# LETTERS

Decoding reveals the contents of visual working memory in early visual areas

Stephenie A. Harrison[1] & Frank Tong[1]

Visual working memory provides an essential link between perception and higher cognitive functions, allowing for the active maintenance of information about stimuli no longer in view[1,2]. Research suggests that sustained activity in higher-order prefrontal, parietal, inferotemporal and lateral occipital areas supports visual maintenance[3–11], and may account for the limited capacity of working memory to hold up to 3–4 items[9–11]. Because higher-order areas lack the visual selectivity of early sensory areas, it has remained unclear how observers can remember specific visual features, such as the precise orientation of a grating, with minimal decay in performance over delays of many seconds[12]. One proposal is that sensory areas serve to maintain fine-tuned feature information[13], but early visual areas show little to no sustained activity over prolonged delays[14–16]. Here we show that orientations held in working memory can be decoded from activity patterns in the human visual cortex, even when overall levels of activity are low. Using functional magnetic resonance imaging and pattern classification methods, we found that activity patterns in visual areas V1–V4 could predict which of two oriented gratings was held in memory with mean accuracy levels upwards of 80%, even in participants whose activity fell to baseline levels after a prolonged delay. These orientation-selective activity patterns were sustained throughout the delay period, evident in individual visual areas, and similar to the responses evoked by unattended, task-irrelevant gratings. Our results demonstrate that early visual areas can retain specific information about visual features held in working memory, over periods of many seconds when no physical stimulus is present.

To investigate the role of early visual areas in working memory, we used functional magnetic resonance imaging (fMRI) to monitor cortical activity while participants performed a delayed orientation discrimination task. During each trial, observers maintained fixation while two sample orientation gratings (~25° and ~115°) were briefly presented in randomized order, followed by a numerical cue indicating whether to remember the first or second grating (Fig. 1a). After an 11-s retention interval, a test grating was presented, and participants indicated which way it was rotated relative to the cued grating (±3° or ±6°). This experimental design allowed us to isolate memory-specific activity. By presenting the same two gratings on every trial, we ensured that stimulus-driven activity could not predict the orientation held in working memory. It was also critical that the memory cue appeared after the presentation of the gratings and not beforehand. Otherwise, subjects could attend more to the appearance of the cued grating, which would enhance orientation-selective responses to that stimulus[17].

Behavioural data confirmed that observers could discriminate small differences in orientation between the cued grating and the test grating. Observers showed equally good performance when the first or second grating had to be remembered (75% and 73% correct, respectively, $T(5) = 1.24$, $P = 0.27$).
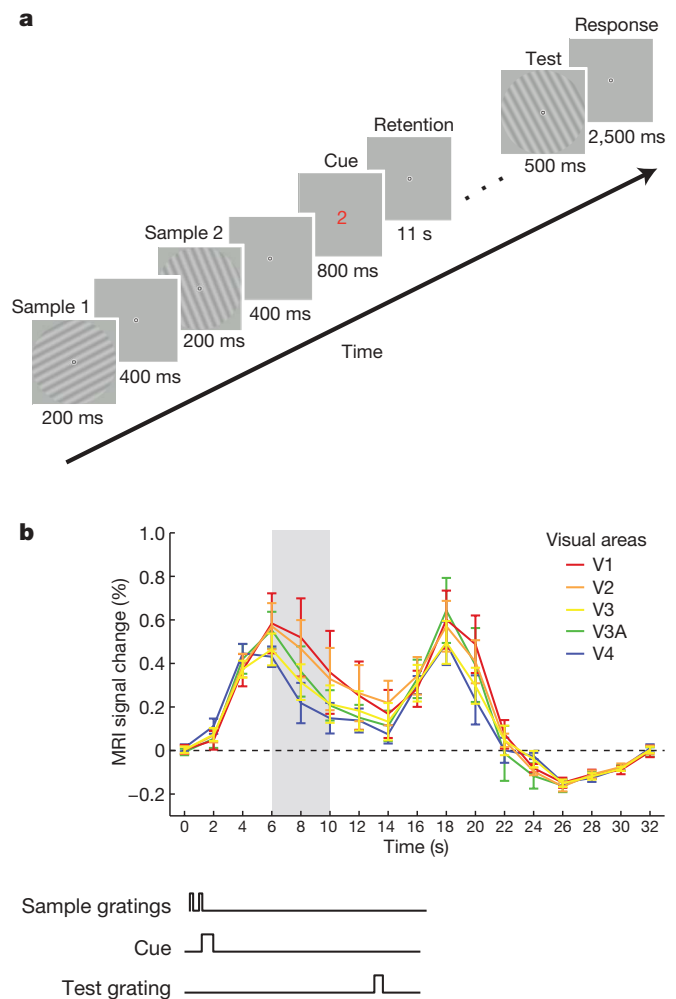
Figure 1 | Design of working memory experiment and resulting time course of fMRI activity. a, Timing of events for an example working memory trial. Two near-orthogonal gratings (25° ± 3°, 115° ± 3°) were briefly presented in randomized order, followed by a numerical cue (green '1' or red '2') indicating which grating to remember. After an 11-s retention period, a test grating was presented, and subjects reported whether it was rotated clockwise or anticlockwise relative to the cued grating. b, The time course of mean BOLD activity ($n = 6$) in corresponding regions of areas V1–V4 during the working memory task (0–16 s) and subsequent fixation period (16–32 s). Error bars indicate ± s.e.m. Time points 6–10 s (shaded grey area) were averaged for subsequent decoding analysis of delay-period activity. The start of this time window was chosen to allow for peak BOLD activity to fully emerge; we selected a conservative end point of 10 s to exclude any potential activity elicited by the test grating.

[1]Psychology Department and Vanderbilt Vision Research Center, Vanderbilt University, Nashville, Tennessee 37240, USA.

We used fMRI decoding methods to determine whether activity in early visual areas might reflect the contents of working memory (see Methods and Supplementary Methods). Although orientation selectivity primarily resides at fine spatial scales in the visual cortex, we have previously shown that pattern classification methods can successfully recover orientation information from cortical activity sampled at coarser resolutions using fMRI[17]. Here we investigated whether activity patterns during the delay period might predict which of the two orientations was held in working memory. For each trial, we calculated the average response of individual voxels over time points 6–10 s (Fig. 1b, grey region), selecting voxels from regions corresponding to 1–4° eccentricity in areas V1 to V4. The activity patterns observed on each trial served as input to a linear classifier with the cued orientation indicating the corresponding label. Classification accuracy was determined using cross-validation methods.

Ensemble activity pooled from areas V1–V4 was highly predictive of the orientation held in working memory, with prediction accuracy reaching 83% (Fig. 2, green curve). Decoding accuracy greatly exceeded chance-level performance of 50% ($T(5) = 18.2$, $P < 10^{-5}$), and proved highly reliable in each of the six participants (performance exceeding 58.75%, $P < 0.05$, binomial test). Notably, decoding was just as effective when the first grating was cued instead of the second (82.1% versus 83.6%, respectively, $T(5) = 1.0$, $P = 0.36$), indicating that this orientation information in the visual cortex was robust to potential interference from a subsequent item. Such robustness to interference has previously been found only in the prefrontal cortex[5]. Individual visual areas showed similar levels of orientation decoding performance ($F(3,15) = 1.71$, $P = 0.21$) ranging from 71–74% accuracy, with every participant showing above-chance decoding in each area. This indicates that maintaining an orientation in working memory is associated with widespread changes in orientation-selective activity throughout the early visual system, including V1, the first stage of orientation processing.

How do these orientation-selective responses for remembered gratings compare with stimulus-driven activity elicited by direct viewing of actual gratings? In a second experiment, participants had to identify letters presented rapidly at fixation while ignoring

low-contrast oriented gratings (25° or 115°) flashing in the surround. Although the gratings were quite faint and task-irrelevant, they nonetheless evoked strong orientation-selective responses in early visual areas (Fig. 2, red curve). Activity in individual areas, V1, V2 and V3, was highly predictive of the orientation of the unattended gratings. Performance was considerably worse for V3A–V4 ($F(3,15) = 20.4$, $P < 10^{-4}$), presumably because activity in higher extrastriate areas is more dependent on visual attention[18]. Next, we evaluated the similarity of orientation-selective activity patterns in the two experiments by training the classifier on one data set and testing it on the other. Generalization performance for activity pooled across V1–V4 was below the performance found in the working memory experiment (Fig. 2, black curve), but was still significantly above chance ($T(5) = 6.0$, $P < 0.005$). Generalization was also better in V1 and V2 than in higher areas ($F(3,15) = 4.5$, $P < 0.05$), perhaps because these early areas exhibit stronger orientation-selective responses under stimulus-driven conditions[17]. Successful generalization across the two experiments is notable given how they differed in both stimulus and task. It seems that retaining an orientation in working memory recruits many of the same orientation-selective subpopulations as those that are activated under stimulus-driven conditions.

Further analyses confirmed that successful orientation decoding could not be explained by global differences in response amplitudes to the two orientations, as decoding applied to the averaged response of each visual area led to chance-level performance (46–57% accuracy, Supplementary Fig. 1a). We also tested for potential effects of global radial bias[19], and found that decoding was significantly impaired by spatially averaging the response of neighbouring voxels corresponding to different radial segments of the visual field (Supplementary Fig. 1b). In contrast, local variations in orientation preference within each radial segment led to high decoding accuracy (Supplementary Fig. 1c), consistent with the notion that much of the orientation information extracted by the classifier resulted from local anisotropies in orientation preference[17] (Supplementary Fig. 2).

Next, we investigated whether orientation-selective activity is maintained throughout the working memory delay period, by performing our decoding analysis on individual fMRI time points. Although individual functional images show poorer signal to noise, we could still detect changes in orientation-selective activity over time in both experiments. Orientation decoding of stimulus-driven activity in areas V1–V4 rose above chance level within 4 s of stimulus onset ($T(5) = 4.13$, $P < 0.01$) and reached asymptotic levels by ~6 s (consistent with the slow time course of the blood-oxygen-level-dependent (BOLD) response); performance remained high as gratings continued to be shown throughout the 16-s stimulus block (Fig. 3a). In comparison, orientation-selective activity in the working memory experiment was delayed by ~2 s, rising significantly above baseline by 6 s ($T(5) = 4.36$, $P < 0.01$) and reaching a plateau by 8 s. This delayed onset is consistent with the fact that observers did not see the task-relevant cue until 1.2 s after the first grating appeared, and required more time to interpret the cue. More notable is the fact that orientation-selective activity persisted throughout the delay period, when no physical stimulus was present, up until presentation of the test grating at time 13 s. Decoding of individual areas led to lower levels of performance; however, a similar pattern of results was found, as is shown for V1 (Fig. 3b).

Interestingly, this maintenance of orientation-selective information throughout the delay period did not seem to depend on a sustained boost in overall BOLD activity. The time course of mean BOLD activity for each visual area revealed a transient response to the first two gratings and a subsequent response to the test grating, with some suggestion of sustained activity in the intervening period (Fig. 1b). However, the level of sustained activity varied widely across subjects. For example, in V1 half of our subjects showed greater than baseline activity late in the delay period, whereas half did not (Supplementary Fig. 3a, b). Nevertheless, orientation-decoding performance was equally good for the two groups (74% versus 75%) and was sustained
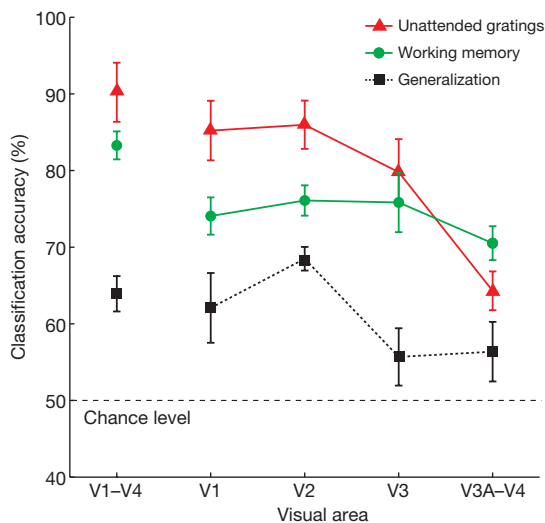
Figure 2 | Orientation decoding results for areas V1–V4. The accuracy of orientation decoding for remembered gratings in the working memory experiment (green circles), unattended presentations of low-contrast gratings (red triangles), and generalization performance across the two experiments (black squares). Error bars indicate ± s.e.m. Decoding was applied to the 120 most visually responsive voxels in each of V1, V2, V3 and V3A–V4 (480 voxels for V1–V4 pooled), as determined by their responses to a localizer stimulus (1–4° eccentricity). Individual areas V3A and V4 showed similar decoding performance but had fewer available voxels, so these regions were combined.
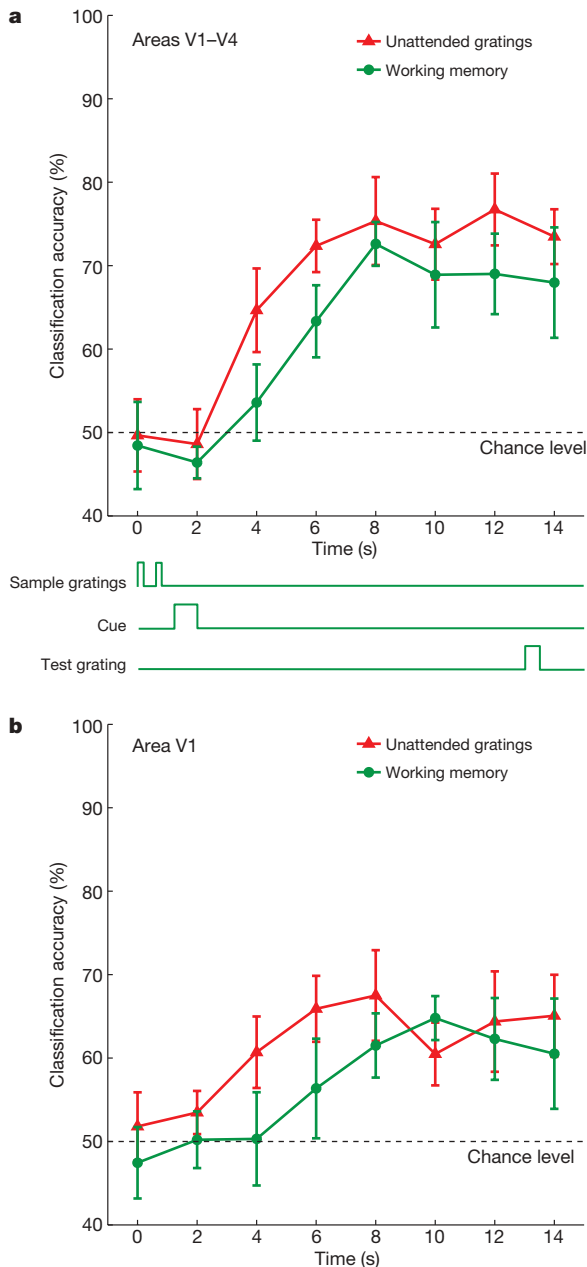
Figure 3 | Time-resolved decoding of individual fMRI time points.
Orientation decoding of unattended stimulus gratings (red triangles), and remembered gratings during working memory (green circles), for activity obtained from areas V1–V4 (**a**) and from V1 only (**b**). Note that orientation information persists throughout the delay period during the working memory task, up until presentation of the test grating at time of 13 s. Error bars indicate ± s.e.m.

memory from activity in early visual areas (Supplementary Fig. 4). The use of randomly selected orientations ensured that long-term memory could not contribute to delayed discrimination; instead, accurate performance could only be achieved by maintaining the task-relevant grating seen on each trial (behavioural accuracy 76.2%). In another experiment, observers were shown two sample orientations followed by a numerical cue, the colour of which indicated whether to make a speeded judgment about the task-relevant orientation or to retain that orientation for subsequent discrimination. Whereas the immediate report task led to unreliable orientation decoding, active maintenance of the task-relevant grating over an extended 15-s delay led to sustained orientation-selective activity in areas V1 to V4 (Supplementary Fig. 5). Furthermore, we tested for effects of visual expectancy by omitting the sample gratings and providing only an initial cue to indicate the approximate orientation ($\sim$25° or $\sim$115°) observers should expect at test. Expectation of a specific future orientation to be discriminated led to good behavioural performance (77.5% correct), but weak orientation-selective responses, as indicated by near chance-level decoding (Supplementary Fig. 6).

We also considered whether eye movements could account for successful decoding of remembered orientations; there are several reasons why this seems unlikely. First, sample gratings were presented for only 200 ms, too briefly for participants to prepare an eye movement within that time; also the working memory cue occurred afterwards, when no other stimulus was present. Second, an eye-tracking control experiment confirmed that all six participants maintained stable fixation when performing the working memory task (see Supplementary Methods). Unlike activity in the visual cortex, eye position signals failed to predict the orientation held in working memory (orientation decoding accuracy, 50.2%, $P = 0.94$). Third, it would be difficult to explain how strategic eye movements during working memory might elicit differential activity patterns that resemble those evoked by unattended gratings when participants had to attend to letters at fixation. Both the stimulus conditions and the strategic demands of the two experiments were profoundly different.

Our results provide new evidence to show that early visual areas can retain specific information about visual features held in working memory. When participants had to remember a precise orientation, this information was maintained in sensory areas, including the primary visual cortex where orientation tuning is strongest. Although V1 is essential for low-level feature processing, there is increasing evidence to suggest a role for V1 in conscious perception[20], attentional selection[18,20] and more complex cognitive functions[21,22]. We find that early visual areas are not only important for processing information about the immediate sensory environment, but can also maintain information in the absence of direct input to support higher-order cognitive functions.

Thus far, there has been little evidence to link V1 activity to visual working memory, perhaps because these tasks do not normally lead to increased activity in the visual cortex[14–16]. One study did find relatively greater V1 activity when monkeys had to report a remembered spatial location by means of an eye movement[23], but this increase in baseline activity could reflect the effects of spatial attention[18,24] or eye movement preparation[25]. Here we found that the overall activity in the visual cortex fell to near-baseline levels after prolonged delays, yet decoding of these low amplitude signals led to reliable prediction of the orientation held in memory.

Our findings suggest a potentially important source of memory-related information that may have been overlooked in previous studies, and indicate promising avenues for future research. Assuming that items in visual working memory are encoded by low levels of population activity, the application of population-decoding methods could help to uncover the underlying neural representations. Previous attempts to decode remembered information from delay-period activity in single neurons have typically led to low or chance levels of performance[5,16,26]. Perhaps if signals from many neurons or neuronal sites were recorded simultaneously to exclude the effects of correlated noise[27], far greater

throughout the delay period (Supplementary Fig. 3c, d). Further analyses supported the notion that the overall BOLD amplitude from a region was unrelated to the amount of memory-related information available in the detailed activity pattern. We found no significant relationship between BOLD amplitudes and decoding accuracy across subjects, or across trials for individual subjects. Thus, it seems that low amplitude signals can nonetheless contain robust memory-related information throughout the entire delay period.

Additional control experiments indicated that this sustained orientation-selective activity reflected active maintenance of the cued orientation throughout the delay period rather than other cognitive processes. When observers were presented with a randomly selected pair of near-orthogonal orientations on every trial, it was still possible to decode which of the two orientations was held in working

information could be uncovered about items retained in memory, as was demonstrated here. The role of synaptic activity in the visual cortex might also be useful to explore, given that the BOLD response is more strongly associated with synaptic than spiking activity[28]. One recent study has reported suggestive evidence of enhanced local field potentials (4–10 Hz) in area V4 of the monkey during a visual working memory task[29]. Curiously, spiking activity did not increase overall but it was more likely to be observed at a specific phase of these slow oscillations, suggesting that the relationship between working memory and spiking activity might go beyond simple changes in firing rate.

It will be interesting for future studies to investigate whether working memory information found in the visual cortex is actively maintained by long-range recurrent interactions between higher-order areas and early visual areas, local recurrent activity within early visual areas, or a combination of both mechanisms. Presumably, prefrontal or parietal areas contributed to the top-down selection process, given that participants had to interpret an abstract cue indicating which of two orientations to hold in memory. However, it has been debated whether feedback signals from higher-order areas would necessarily reflect the contents of working memory[8]. Most network models of working memory have emphasized the importance of local recurrent activity[30]. In these models, a specific pattern of activity can be sustained after stimulus removal if units tuned to similar features share strong excitatory connections, balanced by broad inhibition from units tuned to other features. It is possible that the functional organization of orientation-selective neurons in the visual cortex could provide an infrastructure for such interactions. The present results demonstrate that early visual areas can indeed sustain information for periods of many seconds, indicating that their function is not restricted to sensory processing but extends to the maintenance of visual features and patterns in memory.

METHODS SUMMARY

Six observers, aged 24–36, with normal or corrected-to-normal vision, participated in this study, after providing written informed consent. The study was approved by the Vanderbilt University Institutional Review Board.

The main study consisted of three fMRI experiments. The working memory experiment involved delayed discrimination of one of two randomly cued orientations (Fig. 1a). Sine-wave gratings were centrally presented at ~25° or ~115° orientation (radius 5°, contrast 20%, spatial frequency 1 cycle per degree, randomized phase). The unattended gratings experiment required participants to report whenever a 'J' or 'K' appeared within a sequence of centrally presented letters (4 letters per s, performance accuracy 87.3%) while task-irrelevant gratings flashed on or off every 250 ms during each 16-s stimulus block. Gratings were identical to those used in the working memory experiment, but presented at lower contrast (4%) to elicit weaker visual responses, as might be expected during working memory. The visual-field localizer experiment consisted of blocked presentations of flickering random dots (dot size, 0.2°; display rate, 10 images per s), presented within an annulus of 1–4° eccentricity. This smaller window was used to minimize selection of retinotopic regions corresponding to the edges of the grating stimuli. Observers were instructed to maintain fixation on a central bull's eye throughout every experiment. Participants completed 8–10 working memory runs (32–40 trials per orientation), 4–5 unattended grating runs (28–35 blocks per orientation), and 2 visual-field localizer runs.

Scanning was performed using a 3.0-Tesla Philips Intera Achieva MRI scanner at the Vanderbilt University Institute of Imaging Science. We used gradient-echo echoplanar T2*-weighted imaging (time to echo (TE), 35 ms; repetition time (TR) 2,000 ms; flip angle, 80°; 28 slices, voxel size, $3 \times 3 \times 3$ mm) to obtain functional images of the entire occipital lobe, as well as posterior parietal and temporal regions. Participants used a bite bar system to minimize head motion.

1. Baddeley, A. Working memory: looking back and looking forward. *Nature Rev. Neurosci.* 4, 829–839 (2003).
2. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281 (1997).
3. Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* 173, 652–654 (1971).
4. Miyashita, Y. & Chang, H. S. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331, 68–70 (1988).
5. 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).
6. 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).
7. Pessoa, L., Gutierrez, E., Bandettini, P. & Ungerleider, L. Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron* 35, 975–987 (2002).
8. Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423 (2003).
9. Todd, J. J. & Marois, R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754 (2004).
10. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751 (2004).
11. Xu, Y. & Chun, M. M. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95 (2006).
12. Magnussen, S. & Greenlee, M. W. The psychophysics of perceptual memory. *Psychol. Res.* 62, 81–92 (1999).
13. Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nature Rev. Neurosci.* 6, 97–107 (2005).
14. Offen, S., Schluppeck, D. & Heeger, D. J. The role of early visual cortex in visual short-term memory and visual attention. *Vision Res.* doi:10.1016/j.visres.2007.12.022 (in the press).
15. Bisley, J. W., Zaksas, D., Droll, J. A. & Pasternak, T. Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.* 91, 286–300 (2004).
16. Zaksas, D. & Pasternak, T. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742 (2006).
17. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nature Neurosci.* 8, 679–685 (2005).
18. Kastner, S. & Ungerleider, L. G. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341 (2000).
19. Sasaki, Y. *et al.* The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* 51, 661–670 (2006).
20. Tong, F. Primary visual cortex and visual awareness. *Nature Rev. Neurosci.* 4, 219–229 (2003).
21. Kosslyn, S. M., Ganis, G. & Thompson, W. L. Neural foundations of imagery. *Nature Rev. Neurosci.* 2, 635–642 (2001).
22. Roelfsema, P. R. Elemental operations in vision. *Trends Cogn. Sci.* 9, 226–233 (2005).
23. Super, H., Spekreijse, H. & Lamme, V. A. A neural correlate of working memory in the monkey primary visual cortex. *Science* 293, 120–124 (2001).
24. Ress, D., Backus, B. T. & Heeger, D. J. Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neurosci.* 3, 940–945 (2000).
25. Geng, J. J., Ruff, C. C. & Driver, J. Saccades to a remembered location elicit spatially specific activation in the human retinotopic visual cortex. *J. Cogn. Neurosci.* 21, 230–245 (2009).
26. Miller, E. K., Li, L. & Desimone, R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* 13, 1460–1478 (1993).
27. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nature Rev. Neurosci.* 7, 358–366 (2006).
28. Logothetis, N. K. *et al.* Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157 (2001).
29. Lee, H., Simpson, G. V., Logothetis, N. K. & Rainer, G. Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45, 147–156 (2005).
30. Wang, X. J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463 (2001).