Behavioral/Systems/Cognitive

The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging

Adam C. Riggall[1] **and Bradley R. Postle**[1,2]
Departments of [1]Psychology and [2]Psychiatry, University of Wisconsin-Madison, Madison, Wisconsin 53706

Does the sustained, elevated neural activity observed during working memory tasks reflect the short-term retention of information? Functional magnetic resonance imaging (fMRI) data of delayed recognition of visual motion in human participants were analyzed with two methods: a general linear model (GLM) and multivoxel pattern analysis. Although the GLM identified sustained, elevated delay-period activity in superior and lateral frontal cortex and in intraparietal sulcus, pattern classifiers were unable to recover trial-specific stimulus information from these delay-active regions. The converse—no sustained, elevated delay-period activity but successful classification of trial-specific stimulus information—was true of posterior visual regions, including area MT+ (which contains both middle temporal area and medial superior temporal area) and calcarine and pericalcarine cortex. In contrast to stimulus information, pattern classifiers were able to extract trial-specific task instruction-related information from frontal and parietal areas showing elevated delay-period activity. Thus, the elevated delay-period activity that is measured with fMRI may reflect processes other than the storage, per se, of trial-specific stimulus information. It may be that the short-term storage of stimulus information is represented in patterns of (statistically) "subthreshold" activity distributed across regions of low-level sensory cortex that univariate methods cannot detect.

Introduction

For decades, a prevailing view has been that working memory (WM) storage is accomplished via sustained, elevated neural activity. Such activity, first identified with extracellular recordings in the nonhuman primate (Fuster and Alexander, 1971; Niki, 1974; Funahashi et al., 1989), has been observed in numerous areas of the human brain with functional magnetic resonance imaging (fMRI; Curtis and D'Esposito, 2003). The idea of a role for delay-period activity in storage is reinforced by its sensitivity to manipulation of memory-influencing factors, such as persistence across varying delay lengths, and variation of signal magnitude with memory load (Vogel and Machizawa, 2004; Postle, 2006; Xu and Chun, 2006).

There are, however, complications with the mnemonic interpretation of delay-period activity. One is cases of activity that appear mnemonic but can be shown to support other functions, such as attention or response preparation (Fuster, 2002; Lebedev et al., 2004). Furthermore, lesion-induced deficits originally interpreted as mnemonic (Jacobson, 1936; Funahashi et al., 1993) have subsequently been reinterpreted as reflecting factors other than memory per se (Malmo, 1942; Tsujimoto and Postle, 2012).

A second complication is that delay-period activity can fail to show properties thought necessary for a mnemonic signal. In one such case, although fMRI activity at several sites was elevated throughout a long delay period (24 s), none showed load sensitivity, leaving uncertain whether these regions contribute to storage (Jha and McCarthy, 2000). In another, monkeys showed excellent short-term memory (STM) for direction of motion despite the absence of directionally tuned neurons in either the middle temporal area (MT) or the prefrontal cortex (PFC) that sustained elevated activity across the delay (Bisley et al., 2004; Zaksas and Pasternak, 2006; Hussar and Pasternak, 2012).

A third complication relates to assumptions of homogeneity of function, often only implied or tacit but nonetheless inherent, in massively univariate analyses of neuroimaging data. When activity is identified in a large volume of "activated" tissue, the extraction of a spatially averaged signal from contiguous voxels necessarily assumes that all are doing the same thing. Furthermore, its interpretation entails assuming that this locally homogeneous activity can be construed as supporting a mental function independent of other brain areas. These assumptions, however, are difficult to reconcile with the increasingly common recognition that neural representations are high dimensional and supported by anatomically distributed, dynamic computations (Kriegeskorte et al., 2006; Norman et al., 2006; Bullmore and Sporns, 2009; Cohen, 2011).

Given these complications, we sought to test core assumptions about elevated delay-period fMRI activity using an

Received March 30, 2012; revised July 13, 2012; accepted July 20, 2012.

Author contributions: A.C.R. and B.R.P. designed research; A.C.R. performed research; A.C.R. analyzed data; A.C.R. and B.R.P. wrote the paper.

This work was supported by the National Institutes of Health Grant R01-MH064498 (B.R.P.).

The authors declare no competing financial interests.

Correspondence should be addressed to Adam C. Riggall, Department of Psychology, University of Wisconsin–Madison, W. J. Brogden Psychology Building, 1202 West Johnson Street, Madison, WI 53706. E-mail: riggall@wisc.edu.

DOI:10.1523/JNEUROSCI.1892-12.2012

 0270-6474/12/3212990-09$15.00/0

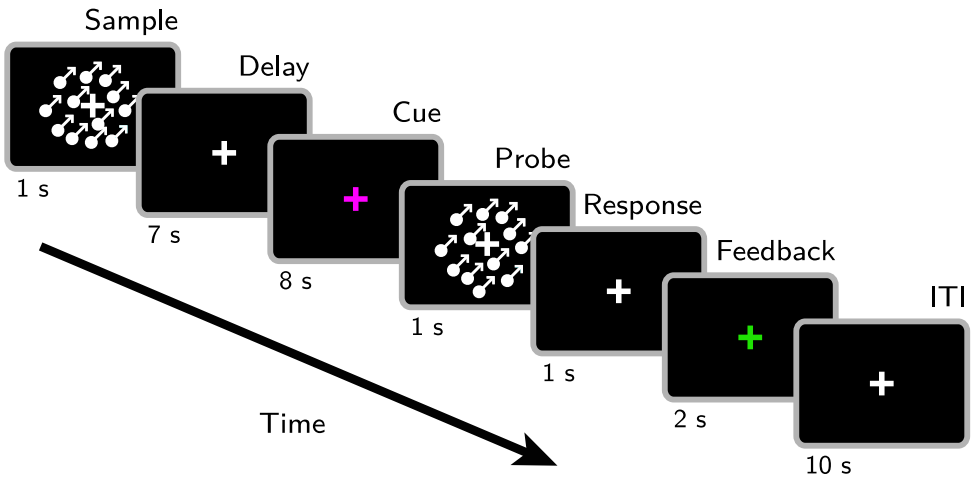

Figure 1. Behavioral task. Participants maintained the direction and speed of a sample motion stimulus over a long delay period. Midway through this delay period, they were cued as to the dimension on which they would be making an upcoming comparison, either direction or speed. At the end of the delay period, they were presented with a probe motion stimulus and had to indicate with a button press whether it matched or did not match the sample stimulus on the cued dimension.

information-based analysis. Multivariate pattern classification was used to test two hypotheses about STM for visual motion information: (1) that elevated delay-period activity carries trial-specific stimulus information and (2) that trial-specific stimulus information can be encoded in subthreshold patterns of activity. The first tests an assumption that has underlain most neuroimaging research on WM. Confirmation of the second would extend findings for visual STM for oriented gratings (Harrison and Tong, 2009; Serences et al., 2009) and offer insight about the physiological basis of STM for visual motion (Hussar and Pasternak, 2012).

Materials and Methods

To test our hypotheses, we scanned subjects (fMRI) while they performed a delayed-recognition task for visual motion (Fig. 1). We then trained pattern classifiers to discriminate the direction of motion from individual time points in the trial and tested to see how these classifiers labeled the data from all the other time points during the trial (creating a decoding time series). A similar approach has been shown to be sensitive to the dynamics of memory content, capturing changes in the memory trace on a continuous basis throughout the trial (Polyn et al., 2005; Lewis-Peacock and Postle, 2008, 2012).

In contrast to analysis methods that assess the magnitude of activity [whether the firing rate of a neuron or the strength of the blood oxygen level-dependent (BOLD) signal of one or a group of voxels], information-based analyses, such as pattern classification, focus on obtaining quantitative measures of the information content within a given area (Kriegeskorte et al., 2006; Kriegeskorte, 2011). This approach, often termed multivoxel pattern analysis (MVPA), uses machine learning methods to identify patterns of activity that are reliably associated with different stimuli or categories of stimuli. The extent to which novel patterns of activity can then be correctly categorized provides a measure of the information available in the underlying voxels (Norman et al., 2006; Haynes, 2011; Jimura and Poldrack, 2012).

Participants. Ten volunteers (five females) between 21 and 28 years of age (mean, 23.8 years) were recruited from the undergraduate and graduate student community of the University of Wisconsin-Madison and were paid for their participation. All subjects had normal or corrected-to-normal vision, no reported history of neurological disease, and no other contraindications for MRI. All subjects gave written informed consent according to the procedures approved by the Health Sciences Institutional Review Board at the University of Wisconsin-Madison. Three subjects (one female) were excluded from Results because of failure to perform the task to criterion level (for more details, see Results).

Behavioral paradigm. Participants were scanned while performing seven runs of a delayed-recognition task for visual motion. A schematic representation of the experiment is illustrated in Figure 1. Each trial began with a fixation cross changing color to white, indicating to subjects that they needed to fixate the cross and prepare for the start of the trial. After 1.5 s, a patch of coherently moving dots was presented (sample, 1 s). Participants were instructed to remember both the direction and the speed of this motion while maintaining fixation for the duration of the trial. The sample was followed by a 15 s delay period. Seven seconds into the delay period, the fixation cross changed color to indicate the dimension on which subjects would be making a match/non-match comparison judgment between the remembered motion and a new set of moving dots, blue indicated they should make the judgment based on the direction of the moving dots, ignoring speed, and magenta indicated that they should make the judgment based on speed, ignoring direction. After the delay, a second patch of coherently moving dots (probe, 1 s) was presented and subjects were required to indicate as quickly as possible with one of two buttons if the probe motion matched or did not match the sample on the cued dimension. After a 1 s response period, the fixation cross changed color to green if the subjects were correct or red if they were incorrect (feedback, 2 s). A 10 s intertrial interval (ITI) followed, during which the fixation cross changed color to gray and subjects were instructed they could break fixation and relax their eyes.

Sample and probe stimuli consisted of circular patches (15° diameter) of coherently moving dots. Sample motion could be in one of four directions (42°, 132°, 222°, 312°) and at one of three speeds (4°/s, 8°/s, 12°/s). Directions were chosen to be off the cardinal axes to reduce potential verbalizations. Probe stimuli on match trials had the same value in the cued dimension and a different value from the non-cued dimension (e.g., for a match trial in which direction was cued the sample might have moved toward 132° at 8°/s, whereas the matching probe might have moved toward 132° at 4°/s). On direction-cued non-match trials, the probe stimulus was rotated (randomly clockwise or counterclockwise) by a degree threshold value from the direction of sample, and the speed was randomly drawn from one of the two values not used in the sample. On speed-cued non-match trials, the probe stimulus speed was changed (randomly increased or decreased) by a proportional threshold value from the speed of the sample, and the direction was randomly drawn from one of the three values not used in the sample. The threshold values for non-match direction and speed were updated using separate adaptive staircases (Levitt, 1971) to keep performance ~75% correct. The logic of this approach was that, by holding task difficulty constant throughout an experimental session, we could assume comparable fidelity of representation on trials for which the response ended up being correct versus incorrect. This would maximize sensitivity by allowing inclusion of all trials in the analyses. Before scanning, subjects practiced a block of 24

trials to familiarize themselves with the experimental procedure and to determine starting threshold values for non-match trials.

Participants completed 168 trials over the course of seven runs while in the scanner. Sample stimuli included all possible pairwise combinations of directions and speeds (12 different combinations, each seen 14 times). Direction was cued on 96 trials, with the remaining 72 cued for speed. This disparity balanced the overall number of individual examples at each dimension value across the cued dimensions. Match and non-match trials were equally likely. The experimental stimuli were controlled by E-Prime 2.0 (Psychology Software Tools) and viewed through fiber-optic goggles mounted on the head coil (SV-7021; Avotec). Participants responded via two buttons on a fiber-optic button box (Psychology Software Tools).

Subjects also performed one block of an area MT+ (which contains both area MT and medial superior temporal area) localizer task, similar to that used by Huk et al. (2002). In summary, participants viewed alternating 18 s blocks of stationary and moving dot patterns (alternating from expanding and contracting every second) within a circular aperture (15°) while maintaining fixation, for a total of eight blocks of each.

Data acquisition and preprocessing. Whole-brain images were acquired with the 3 T scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. For all subjects, a high-resolution T_1-weighted image was acquired with a fast spoiled gradient-recalled-echo sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle, 156 axial slices, 256 × 256 in-plane, 1.0 mm isotropic). A gradient-echo, echo-planar sequence (2 s TR, 25 ms TE) was used to acquire data sensitive to the BOLD signal within a 64 × 64 matrix (39 sagittal slices, 3.5 mm isotropic). Seven runs of the delayed-recognition task were obtained for each subject, each lasting 12 min, 8 s (364 volumes). All task runs were preceded by 10 s of dummy pulses to achieve a steady state of tissue magnetization. One run of the MT+ localizer was obtained for each subject, lasting 4.8 min (144 volumes).

The functional data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). All volumes were spatially aligned to the final volume of the final run using a rigid-body realignment and corrected for slice time acquisition. Linear, quadratic, and cubic trends were removed from each run to reduce the influence of scanner drift. For univariate analyses, data were spatially smoothed with a 6 mm FWHM Gaussian kernel and transformed into Talairach space (Talairach and Tournoux, 1988). For classification analyses, data were *z*-scored separately within run for each voxel. Data were not smoothed and were left in their native space.

Univariate analyses. Each within-trial event of the delayed-recognition task (i.e., sample, pre-cue delay, cue, post-cue delay, probe, response) was modeled separately for direction- and speed-cued trials. Sample and probe were modeled as 1 s boxcars, each delay as a boxcar of appropriate duration, and the cue and response as impulses. All were convolved with a canonical hemodynamic response function. Each of these independent regressors was entered into a modified general linear model (GLM) for analysis using AFNI. For the present purposes, a more generous boxcar-shaped covariate was used to model the delay periods [rather than a more conservative mid-delay delta function (Zarahn et al., 1999; Postle et al., 2000)] to ensure that we would not miss any delay-active voxels, although at the expense of likely also being sensitive to some variance that is attributable to the sample presentation. In this way, we implemented a generous feature selection step that included as many "delay active" voxels as possible for MVPA, thereby being careful not to exclude any such voxels that may potentially carry stimulus-specific information. The localizer was modeled with boxcars for both stationary and moving dot patterns. This localizer was used to ensure that the regions of interest used for the MVPA included MT+.

Pattern classification analyses. Classification was performed using the Princeton Multi-Voxel Pattern Analysis (www.pni.princeton.edu/mvpa) toolbox and custom routines in MATLAB (MathWorks). Preprocessed fMRI data from individual trial time points were used to train separate classifiers to classify the direction of motion (four possible directions) or the speed of motion (three possible speeds) in the sample and, by inference, the direction/speed of motion in memory (stimulus-specific classification), or to classify whether the subject had been cued that direction or speed was the relevant dimension on a given trial (trial-dimension classification).

Classification was accomplished using L2-regularized logistic regression, a linear classification approach used widely in the machine learning community well suited for application to fMRI data because it tends to generalize well after learning in high-dimensional feature spaces with limited training examples (Pereira et al., 2009). The λ penalty term, which reduces the contribution of less informative voxels to classification and thus improves generalization, was determined ($\lambda = 25$) by repeating the whole-brain testing procedure described below for penalty terms at powers of 10 from -4 to 4 and then at a finer-grained resolution within the best interval. The penalty term was chosen to maximize the mean decoding performance across all subjects. During decoding, a trial was considered correctly classified if the correct direction/speed had the highest likelihood estimate (winner-take-all classification). Overall classification accuracy was determined using leave-one-trial-out cross-validation, in which the classifier was repeatedly trained on data from all but one trial, and then tested on the left out trial, rotating through all the trials as the left-out testing trial.

Stimulus-specific classifiers were always trained with data from direction-cued trials or speed-cued trials, never both. Testing of trained classifiers was done on trials of both type. This allowed us to compare how representations changed when subjects were cued that they would be judging speed, and thus direction was no longer relevant, and vice versa. Trial-dimension classifiers were trained and tested on all data.

To examine the dynamics of the memory trace, each classifier was trained using data from only a single time point in the trial (e.g., the first volume acquisition after the target) and then tested on all time points in the left-out trial (i.e., including both time points before and after the training time point). The result of this procedure is a time course of decoding accuracy for the entire trial. By doing the initial training of the classifier using different time points in the trial (e.g., a time point just after the sample, a time point in the later part of the delay, etc.), it was possible to estimate the stability of a given representation throughout the duration of the trial.

Classification was initially performed on whole-brain data that had undergone a basic feature-selection step whereby only those voxels that showed a main effect for task ($t > 2$) in the univariate GLM were included. This step was included to reduce the chances of overfitting during training. Subsequent region of interest (ROI)-based analyses used only those voxels within individual ROIs, created from the intersection of anatomically defined ROIs and voxels that showed either significant sustained delay-period activity or no delay-period activity, depending on the specific hypothesis. Four anatomically defined ROIs were hand drawn for each subject by tracing gray matter on the high-resolution anatomical scans: frontal, parietal, lateral occipital and temporal, and medial occipital. The frontal region included the entire precentral sulcus (PCS) and the posterior portion of the inferior frontal sulcus (IFS). For finer-grained analyses, this was subdivided into three frontal ROIs that showed robust delay-period activity: (1) the superior rostral bank of the PCS bounded superiorly by the intersection of the superior frontal sulcus (SFS); (2) a more inferior portion of the rostral bank of the PCS bounded ventrally by the intersection with the IFS; and (3) the caudal third of the IFS. The parietal region included the entire intraparietal sulcus (IPS) and the superior parietal lobule (SPL). For finer-grained analyses, it was also subdivided into three ROIs: (1) medial–caudal IPS comprising the descending segment and the caudal half of the horizontal segment; (2) dorsolateral IPS comprising the rostral half of the horizontal segment and the ascending segment; and (3) SPL. The lateral occipital and temporal region included all of the lateral occipital gyrus, the fusiform gyrus, the posterior portion of the middle and inferior temporal gyri, and the posterior portion of the inferior temporal sulcus. The medial occipital region covered the medial portion of the occipital lobe from the lingual sulcus to the occipitoparietal sulcus, including all of the calcarine sulcus.

The significance of classifier performance was determined using a random permutation test (Golland and Fischl, 2003) to determine the likelihood of observing a specific accuracy under the null hypothesis that there is no relationship between the data and the specific class labels used to train the classifier (directions/speeds of motion). A null distribution

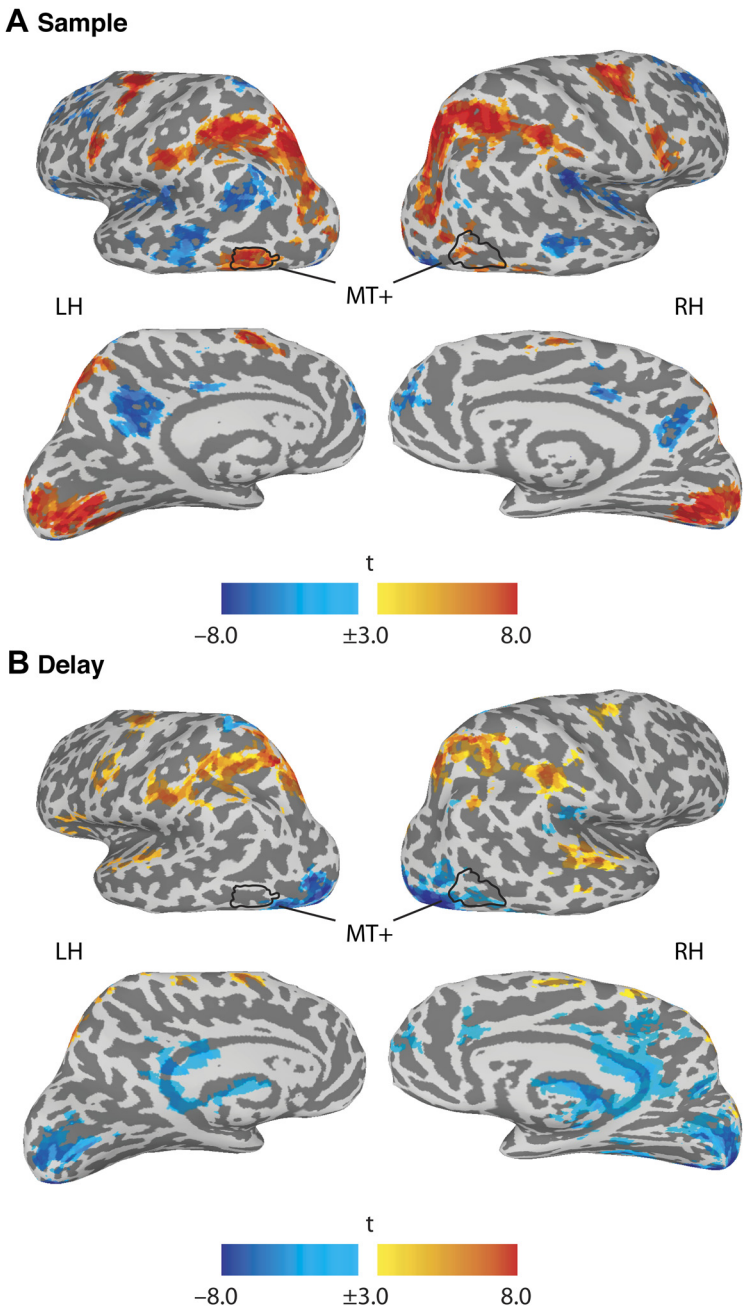

Figure 2. Univariate GLM results. Sample-evoked (***A***) and delay-related (***B***) activity, as estimated from a group-level GLM, thresholded at $p < 0.05$, FDR corrected, and displayed on a representative subject's inflated surface. Note that images are for illustrative purposes only, because all analyses were performed on single-subject data. Superimposed is an outline of MT+ as defined by the localizer. Note for this region that it is robustly activated by the sample but that its activity does not differ from the baseline during the delay period. A qualitatively similar pattern is observed in calcarine and pericalcarine cortex. LH, Left hemisphere; RH, right hemisphere.

was generated by rerunning the entire classification cross-validation procedure 1000 times, randomly shuffling the class labels each time. A p value was then computed by determining the proportion of permuted accuracies that were higher than the observed accuracy. This procedure was repeated for all classification results.

Results

Behavioral results

Task difficulty was equated across subjects by feeding real-time, trial-by-trial performance information to a staircasing algorithm that dynamically adjusted the difficulty of non-matching probe stimuli (direction and speed independently) so as to maintain performance at a target level of 75% correct. Overall mean behavioral performance ($n = 7$) was 80.38% correct for direction trials and 80.32% correct for speed trials, both slightly better than the target performance level because several subjects reached a predefined minimum threshold value, at which point the staircase procedure could no longer reduce the threshold to further increase task difficulty. The average ± SD non-match direction threshold was 10.9° ± 5.33°. The average ± SD non-match speed threshold was a 38% ± 15.3% change. Three subjects were dropped from the study because their inordinately high thresholds indicated that they were responding at random.

Univariate results

To test the first hypothesis that elevated delay-period activity carries trial-specific stimulus information and can thus be interpreted as a neural correlate of memory storage, we first identified areas showing elevated, sustained delay-period activity by solving a massively univariate GLM for each subject. The resultant individual thresholded statistical maps were then used, on a subject-by-subject basis, to select the voxels used for training the classifiers. Group-level statistical maps for the sample-evoked (Fig. 2*A*) and delay-period (Fig. 2*B*) activity illustrate several important characteristics. Activity evoked by the sample was widespread, located bilaterally in posterior visual areas, lateral occipitotemporal regions including the inferior temporal sulcus (including the putative MT+ complex identified with the localizer), IPS, posterior IFS and SFS, and PCS. Delay-period activity was more circumscribed, with clusters of significant activity [$p < 0.05$, false discovery rate (FDR) corrected] bilaterally in the IPS, IFS, and PCS. This pattern of elevated activity agrees with many other findings from studies of visuospatial WM (Curtis and D'Esposito, 2003).

Whole-brain pattern classification

Before directly testing our hypotheses, it was necessary to demonstrate that motion information could be decoded on a time point-by-time point basis. All analyses were performed on single-subject data, with statistical reliability subsequently assessed across the sample. For each subject's data, a set of four classifiers was trained to discriminate the direction of the sample motion stimulus (and, by inference, the remembered motion direction; four possible directions), each using feature-selected whole-brain data from direction-cued trials and restricted to a single time point within the trial. The first classifier ("sample") was trained on data from only the time point 4 s after

the sample onset, which corresponds to the peak of the sample-evoked response. The second classifier ("late delay") was trained on data from the time point 16 s after sample onset, corresponding to the post-cue portion of the delay period, before the onset of the probe stimulus. The third classifier ("probe") was trained on data from the time point 20 s after sample onset (4 s after probe onset), corresponding to the peak of the probe-evoked response. The final classifier ("ITI") was trained on data from the time point 26 s after sample onset, corresponding to the middle of the ITI. This classifier was included as a control, because we would not expect there to be any stimulus-specific information retained once the trial had been completed.

Each trained classifier was then used to construct a decoding time course of direction representation on every trial time point from the held-out direction-cued trials. This approach allowed us to detect changes in the neural representation of information across the trial. For example, by training a classifier on data from very early in the trial (i.e., sample), we would expect to capture early, sensory-based representations. By testing such a classifier on every time point in the trial, we could determine whether such a representation remains stable throughout the trial or whether it deteriorates over time. Similarly, training a classifier with data from a time point late in the delay period (late delay) might capture a recoded representation (e.g., verbal or numeric or clock face), which would be expected to be absent at the beginning of the trial and to strengthen over the course of the delay period when we look at the time point-by-time point decoding performance of that classifier. By limiting our training data to a single time point during the trial, we hoped to maximize our ability to resolve time point-by-time point changes, at the expense of lower signal-to-noise for the classifier inputs.

Results from this analysis can be seen in Figure 3*A*. The mean decoding accuracy for the sample classifier was significantly above chance (25%, $p < 0.05$, permutation test) throughout the entire delay period. Similar results were obtained with the late delay and probe classifiers. These results suggest that the memory representation remains relatively stable and unchanging across the delay period. Decoding accuracies for the ITI classifier were always near chance, which was expected given the absence of any stimulus-related information for the classifier to learn (i.e., signal from 8 s after the offset of the memory probe would not be expected to carry information about the stimulus from the preceding trial). As with the univariate analyses, all trials were included in the analyses. Follow-up analyses using only correct trials produced qualitatively similar results.

To assess the specificity of the classification, the set of classifiers trained to discriminate direction with data from direction-cued trials were also used to decode direction information from speed-cued trials (Fig. 3*B*). Classification with the sample, late delay, and probe classifiers was above chance for time points before the cue, when subjects needed to hold both speed and direction information in memory, but fell to chance levels after the cue, suggesting that subjects discarded direction information when it was no longer relevant to the current trial.

Two features in the data confirm that successful late-delay decoding of direction-cued trials represents the sustained retention of stimulus information and does not reflect an artifact of the slow recovery of the hemodynamic response or effects of motion adaptation. First, if the results were purely driven by the residual hemodynamic response to the sample stimulus or to adaptation, we would expect to see similar decoding performance on both direction-cued and speed-cued trials, because the sample stimuli are identical across these two conditions. Second, such "residual" effects would also result in successful decoding by the ITI classifier, which was trained with data 10 s after the probe stimulus (compared with 16 s after the sample for the late delay), which it clearly does not.

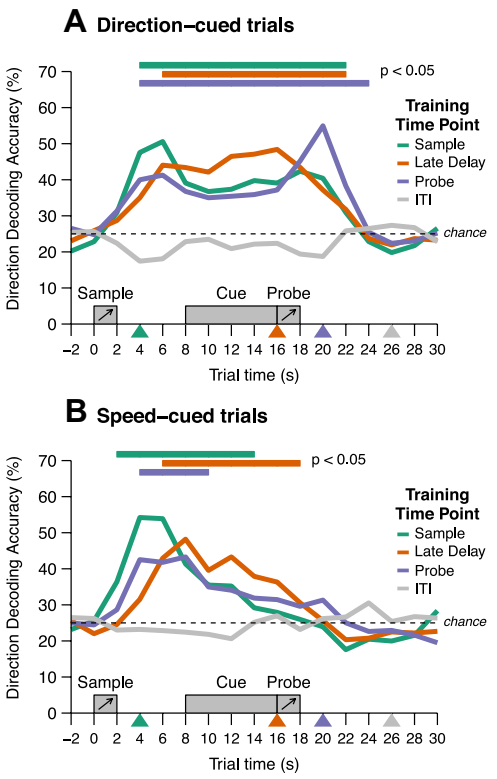

Figure 3. Whole-brain direction-decoding results. Decoding time courses after training on whole-brain data from direction-cued trials. ***A***, Decoding of direction information from direction-cued trials. ***B***, Decoding of direction information from speed-cued trials. Each waveform represents the mean direction-decoding accuracy across subjects ($n = 7$) for a classifier trained with data limited to a single time point in the trial and then tested on all time points in the holdout trials (e.g., the green line illustrates the decoding time course from a classifier trained on only data from time point 4, indicated by the small green triangle along the *x*-axis.) Horizontal bars along the top indicate points at which the decoding accuracy for the corresponding classifier was significantly above chance ($p < 0.05$, permutation test). Schematic icons of trial events are shown at the appropriate times along the *x*-axis. Data are unshifted in time.

The whole-brain classification procedure was also applied to the stimulus dimension of speed, using data from only speed-cued trials. Unlike direction, however, classifier performance never exceeded chance for any time point during the trial. Because subjects performed these trials at the same level of proficiency as direction-cued trials, we interpret this null result to mean that the representation of speed, at least across the range used in this experiment, may be performed on too fine a spatial scale neurally to be discriminated with our fMRI procedure. It is also possible that, by collapsing across different directions when classifying the speed trials, we added too much noise to the signal to classify, given the close relationship between speed and direction (Born and Bradley, 2005). Additionally, the use of only three speeds may have encouraged a coding strategy that varied over time (e.g., verbal labels that changed as the stimuli became more familiar to subjects). The remainder of Results and Discussion will focus on decoding direction information.

ROI-constrained classification

To test our first hypothesis—that elevated delay-period activity carries stimulus-specific information—we repeated the classifi-

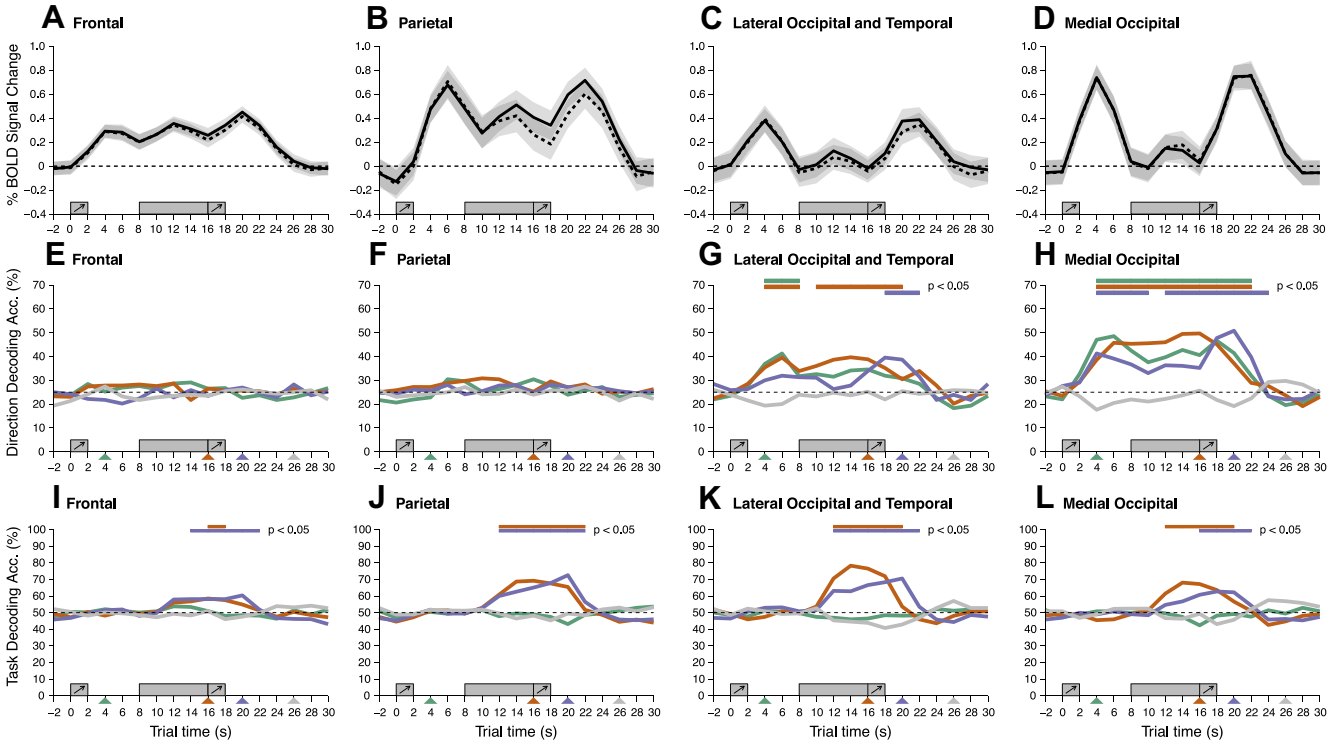

Figure 4. ROI BOLD and decoding time courses. ***A–D***, Average ROI BOLD activity. Data from direction-cued trials use solid lines, and speed-cued trials use dashed lines; bands cover average SE across subjects. ROI stimulus-direction decoding results (***E–H***) and ROI trial-dimension decoding results (***I–L***). Graphical conventions same as Figure 3. All averaged across individual data from seven subjects.

cation procedure that we used with the whole-brain data but applied it to only those voxels in the frontal or parietal cortices that showed elevated delay-period activity as identified with the univariate analysis. As can be seen in the average BOLD time series in Figure 4, *A* and *B*, each of these areas showed elevated activity that was sustained throughout the delay period. However, decoding performance for motion direction never differed from chance in either area, regardless of the time point used to train the classifier (Fig. 4*E*,*F*). Therefore, we failed to find evidence that the sustained, delay-period BOLD activity in these regions carried stimulus-related information; the first hypothesis was not supported. To rule out the possibility that using such large ROIs may have obscured the presence of stimulus representation in smaller regions, we repeated the ROI classification with the smaller, more specific ROIs described in Materials and Methods. Results in all smaller ROIs mirrored those of the larger regions: no evidence for stimulus-related information was found in any of the regions.

To test our second hypothesis—that stimulus-specific information can be recovered from subthreshold patterns of activity (i.e., activity whose signal intensity does not surpass a statistical threshold in conventional univariate analysis)—we repeated the classification procedure as above, limiting the training data to only voxels in medial occipital or lateral occipital and temporal cortex that showed no evidence of elevated activity during the delay period in the GLM. As can be seen in the average BOLD time series in Figure 4, *C* and *D*, these regions showed large sample- and probe-evoked responses, as well as smaller cue-evoked responses, but no sustained, elevated delay-period activity. Decoding performance from these regions, however, was significantly above chance throughout the delay period (Fig. 4*G*,*H*).

Although these posterior regions did not show sustained delay-period activity at the group level (Fig. 2*B*), there were voxels in the individual-subject data of each subject that did. The results of pattern classification did not change appreciably when these voxels were included or excluded from the analyses. Additionally, when the classifiers were trained only on these posterior delay-active voxels, in no subject was decoding performance sustained at above-chance levels across the delay period. Overall, these findings were consistent with the second hypothesis, that brain regions can carry stimulus-specific information in a sustained manner despite the absence of sustained, above-baseline levels of activity.

These results form a double dissociation, with frontal and parietal regions showing elevated, sustained delay-period activity but no delay-period stimulus representation, and posterior regions the converse. One possible concern about applying these results to our understanding of WM storage, however, is that doing so requires the acceptance of the null MVPA findings in frontal and parietal cortex. Might it be the case, for example, that these regions are simply less amenable to MVPA (e.g., perhaps because they represent information at a finer grain of spatial detail than is measurable with our fMRI methods)? To address this possibility, we trained a new set of classifiers on a different discrimination—cue identity (i.e., whether the relevant stimulus dimension was direction or speed)—and repeated the procedure with each of the four ROIs. As shown in Figure 4*I–L*, the relevant trial dimension was decodable from each of the four ROIs. Importantly, for each, this was only true for the late delay and probe time-point-trained classifiers and only for time points after cue presentation. These results indicate that frontal and parietal regions are not inherently "undecodeable" with our methods and, thus, lend more credence to the possibility that they did not rep-

resent stimulus-specific information during the delay period of our WM task.

Discussion

The aim of the present study was to test long-standing views about the relationship between the short-term retention of information and sustained delay-period activity. Using an information-based analysis approach with fMRI data collected during a delayed-recognition task for visual motion, we tested two hypotheses: (1) that sustained, elevated delay-period activity carries stimulus-specific information; and (2) that stimulus information can be encoded in distributed patterns of subthreshold activity. To test the first hypothesis, we trained pattern classifiers with BOLD signal from frontal and parietal areas that showed sustained, elevated delay-period activity. We failed to find evidence that these voxels carried stimulus-specific information during the delay period. To test the second hypothesis, we applied the same procedure to BOLD data from posterior regions that showed robust responses to visual stimuli but no elevated delay-period activity. The classifiers were successfully able to decode the remembered direction throughout the delay period, providing strong evidence in support of this hypothesis.

The first finding can be seen as a failure to support an enduring assumption in cognitive neuroscience, albeit one that is increasingly being called into question (Curtis and D'Esposito, 2003; Lebedev et al., 2004; Curtis and Lee, 2010; Lewis-Peacock and Postle, 2012). Although on its own it might be qualified as a null result, there are several factors that must influence its interpretation. Most saliently, it is paired with a positive result using the same method and derived from statistically "subthreshold" voxels located in areas that are active during the perception of the to-be-remembered information. Empirical evidence thus shows that this method is sensitive. Indeed, although there remains some controversy about the physiological and representational factors that underlie the patterns of activity that correspond to different brain states (Freeman et al., 2011; Thompson et al., 2011), we are not familiar with any suggestion that there may exist brain states to which MVPA is less sensitive than traditional analysis of activation levels of individual voxels or groups of voxels. To the contrary, the near-consensus view is that MVPA methods are more sensitive than traditional activation-based analyses (Kriegeskorte et al., 2006; Norman et al., 2006; Haynes, 2011; Jimura and Poldrack, 2012; Lewis-Peacock and Postle, 2012).

Furthermore, although we cannot rule out the possibility that stimulus information might be represented in frontoparietal cortex at either a spatial scale that is too fine to be detected with our fMRI methods or perhaps via a signal to which BOLD is relatively insensitive (e.g., low-frequency oscillations in local field potentials), we did demonstrate that this is not a limitation for the decoding of trial-specific task instruction-related information. From this perspective, our results are consistent with, for example, the finding from monkeys that PFC and posterior parietal cortex represent the category to which a stimulus belongs (Freedman and Assad, 2006; Swaminathan and Freedman, 2012). It is also worthy of note that, although MVPA has been applied successfully to sensory processing in topographically organized cortex [e.g., the decoding of orientation (Harrison and Tong, 2009; Serences et al., 2009)], it has also been successfully applied to "higher-level" processing in polymodal cortex. Thus, for example, MVPA has demonstrated contextual reinstatement during episodic memory retrieval (Polyn et al., 2005), the recognition of individual faces (Kriegeskorte et al., 2007), and neural correlates of free choice (Soon et al., 2008), all entailing the decoding of information from polymodal temporal, parietal, and/or frontal cortex.

Consistent with our preferred interpretation of the null findings in frontal cortex are several factors. First, there are the results from extracellular recording in monkeys performing a similar task with similar stimuli, in which no evidence for direction-selective persistent activity was found in the PFC throughout the delay period (Zaksas and Pasternak, 2006; Hussar and Pasternak, 2012). Second, a similar pattern to the MVPA results that we describe here has been reported for STM for four categories of visual objects (Linden et al., 2012) and for complex artificial visual stimuli (Christophel et al., 2012). Third, the fact that STM can be intact despite lesions of PFC (D'Esposito and Postle, 1999; Tsujimoto and Postle, 2012) is consistent with the failure to find physiological evidence for STM representations in this region.

The frontoparietal network that has been a focus of this study is known to support the endogenous control of attention (Corbetta and Shulman, 2002; Beck and Kastner, 2009; Noudoost et al., 2010). Interestingly, one account of WM storage is that it is supported by this same top-down mechanism (Curtis and D'Esposito, 2003; Armstrong et al., 2009; Postle, 2011). From this perspective, the sustained delay-period activity observed in this study may correspond to a control signal that does not vary with stimulus identity. Future work would need to reconcile this possibility with the finding that multivariate patterns of frontoparietal activity do discriminate between directions of motion during a sustained attention task (Liu et al., 2011). In addition to specifically memory-related functions, many other functions might be supported by sustained delay-period activity of frontal and parietal regions. Because the frontal and parietal activity observed in the present study (Fig. 4) resembles activity that has been reported in countless previous neuroimaging studies (Curtis and D'Esposito, 2003), it may well be that it does not correspond to a stimulus-specific or even task-specific function. More general demands that many cognitive tasks (including WM tasks) have in common include decision making (Curtis and Lee, 2010), prioritizing certain task-relevant representations and/or processes over others (Miller and Cohen, 2001), monitoring the environment to control the processing of potentially interfering exogenous events (Chao and Knight, 1998; Postle, 2005), actively representing a "behavioral set" (Woolgar et al., 2011), and monitoring behavior so as to prevent prepotent responses (Knight and D'Esposito, 2003), including perseverative responses (Milner, 1963; Tsujimoto and Postle, 2012). (Note that, although the behavioral set account might be consistent with the successful decoding of cue identity in frontal and parietal regions, this explanation does not generalize to the first portion of the delay period.) This is, of course, an incomplete list.

One important question for future study is the nature of the mental codes with which subjects represent motion information across the delay period. In the monkey, a psychophysical study using backward masking provided evidence that the initial memory trace is perceptually based, retaining a high-fidelity representation of the sample (including such trial-irrelevant information as the local velocity of individual dots in the random-dot motion stimulus). However, this representation only endured a few hundred milliseconds into the delay period, perhaps because, in this study, the animals could predict the major features of the impending memory probe (Zaksas et al., 2001). Although the BOLD signal did not afford high temporal resolution in the present study, results with classifiers trained on different portions of the trial suggested that the mnemonic representation is relatively sta-

ble. We cannot know with certainty, however, whether this representation was primarily perceptual, motoric, or categorical in nature, or perhaps some combination of these. Our working assumption is that the mnemonic representation of direction was perceptually based, because it is from visual regions that we were able to recover stimulus direction information. Had subjects used, for example, a covert eye-movement strategy, we would have expected to have been able to decode stimulus information from frontal and parietal regions (Ikkai and Curtis, 2011). The same reasoning makes us skeptical that subjects depended on a verbal strategy for remembering either direction or speed. We did not, however, monitor eye movements, nor did we take steps to discourage covert speech.

The results presented here highlight the differing conclusions that can be drawn from activation- versus information-based analyses of the same dataset. In so doing, they raise questions about the longstanding belief that information retained during WM is stored via sustained delay-period activity, preferentially in frontal and parietal cortex. Instead, the memory trace may be represented in patterns of subthreshold levels of activity distributed across regions of low-level sensory cortex.

References

Armstrong KM, Chang MH, Moore T (2009) Selection and maintenance of spatial information by frontal eye field neurons. J Neurosci 29:15621–15629.

Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Res 49:1154–1165.

Bisley JW, Zaksas D, Droll JA, Pasternak T (2004) Activity of neurons in cortical area MT during a memory for motion task. J Neurophysiol 91:286–300.

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Chao LL, Knight RT (1998) Contribution of human prefrontal cortex to delay performance. J Cogn Neurosci 10:167–177.

Christophel TB, Hebart MN, Haynes JD (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci 32:12983–12989.

Cohen MX (2011) It's about time. Front Hum Neurosci 5:2.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 7:415–423.

Curtis CE, Lee D (2010) Beyond working memory: the role of persistent activity in decision making. Trends Cogn Sci 14:216–222.

D'Esposito M, Postle BR (1999) The dependence of span and delayed-response performance on prefrontal cortex. Neuropsychologia 37:1303–1315.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not columns. J Neurosci 31:4792–4804.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas." J Neurosci 13:1479–1497.

Fuster JM (2002) Physiology of executive function: the perception-action cycle. In: Principles of frontal lobe function (Struss DR, Knight RT, eds), pp 96–108. Oxford: Oxford UP.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Golland P, Fischl B (2003) Permutation tests for classification: towards statistical significance in image-based studies. Inf Process Med Imaging 18:330–341.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Haynes JD (2011) Multivariate decoding and brain reading: introduction to the special issue. Neuroimage 56:385–386.

Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. J Neurosci 22:7195–7205.

Hussar CR, Pasternak T (2012) Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. J Neurosci 32:2747–2761.

Ikkai A, Curtis CE (2011) Common neural mechanisms supporting spatial working memory, attention and motor intention. Neuropsychologia 49:1428–1434.

Jacobson C (1936) Studies of cerebral functions in primates. I. The functions of the frontal association areas in monkies. Comp Psych Monog 13:1–60.

Jha AP, McCarthy G (2000) The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. J Cogn Neurosci 12 [Suppl 2]:90–105.

Jimura K, Poldrack RA (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50:544–552.

Knight RT, D'Esposito M (2003) Lateral prefrontal syndrome: a disorder of executive control. In: Neurological foundations of cognitive neuroscience (D'Esposito M, ed), pp 259–279. Cambridge, MA: Massachusetts Institute of Technology.

Kriegeskorte N (2011) Pattern-information analysis: from stimulus decoding to computational-model testing. Neuroimage 56:411–421.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104:20600–20605.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49 [Suppl 2]:467+.

Lewis-Peacock JA, Postle BR (2008) Temporary activation of long-term memory supports working memory. J Neurosci 28:8765–8771.

Lewis-Peacock JA, Postle BR (2012) Decoding the internal focus of attention. Neuropsychologia 50:470–488.

Linden DE, Oosterhof NN, Klein C, Downing PE (2012) Mapping brain activation and information during category-specific visual working memory. J Neurophysiol 107:628–639.

Liu T, Hospadaruk L, Zhu DC, Gardner JL (2011) Feature-specific attentional priority signals in human cortex. J Neurosci 31:4484–4495.

Malmo RB (1942) Interference factors in delayed response in monkeys after removal of frontal lobes. J Neurophysiol 5:295–308.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Milner B (1963) Effects of different brain lesions on card sorting: the role of the frontal lobes. Arch Neurol 9:100–110.

Niki H (1974) Differential activity of prefrontal units during right and left delayed response trials. Brain Res 70:346–349.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.

Noudoost B, Chang MH, Steinmetz NA, Moore T (2010) Top-down control of visual attention. Curr Opin Neurobiol 20:183–190.

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199–S209.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Postle BR (2005) Delay-period activity in the prefrontal cortex: one function is sensory gating. J Cogn Neurosci 17:1679–1690.

Postle BR (2006) Working memory as an emergent property of the mind and brain. Neuroscience 139:23–38.

Postle BR (2011) What underlies the ability go guide action with spatial

information that is no longer present in the environment? In: Spatial working memory (Vandierendonck A, Szmalec A, eds), pp 897–901. Hove, UK: Psychology Press.

Postle BR, Zarahn E, D'Esposito M (2000) Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. Brain Res Brain Res Protoc 5:57–66.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. Nat Neurosci 11:543–545.

Swaminathan SK, Freedman DJ (2012) Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat Neurosci 15:315–320.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.

Thompson R, Correia M, Cusack R (2011) Vascular contributions to pattern analysis: comparing gradient and spin echo fMRI at 3T. Neuroimage 56:643–650.

Tsujimoto S, Postle BR (2012) The prefrontal cortex and delay tasks: a reconsideration of the "mnemonic scotoma." J Cogn Neurosci 24:627–635.

Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. Nature 428:748–751.

Woolgar A, Thompson R, Bor D, Duncan J (2011) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. Neuroimage 56:744–752.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Zaksas D, Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. J Neurosci 26:11726–11742.

Zaksas D, Bisley JW, Pasternak T (2001) Motion information is spatially localized in a visual working-memory task. J Neurophysiol 86:912–921.

Zarahn E, Aguirre GK, D'Esposito M (1999) Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI. Brain Res Cogn Brain Res 7:255–268.