

Dissociable neural mechanisms supporting visual short-term memory for objects

Yaoda Xu¹ & Marvin M. Chun¹

Using visual information to guide behaviour requires storage in a temporary buffer, known as visual short-term memory (VSTM)¹, that sustains attended information across saccades and other visual interruptions. There is growing debate on whether VSTM capacity is limited to a fixed number of objects^{2,3} or whether it is variable^{4,5}. Here we report four experiments using functional magnetic resonance imaging that resolve this controversy by dissociating the representation capacities of the parietal and occipital cortices. Whereas representations in the inferior intraparietal sulcus (IPS) are fixed to about four objects at different spatial locations regardless of object complexity, those in the superior IPS and the lateral occipital complex are variable, tracking the number of objects held in VSTM, and representing fewer than four objects as their complexity increases. These neural response patterns were observed during both VSTM encoding and maintenance. Thus, multiple systems act together to support VSTM: whereas the inferior IPS maintains spatial attention over a fixed number of objects at different spatial locations, the superior IPS and the lateral occipital complex encode and maintain a variable subset of the attended objects, depending on their complexity. VSTM capacity is therefore determined both by a fixed number of objects and by object complexity.

Behavioural studies have reported that VSTM can store up to about four objects independent of the number of features that must be encoded for each object³, indicating that VSTM capacity might be limited to a fixed number of objects³. However, new evidence shows that fewer than four objects are held in VSTM when object complexity increases^{4,5}, indicating that VSTM capacity is variable and modulated by the complexity of visual objects encoded⁵.

A network of brain areas supports VSTM (that is, visual working memory), including the frontal/prefrontal, parietal and occipital cortices^{6–19}. Although the frontal/prefrontal cortex shows increased activity with memory load⁹, a recent study indicates that the (superior) IPS might have the most important function in determining VSTM capacity for visual objects^{17,18}, because functional magnetic resonance imaging (fMRI) activations within this region correlated most strongly with the number of objects held in VSTM (see also a related electrophysiological study¹⁹). However, these recent findings do not address the fundamental debate over whether neural processing capacity is limited by a fixed number of objects or whether it varies with increasing visual object complexity^{2–5}. In addition, although the role of the frontal/prefrontal cortices in the control and maintenance of VSTM has been well described^{7–14}, the nature of VSTM object representation in occipital and other parietal regions is less well understood and deserves further investigation^{6,11,15,16}.

The occipital visual areas have been frequently reported to be involved in VSTM tasks^{9–13,16}, revealing that VSTM storage may be mediated by the same brain structures processing perceptual

information^{15,16}. Among the different visual areas, the lateral occipital complex (LOC) is important in visual object perception such that it exhibits higher activations for objects than for noise images^{20–22} and participates in successful visual object recognition²³. Whether LOC activity directly correlates with the number of objects successfully retained in VSTM, however, has not been thoroughly established.

Besides LOC, a higher activation for objects than for noise images has also been reported in the inferior IPS, which has been attributed to greater visual attention to objects than to noise images²². Although previous research has indicated that parietal attention mechanisms are involved in VSTM for spatial information^{12,24}, given that attention is essential in VSTM for objects as well, the role of the parietal attention mechanisms in maintaining visual objects requires more detailed investigation.

To address these issues, using fMRI, we examined the roles of the parietal and occipital cortices in VSTM for visual objects. In the first experiment, observers were required to remember either simple or relatively complex (but still easily discriminable) shape features from the same set of objects. Observers viewed one to four, or six black, elongated objects in a sample display and, after a brief delay, detected a shape feature change in a test display. The simple and the relatively complex (hereafter referred to as ‘complex’) shape features were defined as the presence or absence of the hole in the centre of the objects and object outlines, respectively (Fig. 1a).

To evaluate behavioural responses, we applied Cowan’s *K* formula²⁵ and transformed change detection accuracies to VSTM capacity estimates as a function of display set size. To evaluate fMRI responses, we used a region of interest (ROI) approach²² and extracted averaged fMRI responses from three functionally defined ROIs (Fig. 2a). The superior IPS ROI was defined as voxels whose activations correlated with VSTM capacity in a separate colour VSTM task (Supplementary Fig. 1a) and whose Talairach coordinates²⁶ matched those reported previously¹⁷. LOC and inferior IPS ROIs were defined as regions in the lateral and ventral occipital cortex and the inferior IPS region, respectively, revealing higher activations for objects than for noise images (Supplementary Fig. 1b–d)^{20–22}.

VSTM capacity was higher for simple than for complex features (Fig. 2b), with a significant interaction between set size and feature complexity ($n = 7$, $P < 0.001$, analysis of variance, two-tailed; this applies to all subsequent statistical tests except where noted). For simple shape features, *K* values increased linearly with set size ($P < 0.01$), differed among set sizes 2 to 6 ($P < 0.001$) and reached a plateau at set size 4 ($P > 0.18$ for the differences between set sizes 4 and 6). For complex shape features, *K* values did not increase with set size ($P > 0.14$) and reached a plateau at set size 2 ($P > 0.45$ for the differences among set sizes 2 to 6). The maximum VSTM capacity was therefore about four objects for simple shape features and only two objects for complex shape features.

fMRI activations in the superior IPS and LOC tracked the capacity

¹Department of Psychology, Yale University, New Haven, Connecticut 06520-8205, USA.

estimates in behavioural performance, increasing with set size for simple shape features ($P < 0.01$), but not for complex shape features ($P > 0.19$) (Fig. 2b). Although their activations differed among set sizes 2 to 6 for simple shape features ($P < 0.05$), they did not differ for complex shape features ($P > 0.54$), with significant interactions between set size and feature complexity ($P < 0.05$). Moreover, for simple shape features, the increase in activations with set size in these brain regions correlated with the increase in K values ($P < 0.01$). In contrast, regardless of feature complexity, activations in the inferior IPS increased from set sizes 1 to 4 ($P < 0.01$) and reached a plateau at set size 4 ($P > 0.17$ for the differences between set sizes 4 and 6), with no interaction between set size and feature complexity ($P > 0.19$) (Fig. 2b). Thus, whereas activations in the inferior IPS tracked a fixed

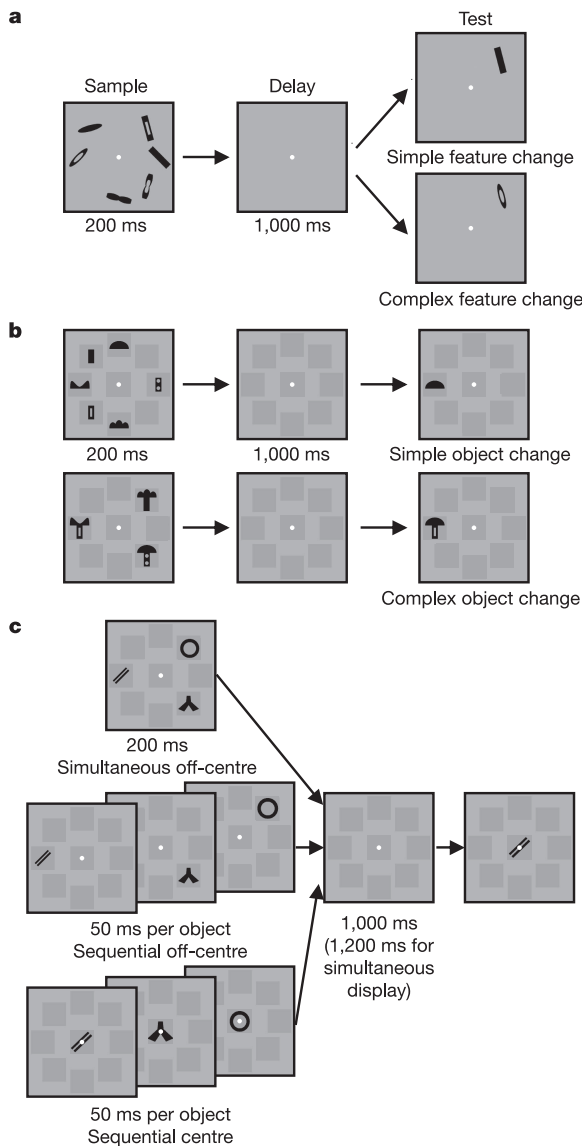


Figure 1 | Example trials from experiments 1, 2 and 4. **a**, In experiment 1, observers viewed the sample displays that consisted of a variable number of black, elongated shapes arranged in a circle, and after a brief blank delay they detected a possible change in the test display at the probed location. In the test display, either a simple feature (presence or absence of the hole in the centre) or a complex shape feature (shape outline) could change. **b**, Experiment 2 was identical to experiment 1, except that each complex object was constructed by attaching two simple objects together. **c**, In experiment 4, observers viewed a variable number of different black objects simultaneously, sequentially at different off-centre locations, or sequentially at the same centre location, and judged whether the probe object in the test display was present in the sample display (sequence).

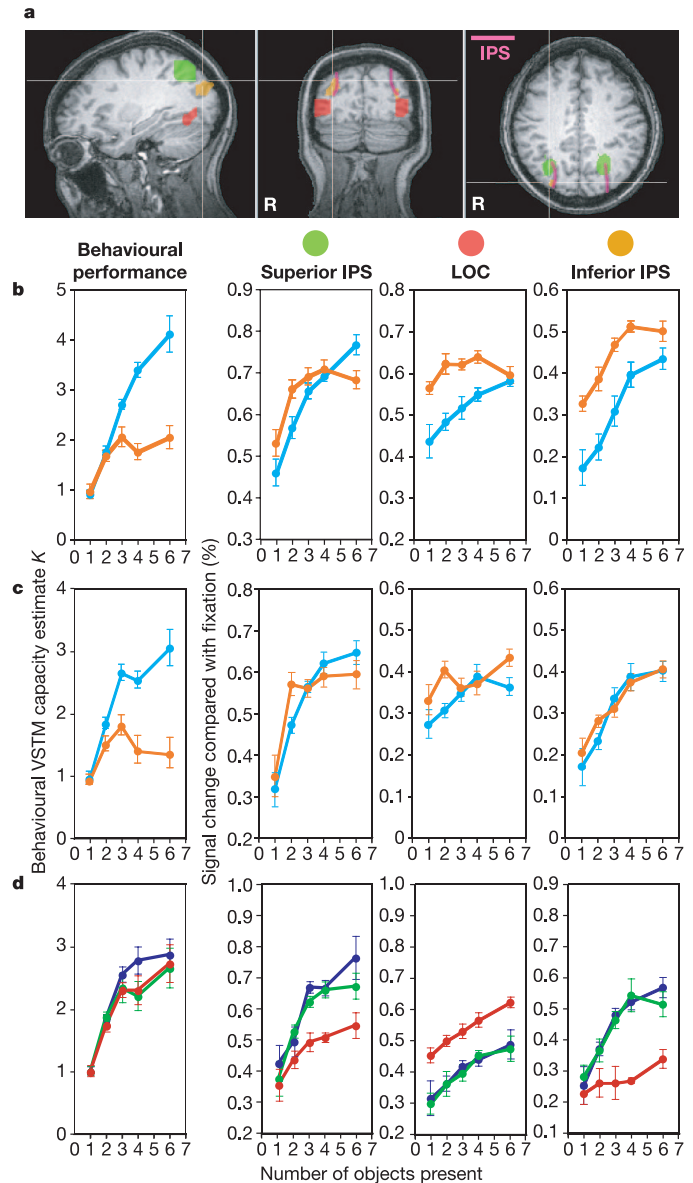


Figure 2 | Results from experiments 1, 2 and 4. **a**, Superior IPS (green), LOC (red) and inferior IPS (orange) ROIs from an example observer in experiment 1. The IPS itself is highlighted in magenta. The averaged Talairach coordinates (x, y, z) for the ROIs are as follows: right/left, +23/-21 -52/-66 +45/+42 for the superior IPS, +26/-21 -80/-85 +30/+26 for the inferior IPS with off-centre presentation, +26/-25 -65/-70 +34/+29 for the inferior IPS with centred presentation (experiment 4), +42/-44 -67/-69 +3/+8 for the LOC with off-centre presentation, and +43/-41 -67/-71 -5/-1 for the LOC with centred presentation (Experiment 4). **b-d**, Averaged behavioural results and peak responses from the ROIs examined in experiments 1 (**b**), 2 (**c**) and 4 (**d**) (means \pm within-subject s.e.m.). In experiment 1, VSTM capacity was about four objects for simple shape features (cyan) and only two objects for complex shape features (orange). In experiment 2, VSTM capacity was about three for simple objects (cyan) and 1.5 for complex objects (orange). In both experiments, whereas a fixed object capacity was observed in the inferior IPS, a variable object capacity tracking the number of objects held in VSTM was observed in the superior IPS and the LOC as object complexity changed. In experiment 4, VSTM capacity was about three objects in all three display conditions: blue, simultaneous off-centre; green, sequential off-centre; red, sequential centred. Whereas responses in the superior IPS and the LOC tracked VSTM capacity in all three display conditions, those in the inferior IPS did so only when objects appeared in different spatial locations. When objects appeared sequentially at the same central location, however, IPS responses did not vary with set size.

number of objects regardless of object complexity, those in the superior IPS and LOC followed the actual number of objects held in VSTM as object feature complexity changed. Our left superior and inferior IPS ROIs coincide with the region whose activity has been shown to correlate with individual differences in VSTM capacity¹⁸.

Because the encoding of simple shape features required a binary discrimination of hole presence or absence at the centre of each shape, observers could have successfully performed the VSTM task by encoding and remembering only the hole-present (or the hole-absent) shapes without retaining features from the other shapes. This potential grouping strategy could therefore have inflated the estimated VSTM capacity and obscured the corresponding brain activations. Although this strategy would only have reduced the differences between set sizes, inconsistent with the pattern of fMRI results reported above, to test this grouping hypothesis directly we conducted a second fMRI experiment that replicated the design of the first experiment with uniquely different simple and complex objects constructed from the same set of shape features (Fig. 1b).

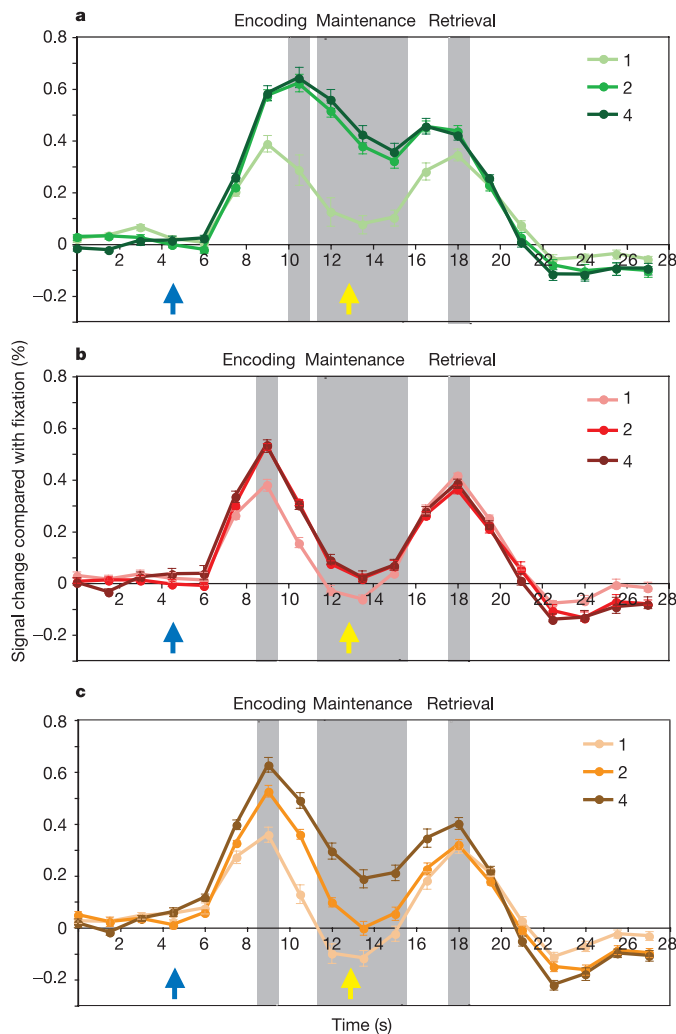


Figure 3 | Results from experiment 3. Activations (means \pm within-subject s.e.m.) from the superior IPS (a), the LOC (b) and the inferior IPS (c). The blue and the yellow arrows indicate onsets of the sample and the test object displays, respectively. Behavioural capacity estimate K values were 0.92 (with a within-subject s.e.m. of 0.14), 1.30 (0.12) and 1.14 (0.22) for set sizes 1, 2 and 4, respectively. Whereas activations in the superior IPS and the LOC followed K values and increased from set sizes 1 to 2 and reached a plateau at set size 2 during both VSTM encoding and maintenance, those in the inferior IPS increased as set size increased from 1 to 4 during both VSTM encoding and maintenance.

As expected, behavioural VSTM capacity was higher for simple than for complex objects (Fig. 2c), with a significant interaction between set size and object complexity ($n = 10$, $P < 0.01$). For simple objects, K values increased with set size ($P < 0.05$), differed among set sizes 2 to 6 ($P < 0.01$), and reached a plateau at set size 3 with K values about 3 objects ($P > 0.29$ for the differences among set sizes 3, 4 and 6). For complex objects, K values did not increase with set size ($P > 0.62$) and reached a plateau at set size 2 with K values about 1.5 objects ($P > 0.49$ for the differences among set sizes 2 to 6). The maximum VSTM capacity was therefore about three objects for simple objects and about 1.5 objects for complex objects.

Replicating the first experiment, superior IPS and LOC activations differed among set sizes 2 to 6 for simple objects ($P < 0.05$), but not for complex objects ($P > 0.73$ for superior IPS, and $P > 0.067$ for LOC); the marginally significant difference was driven by the inexplicably high activation for set size 6 compared with otherwise similar activations among set sizes 2 to 4, $P > 0.28$). The interactions between set size and object complexity in these areas were significant ($P < 0.05$). Moreover, for simple objects, the increase in activations with set size in these brain regions correlated with the increases in K values ($P < 0.05$). In contrast, regardless of object complexity, activations in the inferior IPS increased from set sizes 1 to 4 ($P < 0.001$) and reached a plateau at set size 4 ($P > 0.16$ for the differences between set sizes 4 and 6), with no interaction between set size and object complexity ($P > 0.56$) (Fig. 2c). Thus, even after grouping cues were removed, the results of the second experiment mirrored those of the first experiment, showing a fixed object capacity in the inferior IPS and variable, complexity-dependent object capacity in the superior IPS and the LOC.

It is possible that the lower VSTM capacity for the complex objects was due to perceptual processing limitations rather than memory limitations. To test this hypothesis we conducted a behavioural experiment that provided more time for perceptual processing. Using the complex objects from the second experiment, we found that VSTM capacity did not improve whether the encoding time for four objects was 200 or 500 ms ($n = 6$, $P > 0.82$; $K = 1.90$ and $K = 1.85$, respectively) or whether four objects were presented simultaneously for 200 ms or sequentially two at a time for 200 ms with a 500-ms blank interval in between. In fact, simultaneous presentation was better than sequential presentation ($n = 6$, $P < 0.05$; $K = 1.90$ and $K = 1.60$, respectively). These results indicate that the reduced VSTM capacity for complex objects was due to limitations in VSTM rather than perception.

To dissociate whether brain activations observed so far reflect VSTM encoding, maintenance, or retrieval and comparison, in a third fMRI experiment we presented either one, two or four complex objects from the second experiment (Fig. 1b) and asked observers to maintain the objects over an extended 8,300-ms delay^{9,10,17}. Activations in the LOC and the superior IPS differed between set sizes 1 and 2 and between set sizes 1 and 4 ($n = 8$, $P < 0.05$) but not between set sizes 2 and 4 ($P > 0.18$) during both encoding and maintenance (Fig. 3a, b). These differences were largely maintained in the superior IPS during retrieval ($P < 0.05$ between set sizes 1 and 2, $P < 0.079$ between 1 and 4, and $P > 0.64$ between 2 and 4), but not in the LOC ($P > 0.18$ for the overall differences among set sizes). Activations in the inferior IPS differed between set sizes 1 and 2, between 1 and 4, and between 2 and 4 ($P < 0.05$) during both encoding and maintenance, and these differences were partly retained during retrieval (Fig. 3c, $P < 0.092$ between set sizes 1 and 4, $P < 0.05$ between 2 and 4, and $P > 0.87$ between 1 and 2). These results indicate that brain activations observed in the first two fMRI experiments mainly reflected activations during VSTM encoding and maintenance, consistent with a previous finding on the superior IPS¹⁷.

Finally, to examine whether LOC and IPS activations during VSTM tasks track object identity or simply the locations occupied by objects in the display and in memory, we conducted a fourth fMRI experiment that varied whether objects shared the same or different

spatial locations. One to four, or six, different black objects were presented in one of three ways: simultaneously, sequentially at different off-centre locations, or sequentially at the same centre location. The task was to judge whether the probe object in the test display was present in the sample display. Because the probe object always appeared at the centre of the test display, only VSTM for object identities but not their locations was needed for this task (Fig. 1c). For all three presentation conditions, K varied significantly with set size ($n = 7$, $P < 0.01$), increased as set size increased from 1 to 3 ($P < 0.05$) and reached a plateau at set size 3 ($P > 0.46$ for the differences among set sizes 3, 4 and 6), with no differences between the different conditions ($P > 0.92$). VSTM capacity was therefore about three objects for all three presentation conditions.

To account for differences in eccentricity for centred and off-centre presentations, we separately localized the part of the LOC and the inferior IPS that respond to centred and off-centre presentations (Supplementary Fig. 1d). Regardless of whether objects occupied the same or different spatial locations, fMRI responses in the superior IPS and LOC varied with set sizes ($P < 0.06$ for LOC in the simultaneous condition and $P < 0.05$ for all others) and correlated with the corresponding K values ($P < 0.05$) in all three presentation conditions. However, inferior IPS exhibited a similar response pattern only when objects appeared at different spatial locations regardless of simultaneous or sequential presentation ($P < 0.05$). When objects appeared sequentially at the same central location, however, inferior IPS responses did not vary with set size ($P > 0.45$). Overall LOC responses were greater for centred (fovea) than off-centre presentations ($P < 0.073$), whereas the opposite was true for superior IPS and inferior IPS responses ($P < 0.05$ and $P < 0.093$, respectively). These results indicate that the LOC represents the visual objects held in VSTM and some object location information, as indicated by the object location eccentricity effect in Fig. 2d. In contrast, inferior IPS representations during VSTM are more spatial in nature²⁷, automatically indexing a fixed number of objects by means of their locations, even when the encoding of spatial location is not required. Meanwhile, responses in the superior IPS in this task resembled both those of the LOC (with responses correlating to the number of objects held in VSTM) and those of the inferior IPS (with a lower overall response for centred than off-centre presentations), indicating that both object identity and some location information might be represented in this brain area during VSTM²⁷.

Together, our data from four fMRI experiments indicate dissociable neural mechanisms in the superior and the inferior IPS and the LOC that work in parallel to support VSTM during visual object encoding and maintenance. Regardless of object complexity, inferior IPS representations are limited by a fixed number of objects at different spatial locations, indicating a possible spatial indexing mechanism that maintains spatial attention over a fixed number of objects at different spatial locations during both encoding and maintenance of VSTM. Consistent with its role in attention^{28,29}, this mechanism may select and determine the maximum number of objects that can be maintained in VSTM and possibly other capacity-limited visual tasks such as subitizing and multiple object tracking^{25,30}. In contrast, the representations in the LOC and the superior IPS are not limited by a fixed number of objects, but rather by object complexity and the total amount of visual information encoded, indicating that these brain areas participate in the detailed representation of visual objects in VSTM during both encoding and maintenance. Thus, VSTM capacity is determined both by a fixed number of objects and by object complexity. These findings not only clarify how parietal and occipital cortices mediate VSTM but also help to resolve a central debate over what determines VSTM capacity²⁻⁵.

METHODS

In experiment 1, nine observers viewed elongated and tilted black objects in a sample display and, after a brief delay, detected a shape feature change in the test display at the probed location in 50% of the trials (Fig. 1a). In different task

blocks, observers encoded either simple shape features (the presence or absence of a hole in the shape) or complex shape features (shape outlines). In experiment 2, each complex object was constructed by attaching a cap-like simple object to a stem-like simple object⁴. Twelve observers encoded either simple or complex objects (Fig. 1b). In experiment 3, nine observers retained the complex objects used in experiment 2 for an extended memory delay period. In experiment 4, eight observers encoded objects that were presented simultaneously, sequentially at different off-centre locations, or sequentially at the same centre location, with the probe object always appearing at the same centre location (Fig. 1c).

MRI data were acquired from a Siemens Trio 3T scanner and analysed with BrainVoyager QX 1.3 (<http://www.brainvoyager.com>). After defining the ROIs separately in each observer, we overlaid the ROIs onto the MRI data from our VSTM experiment and extracted time courses from each observer. The resulting peak responses were then averaged across observers. Behavioural data collected during the scans were transformed to VSTM capacity estimates by using Cowan's K formula²⁵: $K = (\text{hit rate} + \text{correct rejection rate} - 1) \times (\text{display set size})$.

More details of the experiments are available in Supplementary Methods.

Received 29 July; accepted 26 September 2005.

Published online 28 December 2005.

- Phillips, W. A. On the distinction between sensory storage and short-term visual memory. *Percept. Psychophys.* **16**, 283–290 (1974).
- Pashler, H. Familiarity and visual change detection. *Percept. Psychophys.* **44**, 369–378 (1988).
- Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
- Xu, Y. Encoding colour and shape from different parts of an object in visual short-term memory. *Percept. Psychophys.* **64**, 1260–1280 (2002).
- Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* **15**, 106–111 (2004).
- Fuster, J. M. & Jervey, J. P. Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J. Neurosci.* **2**, 361–375 (1982).
- Goldman-Rakic, P. S. In *Handbook of Physiology: The Nervous System, Higher Functions of the Brain* (eds Mountcastle, V. B., Plum, F. & Geiger, S. R.) 373–417 (American Physiological Society, Bethesda, Maryland, 1987).
- Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
- Cohen, J. D. *et al.* Temporal dynamics of brain activation during a working memory task. *Nature* **386**, 604–608 (1997).
- Courtney, S. M., Ungerleider, L. G., Keil, K. & Haxby, J. V. Transient and sustained activity in a distributed neural system for human working memory. *Nature* **386**, 608–611 (1997).
- Desimone, R. Neural mechanisms for visual memory and their role in attention. *Proc. Natl Acad. Sci. USA* **93**, 13494–13499 (1996).
- Smith, E. E. & Jonides, J. Neuroimaging analyses of human working memory. *Proc. Natl Acad. Sci. USA* **95**, 12061–12068 (1998).
- Pessoa, L., Gutierrez, E., Bandettini, P. A. & Ungerleider, L. G. Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron* **35**, 975–987 (2002).
- Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* **9**, 415–423 (2003).
- Jonides, J., Lacey, S. C. & Nee, D. E. Processes of working memory in mind and brain. *Curr. Dir. Psychol. Sci.* **14**, 2–5 (2005).
- Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nature Rev. Neurosci.* **6**, 97–107 (2005).
- Todd, J. J. & Marois, R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* **428**, 751–754 (2004).
- Todd, J. J. & Marois, R. Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cogn. Affect. Behav. Neurosci.* **6**, 144–155 (2005).
- Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* **428**, 748–751 (2004).
- Malach, R. *et al.* Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci. USA* **92**, 8135–8139 (1995).
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y. & Malach, R. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21**, 191–202 (1998).
- Kourtzi, Z. & Kanwisher, N. Cortical regions involved in perceiving object shape. *J. Neurosci.* **20**, 3310–3318 (2000).
- Grill-Spector, K., Kushnir, T., Hendler, T. & Malach, R. The dynamics of object-selective activation correlate with recognition performance in human. *Nature Neurosci.* **3**, 837–843 (2000).
- Awh, E. & Jonides, J. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.* **10**, 433–437 (1999).
- Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**, 87–114 (2001).

26. Talairach, J. & Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain* (transl. Rayport, M) (Thieme Medical, New York, 1988).
27. Sereno, M. I., Pitzalis, S. & Martinez, A. Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in human. *Science* **294**, 1350–1354 (2001).
28. Kanwisher, N. & Wojciulik, E. Visual attention: insights from brain imaging. *Nature Rev. Neurosci.* **1**, 91–100 (2000).
29. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nature Rev. Neurosci.* **3**, 215–229 (2002).
30. Culham, J. C. *et al.* Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron* **32**, 737–745 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Widders for assistance in MRI scanning, and S. J. Luck and R. Marois for comments on earlier versions of this manuscript. This research was supported by an NIH grant to M.M.C. and in part by an NSF grant to Y.X.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.X. (yaoda.xu@yale.edu).