features, such as orientation and color. These individual features may form the units of VSTM. According to this view, the number of features necessary to characterize a single item determines how many items can be reliably stored in VSTM (Phillips, 1974). For example, VSTM may store four to five colors or a similar number of different shapes.

The second view of VSTM capacity goes beyond individual features. The visual world comprises meaningful objects, and each object comprises multiple features (e.g., a long, red, vertical line). Hence, the units of VSTM may include an aggregation of features that form particular objects, a visual correlate of "chunking" (Miller, 1956). According to this account, increasing the number of fea-

tures defining a given object would not influence the capacity of VSTM defined in terms of the number of objects (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001).

Finally, features and objects tend to occupy distinct locations in visual space, and many visual operations, such as target selection and saccadic eye movements, are spatially directed. Moreover, spatial locations are an important attribute in VSTM. Observers can remember a limited number of spatial locations in working memory (Jonides et al., 1993; McCarthy et al., 1994), and this may be separable from working memory for the identity or shapes of objects (Funahashi, Bruce, & Goldman-Rakic, 1989; Jiang, Olson, & Chun, 2000; McCarthy et al., 1996; Simons, 1996; Smith & Jonides, 1997; Wilson, Scalaidhe, & Goldman-Rakic, 1993). Thus, there is a reasonable possibility that the number of spatial locations occupied by relevant visual stimuli may be a limiting factor.

Recently, Luck and Vogel (1997) provided intriguing evidence that the capacity of VSTM can be understood in terms of integrated visual objects rather than individual features. In their study, participants made a same-different judgment after seeing two temporally separated displays of multiple visual objects. On half of the trials, the two displays were identical, whereas in the remaining trials they differed by a single feature. They found that the accuracy of response was determined by the number of objects in the display, with performance beginning to drop as the number of objects increased beyond four. Interestingly, the number of features defining each object was not relevant, because participants were equally good at retaining objects defined by a conjunction of four features as they were for single-feature objects. These results exclude the possibility that VSTM capacity is defined by the number of features.

One limitation of Luck and Vogel's (1997) study was that it could not be determined whether the capacity of

We thank Steve Luck, Tom Sanocki, and an anonymous reviewer for their helpful suggestions on the manuscript, and Yuhong Jiang for useful discussions. This study was supported in part by McDonnell-Pew Cognitive Neuroscience Grant 99-27 from the James S. McDonnell Foundation (D.L.) and NSF Grant BCS-9817349 (M.M.C.). Correspondence should be addressed to D. Lee, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627 (e-mail: dlee@cvs.rochester.edu).

VSTM is defined by the number of objects or spatial locations. In all of their experiments, only one object appeared in any given location. This problem is reminiscent of an important debate in the visual attention literature: Does attention select objects or spatial locations? According to space-based accounts, attention is directed to a particular region in space, and visual stimuli in the attended location receive further processing that is not available for stimuli appearing in unattended regions (Eriksen & Yeh, 1985; Posner, 1980). In contrast, according to object-based accounts of attention, individual features are grouped into various objects without attention, and these integrated objects are the entities to which attention is directed (Duncan, 1984; Kanwisher & Driver, 1992; Vecera & Farah, 1994). Ordinarily, it is difficult to distinguish between space-based and object-based accounts of attention, because different objects usually occupy different regions in the visual environment. Nevertheless, many studies have shown that attention can be based on objects rather than spatial locations (Chun & Cavanagh, 1997; Driver & Baylis, 1989; Duncan, 1984; Kahneman, Treisman, & Gibbs, 1992; Kanwisher & Driver, 1992; Kramer & Jacobson, 1991; Moore, Yantis, & Vaughan, 1998; Neisser & Becklen, 1975; Vecera & Farah, 1994). As a classic example, Duncan exposed his participants to a brief presentation of a box and a line overlapping in the same region of space. The box and the line were each defined by two features. Duncan found that accuracy was higher when the participants were required to report two features that belonged to either the box or the line than when they were required to report one feature from the box and a second feature from the line. This deficit in performance was interpreted as evidence that there was a cost for attending to multiple objects, even when spatial location was held constant.

In the present study, we investigated further whether VSTM capacity is object-based or space-based. To address this issue, we adopted the same stimuli that have been used successfully in the selective attention literature to distinguish these factors (Duncan, 1984). As we have suggested above, these two alternatives have not been fully distinguished in past research.

EXPERIMENT 1

It is possible that VSTM capacity is defined by the number of spatial locations in which visual items are presented. In Luck and Vogel's (1997) study (see also Vogel et al., 2001), integrated objects occupied unique spatial locations. In addition, other studies have shown that working memory has a dissociable component for encoding spatial locations (Funahashi et al., 1989; Jonides et al., 1993; Smith & Jonides, 1997; Wilson et al., 1993). Thus, it remained possible that VSTM capacity is based on the number of spatial locations that are occupied by visual items. To test this, we adopted a method used by

Vecera and Farah (1994), who manipulated the number of spatial locations in which visual stimuli were presented without changing either the number of features or that of objects.

In a selective attention task, Vecera and Farah (1994) found that the reporting of features from two different objects was not influenced by whether these objects occupied the same location or not. Similarly, we reasoned that if the VSTM capacity was truly based on objects rather than spatial locations, performance would not be affected by changing the number of locations in which a fixed number of objects and features were presented. In the following three experiments, the sample array included three lines and three boxes. In half of the trials (the *together* condition), each stimulus item consisted of a pair of one line and one box superimposed. These three object pairs occupied a total of three spatial locations. In the remaining half of the trials (the *separate* condition), the lines and boxes occupied separate locations, occupying a total of six spatial locations. Changes could occur to any of the four (two line and two box) features with equal probabilities, so the participants were required to remember all of the features presented.

Method

Participants. Ten observers (4 males and 6 females; age, 19–20) were recruited to participate in this experiment. All participants reported normal or corrected-to-normal vision.

Stimulus. In all trials, the sample array included three lines and three boxes. In half of the trials (the *together* condition), each of three locations contained a superimposed line and box pair. In the other half of the trials (the *separate* condition), lines and boxes were presented in alternating locations in the imaginary circle (Figure 1), occupying six locations. The stimuli in the sample array were modeled after those used in the experiments of Duncan (1984) and Vecera and Farah (1994). The width of the box was always 0.72° of visual angle and its height was either 0.93° (short) or 1.29° (tall). In addition, the box had an opening (0.21° wide) in the center of either its left or its right side. The line was tilted 7.1° from the vertical either to the left (counterclockwise) or to the right (clockwise), and its length was always 1.65° . The line was either dashed or dotted, consisting of six dashes ($2.1' \times 15.0'$) or six squares ($6.4' \times 6.4'$). All visual stimuli were displayed on a computer monitor (32×24 cm) located approximately 50 cm away from the participant's eyes. The screen resolution was $1,024 \times 768$ pixels, so a single pixel corresponded to about $2.1' \times 2.1'$ of visual angle. Throughout the experiment, the background of the monitor was black, and the stimuli were arrayed on an imaginary circle ($\text{rad} = 2.86^\circ$) positioned at the center of the monitor.

Procedure. The sample array was presented for 200 msec. It was followed by an 800-msec delay before the test stimulus was presented. In the *together* condition, the test stimulus was a line and box pair, whereas in the *separate* condition, it was either a line or a box. The participants' task was to press one key when the test stimulus was identical to the item that had appeared in the same location in the sample array. They were instructed to press another key if the test stimulus was different. The test stimulus was identical to the corresponding item in the sample array in 50% of the trials. In the remaining 50% of the trials, the test stimulus differed from the corresponding item in the sample array on one of the four dimensions described above. The test stimulus remained visible until the

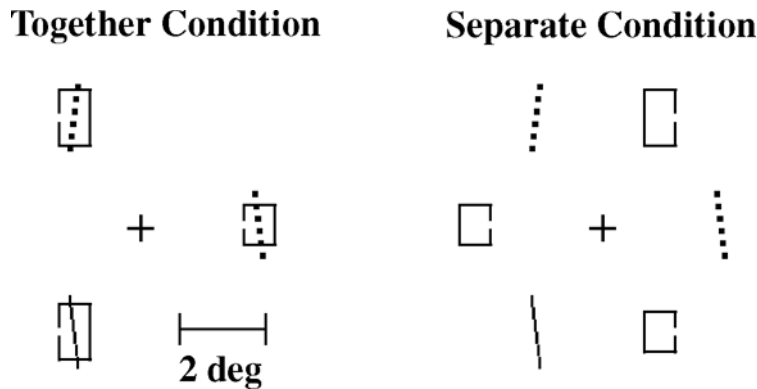


Figure 1. An example of the sample array used in the together (left) and separate (right) conditions of Experiment 1.

subject made a response. Each subject performed 10 trials in each condition as practice and then performed a total of 400 trials. In both cases, condition was randomized for each trial.

Statistical significance was determined with a paired *t* test. For binomial proportions, such as response accuracy, arcsine transformation (i.e., $\arcsin \sqrt{x}$) is recommended to increase the variance of proportions near zero or one (Snedecor & Cochran, 1989). Therefore, this transformation was applied to the proportion of correct responses for each condition. In general, this provided more conservative estimates (i.e., larger *p* values).

Results and Discussion

Response accuracy did not differ significantly for the together and separate conditions (Figure 2) [paired *t* test, $M = 0.007$, $SE = 0.016$, $t(9) = 0.452$, $p = .662$].¹ Because the number of locations was varied while the numbers of objects and features were held constant, these results suggest that number of spatial locations occupied by visual information does not influence VSTM performance.

However, at least two possible factors might have offset the potential effects of spatial location in Experiment 1. First, in the separate condition, the lines and the boxes were presented in separate spatial locations, thus making

it unnecessary to perceptually segregate these two objects in a given location as in the together condition. This possibility was tested in Experiment 2, where lines and boxes were presented in two different colors, enhancing perceptual segregation of the overlapping objects as has commonly been done in previous studies (DeSchepper & Treisman, 1996; Rock & Gutman, 1981; Tipper & Cranston, 1985). Second, because the test stimulus in the together condition always included two objects, this might have made the decision making (comparison) process more difficult in the together condition, thus offsetting any possible benefit obtained from the use of fewer spatial locations. In Experiment 3, this possibility was tested by presenting only one object as a test stimulus in both conditions. Using single probes minimizes decision effects in assessing VSTM capacity (Palmer, 1990).

EXPERIMENT 2

Method

A total of 8 new observers participated in this experiment (5 males and 3 females; age, 18–21). The stimuli and the procedure were identical to those used in Experiment 1, except that all the lines were presented in red, whereas all the boxes were presented in green.

Results and Discussion

As in Experiment 1, response accuracy was not significantly different for the together and separate conditions (Figure 2) [paired *t* test, $M = 0.018$, $SE = 0.02$, $t(7) = 0.933$, $p = .382$]. To the extent that our color manipulation eased the segregation of the lines and boxes, these results provide additional support to the conclusion that the number of locations is not an important factor in determining the VSTM capacity.

EXPERIMENT 3

Method

A total of 8 new observers were recruited (4 males and 4 females; age, 18–20). The stimuli were identical to those used in Experiment 1, except that only one object (a line or a box) was presented as the test stimulus. In both together and separate conditions, this object was identical to the same object presented in the same loca-

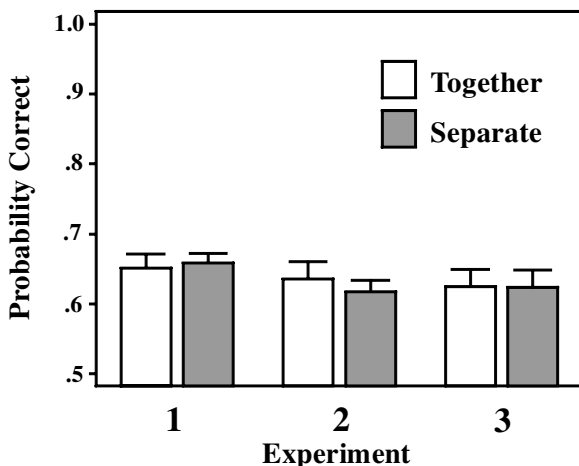


Figure 2. Response accuracy in Experiments 1–3. Error bars represent *SE*.

tion in 50% of the trials, whereas one feature of the object changed in the remaining trials.

Results and Discussion

Response accuracy was not significantly different for the together and separate conditions (Figure 2) [paired t test, $M = 0.007$, $SE = 0.010$, $t(7) = 0.066$, $p = .949$]. These results, combined with those of Experiment 1, suggest that the presence of an extraneous object in the test stimulus did not influence the outcome of these experiments. In sum, VSTM capacity does not seem to be influenced by the number of spatial locations occupied by visual information.

GENERAL DISCUSSION

The primary goal of this study was to determine whether the VSTM capacity can be described better in terms of the number of objects or in terms of the number of spatial locations. In three experiments, we manipulated the number of spatial locations independently of the number of features or that of objects. The results clearly demonstrated that the number of spatial locations does not influence the VSTM performance. These findings strengthen the claim that VSTM stores integrated objects (Luck & Vogel, 1997). Objects were confounded with spatial locations in Luck and Vogel's study, so it was important to rule out this factor, because a considerable amount of research has been devoted to working memory for spatial locations (Funahashi et al., 1989; Jonides et al., 1993; Wilson et al., 1993). Of course, we are not suggesting that working memory does not encode spatial location information; people as well as other animals are clearly capable of remembering the locations of objects and events. Our point is simply that the number of spatial locations per se does not influence VSTM capacity; rather, the number of objects occupying various locations is more important.

A fundamental question that remains unanswered is how to distinguish attentional selection processes from memory encoding and maintenance processes. Our results replicated selective attention manipulations to show that VSTM performance is unaffected by increases in the number of spatial locations that must be attended (Vecera & Farah, 1994). Nevertheless, our results cannot determine whether object-based attention and VSTM encoding and maintenance employ common mechanisms or whether selection mechanisms feed their output into later VSTM processes. Further empirical and theoretical work is needed to address this important issue.

Regardless of progress on the attention versus memory debate, research can proceed on another related, fundamental question. To the extent that VSTM is object based, it is important to understand what defines an object. In the present study, VSTM performance was not affected by placing the lines and boxes in different areas, reinforcing the assumption that these stimuli were treated as independent visual objects. However, objects encountered in the natural environment are usually more com-

plicated. Further insights into the nature of "object" can be found in a recent study by Xu and Potter (1999). They demonstrated that VSTM can encode multiple features only when these features are integrated into a single object defined by the same coherent boundary (e.g., a colored, oriented bar). When the features (such as orientation and color) are distributed into different "parts" (consider a Saturn-like object with an oriented line passing through a round colored ball, or, as another example, a colored beach ball with a black oriented stripe), then VSTM treats those orientation and color features as if they were from different objects. Thus, VSTM capacity is dependent on how perceptual and attentional mechanisms parse the visual input into different objects, suggesting a tight link between working memory maintenance and attentional selection processes.

REFERENCES

- BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.
- BADDELEY, A. D., & HITCH, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- CHUN, M. M., & CAVANAGH, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science*, **8**, 74-79.
- DESCHÉPPER, B., & TREISMAN, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 27-47.
- DRIVER, J., & BAYLIS, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 448-456.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.
- ERIKSEN, C. W., & YEH, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 583-597.
- FUNAHASHI, S., BRUCE, C. J., & GOLDMAN-RAKIC, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, **61**, 331-349.
- JIANG, Y., OLSON, I. R., & CHUN, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 683-702.
- JONIDES, J., SMITH, E. E., KOEPE, 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.
- KANWISHER, N., & DRIVER, J. (1992). Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science*, **1**, 26-31.
- KRAMER, A. F., & JACOBSON, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics*, **50**, 267-284.
- LUCK, S. J., & VOGEL, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390**, 279-281.
- MCCARTHY, G., BLAMIRE, A. M., PUCE, A., NOBRE, A. C., BLOCH, G., HYDER, F., GOLDMAN-RAKIC, P., & SHULMAN, R. G. (1994). Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proceedings of the National Academy of Sciences*, **91**, 8690-8694.
- MCCARTHY, G., PUCE, A., CONSTABLE, R. T., KRYSTAL, J. H., GORE, J. C., & GOLDMAN-RAKIC, P. (1996). Activation of human prefrontal cortex during spatial and nonspatial working memory tasks measured by functional MRI. *Cerebral Cortex*, **6**, 600-611.
- MILLER, G. (1956). The magical number seven plus or minus two: Some

- limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.
- MOORE, C. M., YANTIS, S., & VAUGHAN, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, **9**, 104-110.
- NEISSER, U., & BECKLEN, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, **7**, 480-494.
- PALMER, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 332-350.
- PASHLER, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, **44**, 369-378.
- PHILLIPS, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, **16**, 283-290.
- POSNER, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3-25.
- ROCK, I., & GUTMAN, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 275-285.
- SIMONS, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, **7**, 301-305.
- SMITH, E. E., & JONIDES, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, **33**, 5-42.
- SNEDECOR, G. W., & COCHRAN, W. G. (1989). *Statistical methods* (8th ed.). Ames: Iowa State University Press.
- SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General & Applied*, **74**, 1-29.
- TIPPER, S. P., & CRANSTON, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology*, **37A**, 591-611.
- 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., & LUCK, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 92-114.
- WILSON, F. A., SCALADHE, S. P., & GOLDMAN-RAKIC, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, **260**, 1955-1958.
- XU, Y., & POTTER, M. C. (1999). *Objects through the eyes of visual working memory*. Paper presented at the 7th Annual Workshop on Object Perception and Memory, Los Angeles.

NOTE

1. In Experiment 1, the probability of reporting a change was significantly higher in the separate condition than in the together condition for both correct (i.e., hit, $M = 0.641$ vs. 0.564) and incorrect (i.e., false alarm, $M = 0.261$ vs. 0.324) responses ($p < .05$). This suggests that the participants used different response criteria for these two conditions. However, d' was not significantly different [paired t test, $t(9) = 0.012$, $p = .99$]. In addition, this pattern of errors was not observed for Experiments 2 and 3.

(Manuscript received April 12, 1999;
revision accepted for publication July 17, 2000.)