

Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex

Kartik K. Sreenivasan^{1,2}, Jason Vytlačil¹, and Mark D'Esposito¹

Abstract

■ The predominant neurobiological model of working memory (WM) posits that stimulus information is stored via stable, elevated activity within highly selective neurons. On the basis of this model, which we refer to as the canonical model, the storage of stimulus information is largely associated with lateral PFC (LPFC). A growing number of studies describe results that cannot be fully explained by the canonical model, suggesting that it is in need of revision. In this study, we directly tested key elements of the canonical model. We analyzed fMRI data collected as participants performed a task requiring WM for faces and scenes. Multivariate decoding procedures identified patterns of activity containing information about the items maintained in WM (faces, scenes, or both). Although information

about WM items was identified in extrastriate visual cortex (EC) and LPFC, only EC exhibited a pattern of results consistent with a sensory representation. Information in both regions persisted even in the absence of elevated activity, suggesting that elevated population activity may not represent the storage of information in WM. Additionally, we observed that WM information was distributed across EC neural populations that exhibited a broad range of selectivity for the WM items rather than restricted to highly selective EC populations. Finally, we determined that activity patterns coding for WM information were not stable, but instead varied over the course of a trial, indicating that the neural code for WM information is dynamic rather than static. Together, these findings challenge the canonical model of WM. ■

INTRODUCTION

Early single-unit investigations into the neural basis of working memory (WM) documented elevated firing in neurons in lateral PFC (LPFC) when a monkey was required to store information on-line to link a stimulus to a subsequent response (Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster, 1973; Fuster & Alexander, 1971; Kubota & Niki, 1971). This activity, termed “delay period activity,” has been interpreted by many (though, notably, not Fuster & Alexander, 1971, or Kubota & Niki, 1971) as representing the short-term maintenance of information about the to-be-remembered stimulus. These observations inspired a highly influential theoretical framework that has motivated several seminal findings in the study of WM and continues to shape the scope and tenor of WM research (Goldman-Rakic, 1995; Wilson, Scalaidhe, & Goldman-Rakic, 1993). We refer to this framework as the canonical model of WM.

There are several key tenets of the canonical model of WM. One tenet that is the subject of recent debate is the notion that LPFC neurons store information about the sensory features of memoranda in the service of WM. This view has been bolstered by the consistent observation of delay period activity in LPFC. However, recently

developed multivariate decoding methods, which rely on supervised learning algorithms to identify patterns of brain activity that represent specific types of information (Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006), offer potentially increased sensitivity relative to traditional univariate methods for localizing information content (Jimura & Poldrack, 2012). These methods have increasingly been applied to the study of how information is represented in WM (Sreenivasan, Curtis, & D'Esposito, in press). Several fMRI studies utilizing decoding methods have identified patterns of visual activity that code for sensory properties of visual items during WM for those items (Han, Berg, Oh, Samaras, & Leung, 2013; Xing, Ledgeway, McGraw, & Schluppeck, 2013; Christophel, Hebart, & Haynes, 2012; Riggall & Postle, 2012; Linden, Oosterhof, Klein, & Downing, 2011; Ester, Serences, & Awh, 2009; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009). Moreover, information about maintained visual items persists in visual cortex throughout the delay period, suggesting that sensory regions participate in the storage of WM information (Riggall & Postle, 2012; Harrison & Tong, 2009). At the same time, data from single-unit studies and one recent fMRI study indicate that multivariate patterns of LPFC activity also encode information about currently maintained visual WM stimuli (Lee, Kravitz, & Baker, 2013; Rigotti et al., 2013; Stokes et al., 2013; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008). Thus, the respective roles of these regions are unresolved. A

¹University of California, Berkeley, ²New York University Abu Dhabi

critical step in resolving the contributions of these regions to WM involves dissociating representations that code for sensory features from those that code for nonsensory features of WM items.

Another tenet of the canonical model is that WM information is encoded by neural populations that are highly selective for the maintained information. In line with this view, univariate analyses of WM data have largely focused on neural populations that respond preferentially to the features of the memoranda. However, in other contexts such as the formation of sensory representations during stimulus perception, information about stimulus properties is coded for by activity in populations with a wide range of selectivity for the properties of the stimulus being represented (Ewbank, Schluppeck, & Andrews, 2005; O'Toole, Jiang, Abdi, & Haxby, 2005; Cox & Savoy, 2003; Haxby et al., 2001). It remains unclear whether WM representations similarly recruit nonselective neural populations.

Perhaps the most central tenet of the canonical model is the idea that elevated, sustained delay period activity is the neural mechanism supporting the storage of WM information. Delay period activity is consistently demonstrated in monkey electrophysiological data as well as fMRI studies in humans (e.g., Zarahn, Aguirre, & D'Esposito, 1999; Courtney, Ungerleider, Keil, & Haxby, 1997) and has become synonymous with the storage of information in WM. A corollary of this property is that WM information is coded for in a static manner over the course of maintenance. That is, storage-related neural activity must persist in a stable form to hold WM representations in an active state. Accordingly, disruptions of delay period activity over time or because of external interference are thought to indicate a corruption of WM storage. Thus, inferences about a region's contribution to WM storage typically depend on the magnitude and temporal stability of delay period activity within that region (Artchakov et al., 2009; Schluppeck, Curtis, Glimcher, & Heeger, 2006; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Jha & McCarthy, 2000; Miller, Erickson, & Desimone, 1996). The relationship between temporally stable delay period activity and WM storage is called into question by recent work that finds evidence for WM information in regions that do not exhibit delay period activity (e.g., Serences et al., 2009). Although compelling, these studies do not preclude the possibility that subpopulations of voxels within their ROIs exhibit robust delay period activity and disproportionately encode WM information. In addition, studies examining population coding of sensory features have observed that information about sensory features is maximal during temporally varying patterns of activation rather than periods of stable population activity (Mazor & Laurent, 2005).

Taken together, the evidence outlined above necessitates a reevaluation of the canonical model of WM. The goal of this study was to critically evaluate key elements of this model. We analyzed fMRI data from 49 healthy

adult participants who performed a delayed recognition task requiring WM for faces, scenes, or both faces and scenes, depending on task instructions. First, we investigated the respective roles of IPFC and visual cortex during WM by directly comparing the nature of the information encoded by these two regions. Next, we systematically examined the degree to which sensory representations of WM stimuli are dependent on activity in neural populations that are highly selective for the maintained items. Finally, we tested the relationship between information storage and stable elevated delay period activity and characterized the temporal properties of WM information storage.

METHODS

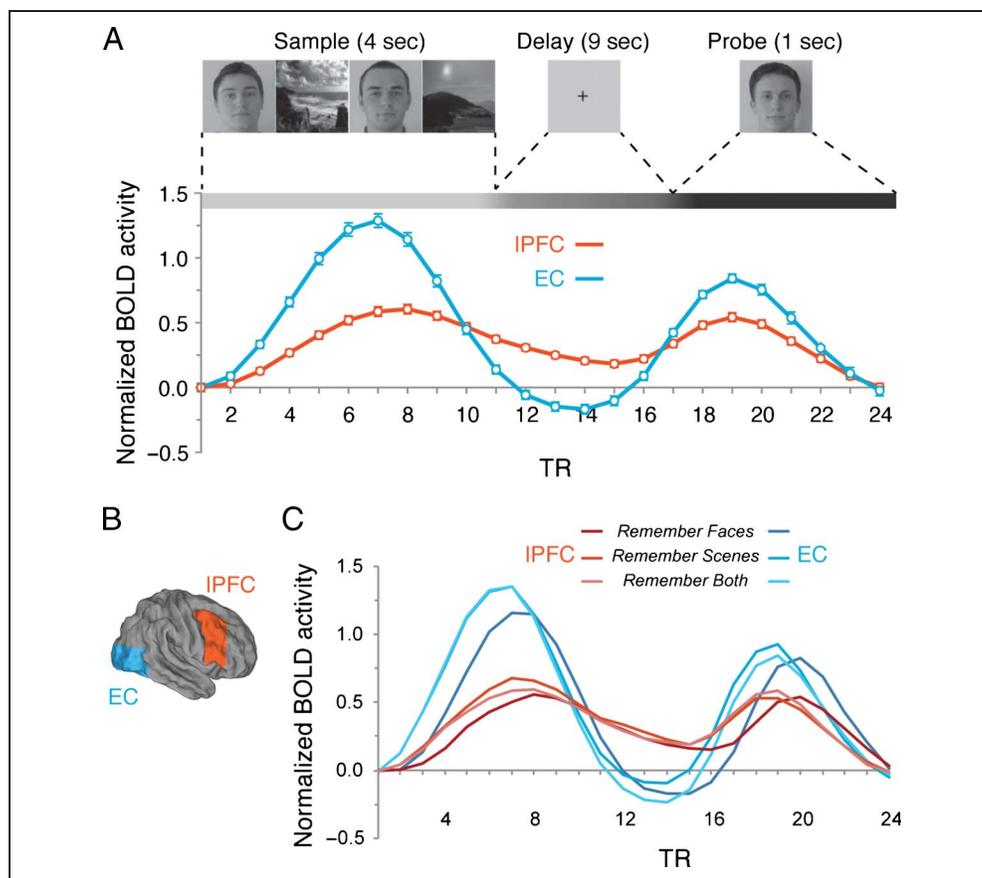
Participants

Data from 49 healthy adult participants, 18–32 years old (mean = 22.6 years; 20 women), were included in this analysis. All participants were right-handed with normal or corrected-to-normal vision and were not taking any medications with psychoactive, cardiovascular, or homeostatic effects. Written informed consent was obtained from all participants according to procedures approved by the University of California, Berkeley Committee for Protection of Human Subjects. Analyses of portions of this data set have been published elsewhere (Cohen, Sreenivasan, & D'Esposito, 2012; Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005).

Behavioral Task

A sample trial of the WM task is depicted in Figure 1A. Participants viewed four sequentially presented sample images (two faces and two scenes). Each sample image was presented for 800 msec with a 200-msec ISI. Participants' task varied according to instructions presented at the beginning of each scanning run. On Remember Faces trials, participants were instructed to remember the two faces and ignore the two scenes; on Remember Scenes trials, participants were instructed to remember the two scenes and ignore the two faces; on Remember Both trials, participants were instructed to remember all four sample images. Participants maintained the relevant sample images in WM over a 9-sec blank delay period. Following the delay period, a single probe image was presented for 1 sec, and participants responded indicating whether the probe image matched one of the relevant sample images (50% probability). The probe image was always a face on Remember Faces trials and was always a scene on Remember Scenes trials. The probe image could be either a face or scene on Remember Both trials. Data from a perceptual control condition that did not require WM were not included in the analyses described here. Trials were separated by a 10-sec intertrial interval. Each scanning run consisted of 10 trials of a single condition. There were

Figure 1. Behavioral task, BOLD time courses, and anatomical ROIs. (A) Top: On each trial, participants were presented with two faces and two scenes and instructed to remember the relevant sample items (faces, scenes, or both faces and scenes). The sample images were followed by a blank delay period, after which participants indicated whether the probe matched one of the relevant sample items. Bottom: Event-related BOLD time series were extracted from each of the ROIs, normalized, and averaged across participants. All error bars are *SEM*. The horizontal grayscale bar indicates the phase of the trial corresponding to BOLD and decoding measures, adjusted for the convolution with the HRF. (B) Analyses focused on a priori anatomical ROIs: IPFC (orange) and EC (blue). ROIs were bilateral; however, only the right hemisphere is shown here. (C) BOLD time courses separated by task condition in IPFC and EC ROIs. Error bars have been omitted for exposition.



three runs for each condition presented over the course of the experiment.

fMRI Data Acquisition and Preprocessing

Imaging data were collected with a 4T Varian (Palo Alto, CA) INOVA scanner equipped with a transverse electromagnetic send-and-receive radio-frequency head coil. Functional data were acquired with a two-shot T2*-weighted echoplanar imaging sequence (18 slices, slice thickness = 5 mm, repetition time [TR] = 2000 msec, echo time [TE] = 28 msec, matrix 64×64 , field of view = 224 mm). Slice-time correction was applied off-line using sinc interpolation. Each shot of half k -space was combined with the bilinear interpolation of the two flanking shots to result in an interpolated TR of 1000 msec. To register functional data to brain anatomy, a T1-weighted gradient-echo multislice anatomical scan with the same slice prescription as the functional images (TR = 200 msec, TE = 5 msec, matrix = 256×256 , field of view = 224 mm) and a high-resolution anatomical 3-D MP-FLASH (TR = 9 msec, TE = 5 msec, matrix = $256 \times 256 \times 128$, field of view = $224 \times 224 \times 198$ mm) were additionally acquired. Functional and anatomical data were preprocessed using FSL 4.1 (FMRIB's Software Library; www.fmrib.ox.ac.uk/

fsf); the MCFLIRT module was used for motion correction, and BET was used to skull-strip the data. All analyses were conducted in individual participant space on unsmoothed data.

ROIs

Anatomical IPFC and extrastriate visual cortex (EC) ROIs are shown in Figure 1B. ROIs were defined on a standard brain (MNI152) and transformed to individual participant space using FSL's FLIRT module for linear registration. The parameters to register the gradient-echo multislice anatomical image to the high-resolution MP-FLASH anatomical image (7 degrees of freedom) and the parameters to register the MP-FLASH to standard MNI152 space (12 degrees of freedom) were combined and inverted to provide the transformation from MNI space to individual participant space. IPFC (mean size = 1680 voxels, *SEM* = 35 voxels) was defined by combining the unthresholded templates of bilateral middle frontal gyrus and bilateral inferior frontal gyrus from the Harvard-Oxford Probabilistic Brain Atlas (FSL; provided by the Harvard Center for Morphometric Analysis). The boundaries of the bilateral EC ROI (mean = 1679; *SEM* = 36) were determined anatomically on the standard template brain and

included the lingual gyrus, the parahippocampal gyrus, posterior portions of the fusiform and inferior temporal gyri extending rostrally to the mid-fusiform gyrus to include the typical location of the fusiform face area (Kanwisher, McDermott, & Chun, 1997), and the surrounding occipital cortex.

Univariate fMRI Analysis

Although our primary analyses involved multivariate decoding methods, we used a traditional univariate general linear model (GLM) to identify canonical delay period activity. To visualize the time course of the BOLD data, individual trial time series were extracted from each ROI, z -scored, and averaged across trials, with the first TR of each trial serving as a baseline (Figure 1A, bottom; see Figure 1C for time courses separated by task condition). It should be noted that z -scored time courses are only presented for visualization purposes; all analyses of delay period activity magnitude were conducted on parameter estimates of the GLM (below). Parameter estimation for events of interest was conducted in AFNI (Cox, 1996). Our model included regressors for sample, delay, and probe events for each task condition (nine events of interest; correct trials only). Sample and probe events were modeled as 4-sec and 1-sec boxcar functions located at sample and probe stimulus onset, respectively. Delay events were modeled as a 1-sec boxcar function located in the middle of the delay period. Regressors for each event type were created by convolving the boxcars with a canonical gamma hemodynamic response function (HRF). Previous analyses have demonstrated that this method of temporally segregating regressors by at least 4 sec results in sufficiently low autocorrelation between events and can therefore produce independent parameter estimates for each regressor (Zarahn et al., 1999; Zarahn, Aguirre, & D'Esposito, 1997). This approach has successfully been used to isolate sample-evoked activity from delay-related activity (Cohen et al., 2012; Yoon, Curtis, & D'Esposito, 2006; Jha, Fabian, & Aguirre, 2004; Pessoa et al., 2002). Nuisance regressors included estimated motion parameters; sample, delay, and probe events for incorrect or missed trials; and the first and second derivatives of the gamma HRF to account for differences in the latency and dispersion of the peak BOLD response.

One of our analyses examined whether delay period activity magnitude was related to decoding evidence for the storage of WM information. To formally investigate this relationship, we divided each anatomical ROI into tertiles based on the magnitude of delay period activity in each voxel. The magnitude of delay period activity in a given voxel was determined by the t value of the delay period parameter estimate from the GLM collapsed across the three conditions, and voxels were assigned

to the top, middle, or bottom delay period tertile ROI according to this value.

Another analysis investigated the degree to which WM information was encoded by category-selective voxels. This required first defining the face and scene selectivity of voxels within an ROI and then removing voxels from the decoding analysis according to their selectivity. Voxels were ranked according to their preference for faces or scenes by analyzing localizer data from an independent scanning run. In this run, 16-sec blocks of rapidly presented face and scene stimuli were interspersed with blank 16-sec blocks, and participants were instructed to indicate stimulus repetitions with a button press. Data acquisition, preprocessing, and model (GLM) parameters were as described above, except that face, scene, and baseline events were modeled as 16-sec boxcar functions convolved with the canonical HRF. Parameter estimates for the face > scene and scene > face contrasts were used to determine the degree of voxels' preference for faces or scenes. The top v percentile of voxels consisted of the top $v/2$ percentile of face- and scene-preferring voxels.

Decoding Analyses

All decoding analyses were carried out using the Princeton MVPA toolbox (www.csmb.princeton.edu/mvpa/) and custom scripts implemented in MATLAB (The MathWorks, Inc., Natick, MA). Before decoding, BOLD data from each voxel were detrended by scanning run, separated into individual trial epochs, and temporally z -scored. No explicit feature selection was implemented beyond the masking of data with anatomical ROIs. We analyzed equivalent numbers of trials across task conditions for each participant. Decoding analysis was implemented using a logistic regression classifier. Training data labeled by task condition (Remember Faces, Remember Scenes, Remember Both) were entered into the classifier, which constructs a model that can discriminate between conditions given the multivoxel patterns of activation as an input. The classifier was then tested on unlabeled test data. Above-chance (>33% accuracy) ability to predict the condition indicates that the multivoxel patterns of activity contained information that discriminated between conditions. Successful decoding during the blank delay period would then indicate that information about the WM items persisted despite the lack of visual input and would be positive evidence for stored WM representations.

Most of our decoding analyses employed a leave-one-trial-out cross-validation scheme: The classifier was trained on data from all but one trial and tested on the remaining trial on each cross-validation fold. This procedure was repeated until each trial in turn served as the testing trial (Pereira, Mitchell, & Botvinick, 2009). Each cross-validation fold resulted in the assignment of a weight value to each voxel in the ROI for each of the three task conditions, indicating the degree to which the activity

within that voxel contributed to the classifier's output for that condition. During testing of the classifier, the vector of voxel BOLD activity was multiplied by the vector of voxels weights for each condition, resulting in a single activation value for each of our three conditions for each cross-validation fold. The testing trial was assigned a classifier guess in a winner-take-all manner. Accuracies of classifier guesses were averaged over cross-validation folds, resulting in a decoding accuracy. We set the ridge penalty (lambda value) for the logistic regression classifier to 0.01. Other penalty values yielded highly similar decoding accuracies.

To examine whether WM information persisted across the trial, we used a temporally resolved decoding approach. This involved creating a classifier for each of the 24 sample points (TRs) in the trial and testing each classifier only on data from the corresponding TR in other trials. The classifier was never trained and tested on data from the same trial. Thus, each training data point was separated from the closest testing data point by 23 TRs. As our focus was on identifying storage-related neural activity, statistical analyses focused on the epoch corresponding to the delay period, which, accounting for the hemodynamic lag of $\sim 4\text{--}6$ sec, was determined to be TRs 11–16 of each trial. This (relatively conservative) range was chosen to minimize the influence of sample- or probe-related activity on classifier estimates; however, results were consistent across less conservative ranges. For all statistical comparisons, the relevant measure was averaged over the six delay TRs. Statistical significance of decoding accuracies was assessed with a one-sample *t* test, with 33% accuracy as chance-level decoding. All comparisons were two-tailed.

One of our objectives was to investigate the nature of information encoded within IPFC and EC ROIs. We reasoned that sensory representations of more similar categories would be encoded in activity patterns that were more similar; thus, for example, patterns encoding sensory representations of faces should be more similar to patterns encoding both faces and scenes than they should be to patterns encoding scenes alone. To examine the similarity of patterns of activity in our task conditions, we examined misclassification rates (Chen et al., 2012; Kriegeskorte, 2008) for the Remember Faces and Remember Scenes conditions. We divided trials on which the classifier had incorrectly guessed the task condition into trials on which the classifier incorrectly guessed Remember Both and trials on which the classifier incorrectly guessed the opposite perceptual category (i.e., when the classifier guessed Remember Faces for a Remember Scenes trial, or when it guessed Remember Scenes for a Remember Faces trial). The proportion of incorrect classifier guesses for Remember Both and the opposite perceptual category were combined across Remember Faces and Remember Scenes conditions. These proportions were entered into a two-way repeated-measures ANOVA with factors of ROI (IPFC vs. EC) and

Classifier Guess (guess Remember Both vs. guess opposite perceptual category).

A separate classification procedure was used to examine the temporal stability of WM population coding. Unlike the previous procedure, which involved constructing a classifier for each TR that was only tested on data from the corresponding TR in other trials, this procedure involved constructing a classifier for each TR and testing each classifier on data from each TR in turn. This temporal cross-generalization procedure (Stokes et al., 2013; Meyers et al., 2008) enabled us to determine whether patterns of activity that encoded WM information at one point during the trial encoded WM information at other points in the trial as well. Temporal cross-generalization precluded the use of a leave-one-trial-out cross-validation approach, since TR 24 of trial $n - 1$ and TR 1 of trial n would be temporally contiguous, in violation of the rule that training and testing data should be independent to avoid biasing the classifier. Instead, we divided the data set into six groups, each of which contained data from each trial type. The classifier was trained on five groups and tested on the sixth using a leave-one-group-out cross-validation procedure, thus ensuring that training and testing data sets were independent. Lambda was set to 100 for this analysis.

RESULTS

Decoding WM Category Information

BOLD data from Remember Faces, Remember Scenes, and Remember Both trials were entered into a logistic regression classifier, which was trained on data labeled with the relevant WM stimulus category for each trial and tested on its ability to distinguish the relevant WM stimulus category in independent, unlabeled data. The logic behind this approach is that if a region represents WM stimulus information, then our classifier should be able to distinguish between task conditions at an above-chance level. We applied the decoding analysis independently to each of the 24 TRs that constituted the data acquired within a trial to examine whether evidence for WM information persisted over the course of the trial. Above-chance decoding accuracy corresponding to the delay period of the trial, when no visual information was present and WM maintenance was ongoing, was taken as evidence for the storage of WM information. Our analyses were restricted to two ROIs, IPFC and EC (Figure 1B), that have been implicated in the storage of visual WM information (Lepsien & Nobre, 2007; Ranganath, Cohen, Dam, & D'Esposito, 2004; Pessoa et al., 2002; Sakai, Rowe, & Passingham, 2002; Petrides, 2000; Zarahn et al., 1999; Fuster, Bauer, & Jervey, 1985). The decoding analysis demonstrated robust, above-chance accuracy across the trial (Figure 2A, left) and in particular during the delay phase of the trial in both EC and IPFC ROIs, $t(48) > 7.3$,

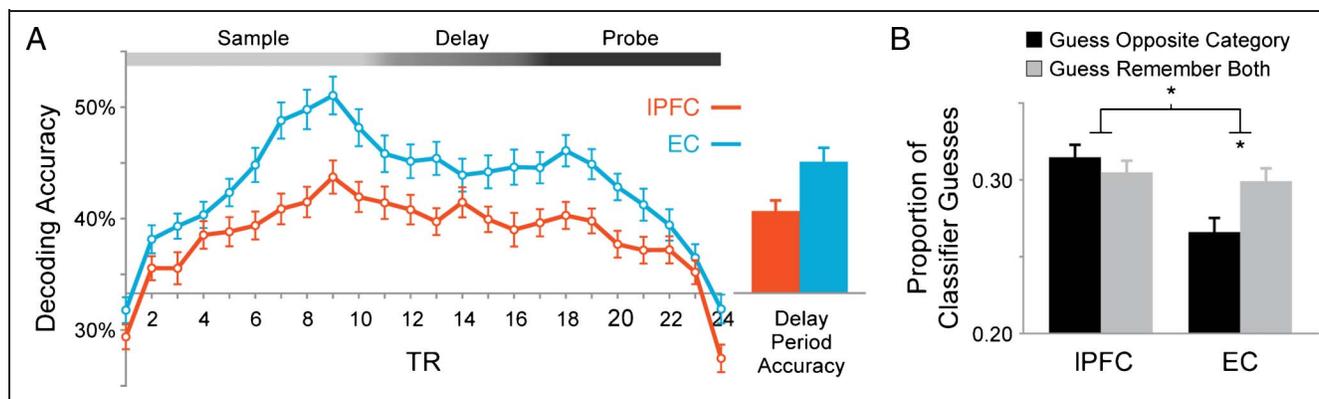


Figure 2. Decoding WM information from IPFC and EC ROIs. (A) Left: Decoding was above chance (33% accuracy) across all three epochs of the trial in both ROIs. Right: To isolate information about WM items during the maintenance phase of the trial, we examined decoding accuracy collapsed across the six TRs corresponding to the delay epoch of the trial. Accuracy was significantly above chance during the delay period in both ROIs, suggesting that information about the WM stimuli was maintained in both ROIs. (B) To distinguish between the storage of a sensory representation versus a nonsensory representation, we examined the misclassification of Remember Faces and Remember Scenes trials during the delay period. The classifier was disproportionately more likely to incorrectly guess Remember Both than the opposite category (i.e., guess Remember Faces on Remember Scenes trials and vice versa) in EC, consistent with a sensory representation, but not in IPFC. $*p < .0001$.

$ps < .0001$, Cohen's $d > 1.05$ (Figure 2A, right), indicating that category representations were maintained in both regions.

The Nature of WM Information in EC and IPFC

One of our primary goals was to distinguish WM representations that were sensory in nature, as would be expected if a region participates in WM storage, from nonsensory representations such as rules, goals, or abstract representations of categories. To do so, we examined the classifier's misclassification rates, which can provide insight into the representational similarity of our categories of interest (Chen et al., 2012; Kriegeskorte, 2008). We reasoned that if a region supports a sensory representation of WM stimuli, then Remember Faces trials should be incorrectly classified as Remember Both trials more often than they should be misclassified as Remember Scenes trials, because the sensory representation of faces is more similar to the representation of faces and scenes than it is to scenes alone. Similarly, Remember Scenes trials should also be disproportionately misclassified as Remember Both trials if activity patterns encode sensory representations. This approach was motivated by previous work demonstrating that visual neurons respond based on visual similarity to their preferred feature whereas IPFC neurons can encode arbitrary and abstract category boundaries independent of visual similarity (Freedman, Riesenhuber, Poggio, & Miller, 2001, 2003). Thus, our prediction was that misclassification rates in EC would be consistent with a sensory representation whereas misclassification rates in IPFC would not distinguish between visually similar categories. We compared the pattern of misclassification in our two ROIs during the delay period by performing a two-way ANOVA on the proportion of misclassified trials with

the factor of ROI and Guess Type (guess Remember Both and guess opposite perceptual category—that is, face guess on scene trials and vice versa). We found a significant ROI \times Guess Type interaction, $F(1, 48) = 10.49$, $p = .002$; $\eta_p^2 = 0.18$ (Figure 2B): A greater proportion of Remember Faces and Remember Scenes trials were misclassified as Remember Both in EC, $t(48) = 3.2$, $p = .003$; $d = 0.45$, whereas there was no significant difference in the proportion of trials misclassified as Remember Both versus the opposite perceptual category in IPFC, suggesting that EC and not IPFC stores a sensory representation of WM items.

Contribution of Selective Neural Populations to WM Information Storage

To investigate whether sensory WM representations were encoded by category-selective populations within EC, we ranked EC voxels according to their category selectivity and removed increasing numbers of voxels from the decoding analysis to determine the degree to which decoding was dependent on category-selective voxels. Similar procedures have previously been used to determine whether representations of object categories depend on selective voxels during perception (Haxby et al., 2001) and attention (Chen et al., 2012). Face and scene selectivity of voxels were determined in each participant in an independent scanning run (see Methods). Figure 3A shows the top 25% of selective EC voxels in two representative participants. Note that these voxels correspond well to previously described face- and scene-dedicated processing modules in EC (Gauthier et al., 2000; Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998; Kanwisher et al., 1997). After identifying these voxels, we repeated the decoding analysis as described above after removing a percentile of the

most selective voxels from the analysis. The analysis was conducted removing 5%, 25%, and 50% of the most category-selective voxels from EC. Although decoding accuracy was reduced as an increasing proportion of category-selective EC voxels were removed from the analysis (Figure 3B, left), decoding accuracy during the delay period remained significantly above chance, even when half of the voxels in EC were removed, $t_s(48) > 7.8$, $p_s < .0001$; $d_s > 1.1$ (Figure 3B, right). From this, we concluded that, although category-selective EC voxels may code for WM information, WM storage recruits distributed EC populations with a broad range of category selectivity.

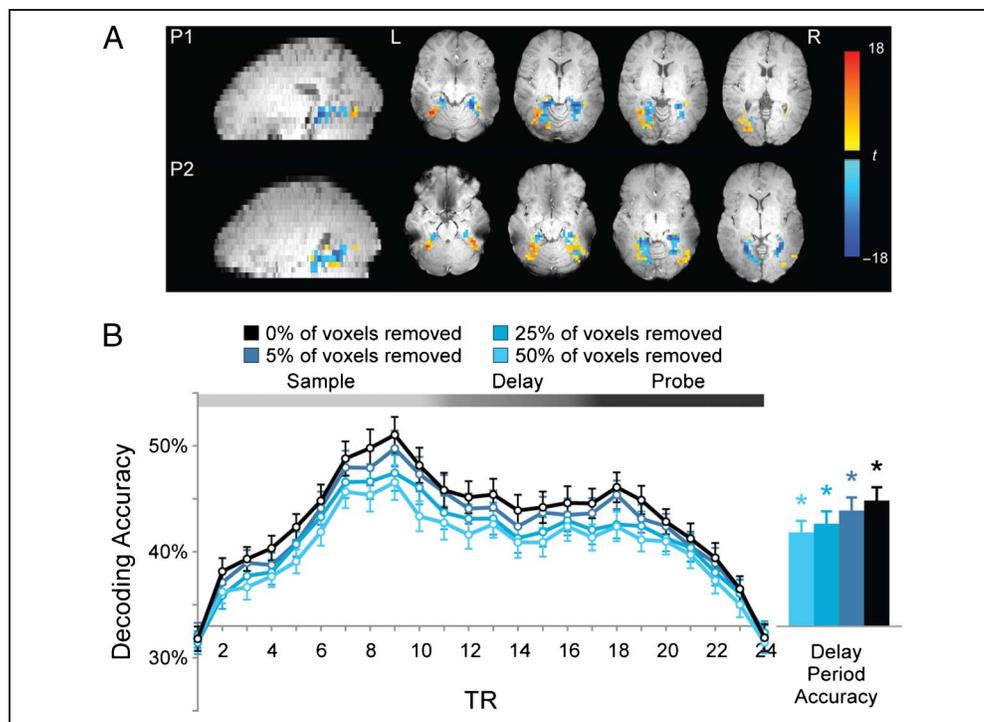
Delay Period Activity and WM Information Storage

To understand the role of delay period activity in WM storage, we investigated the relationship between our decoding metrics and the magnitude of delay period activity in LPFC and EC ROIs. Individual voxels within each ROI were assigned to strata according to the magnitude of delay period activity as determined by the delay period parameter estimates of our univariate model (see Methods). We created three strata within each ROI, with the top tertile demonstrating robust delay period activity, the middle tertile showing an absence of delay period activity, and the bottom tertile demonstrating below-baseline levels of activity during the delay (Figure 4A). If delay period activity is related to WM information storage, then the top tertile should demonstrate greater evidence for WM information storage, as evinced by higher decoding accuracy during the delay period. Decoding analyses performed sep-

arately in each tertile ROI showed that decoding accuracy was consistent across tertiles (Figure 4B, left) and did not differ significantly during the delay in either ROI ($F_s < 0.67$, $p_s > .5$, $\eta_p^2 < 0.02$; Figure 4B, right).

In a complementary analysis, we examined the relationship between the magnitude of delay period activity in a voxel and the degree to which that voxel was considered informative by the classifier during the delay period. We extracted the delay period classifier weights (see Methods) for each of the three conditions (Remember Faces, Remember Scenes, Remember Both) from our original decoding analysis. To arrive at a single delay period weight value per voxel per condition, weights were averaged over cross-validation folds and then over the six delay TRs. Both positive and negative weight values can indicate that a voxel is highly informative to the classifier; we therefore examined the correlation between the absolute magnitude of the weights and the univariate model's estimate of delay period activity in the same condition. This yielded three correlation values per ROI, which were each averaged across participants. If the magnitude of delay period activity in a voxel is an indication of the degree to which it was informative to our decoding analysis, a positive correlation should be expected. Consistent with the analysis above, the correlation coefficients were between -0.01 and 0 , indicating no relationship between a voxel's contribution to the classifier and its delay period activity magnitude. Results were qualitatively similar when using the raw weight values. Together, these analyses present a formal dissociation between the magnitude of delay period activity and WM storage.

Figure 3. Relationship between category selectivity and WM storage in EC. (A) An independent scanning run was used to identify category-selective voxels. The top 25% of EC voxels ranked by category selectivity are shown here for two representative participants. The contrast shown for these voxels is faces > scenes; warm colors indicate voxels selective for faces, whereas cool colors indicate voxels selective for scenes. (B) The decoding analysis was repeated as increasing percentages of voxels were removed from the ROI based on their degree of category selectivity. Accuracy remained well above chance despite the removal of up to 50% of EC voxels (decoding accuracy at each time point shown on the left; accuracy collapsed over the delay period shown on the right).



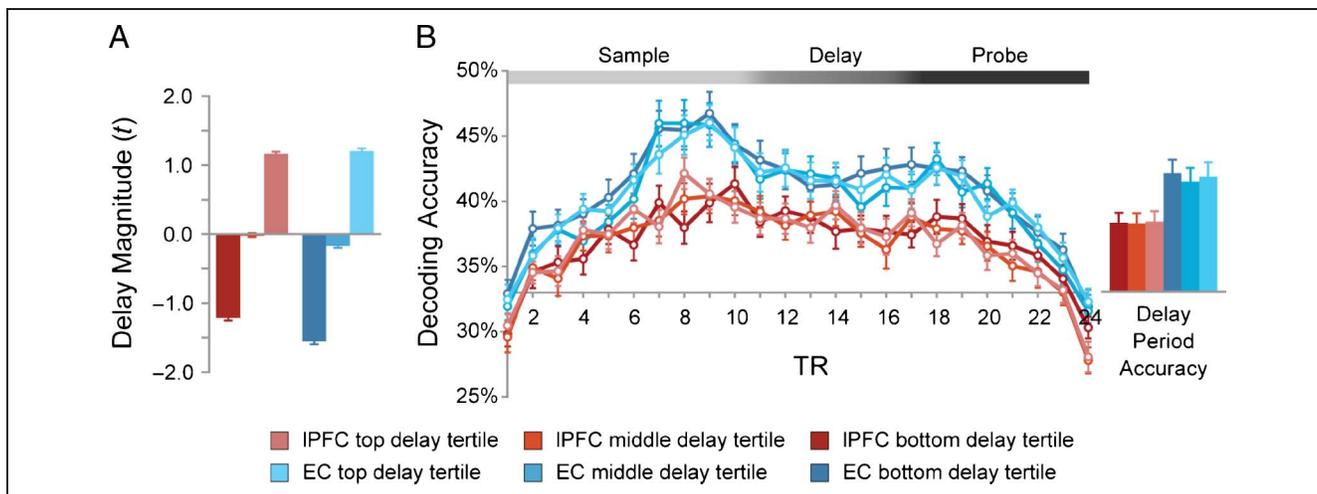


Figure 4. Correspondence between delay period activity and WM storage. (A) Anatomical IPFC and EC ROIs were divided into tertiles based on the magnitude of delay period activity. The top tertile in each ROI exhibits delay period activity that is well above baseline, whereas the bottom tertiles show below-baseline levels of delay period activity. (B) Decoding was carried out separately for each tertile. Accuracy did not differ as a function of delay period magnitude in either ROI, indicating a dissociation between delay period activity magnitude and WM storage. The graph on the right depicts decoding accuracy during the delay period for both ROIs (both one-way ANOVAs with factor of tertile were nonsignificant).

Temporal Stability of WM Information Storage

The above analyses dissociate the magnitude of delay period activity and WM storage. A separate but related question is whether sustained WM representations rely on stable multivoxel patterns of activity. Patterns of voxels

with a wide range of activation levels could stably encode a stimulus independent of their delay period activity magnitude. Our previous decoding analyses revealed WM category information in EC that persisted throughout the trial; however, they did not distinguish whether this information was encoded via patterns of activity that were

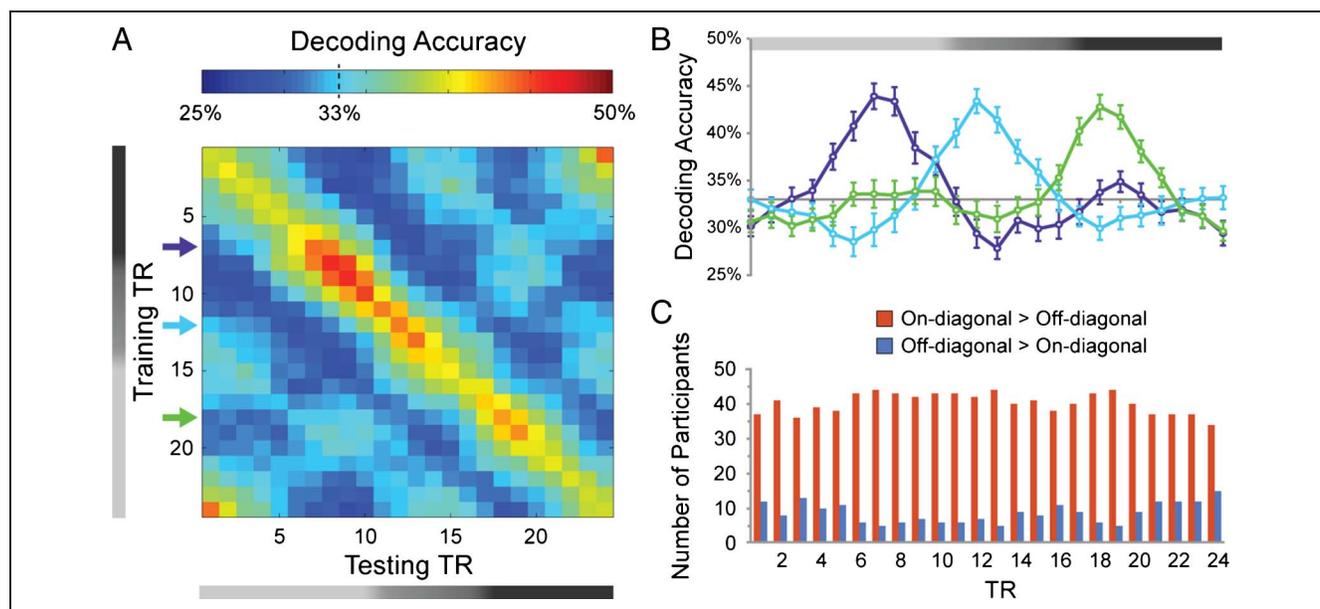


Figure 5. Temporal dynamics of WM storage. (A) Temporal cross-generalization analysis involved training and testing a classifier at each of the 24 TRs that comprise a trial, resulting in a 24×24 matrix of decoding accuracies. Note that the diagonal of the matrix is not equivalent to the plot in Figure 2A because of slightly different decoding procedures employed in the two analyses (see Methods). (B) Decoding accuracy is shown for three training TRs (TR 7, TR 12, and TR 18) indicated by the arrows in A. (C) The number of participants for whom training and testing the classifier on data from the same parts of the trial (the on-diagonal elements of the matrix in A) yielded higher classification accuracy than when data from different parts of the trial (the off-diagonal elements of the matrix) is greater than chance (binomial test; $p < .005$) indicating that patterns coding for WM representations shift across encoding, maintenance, and response.

stable throughout the trial or whether storage was carried out by patterns of activity that shifted over the course of a trial. To investigate this question, the decoding analysis was modified to train the classifier on data from one TR and test on each of the 24 TRs in turn. This procedure was repeated such that each TR served as the training TR for one iteration of testing, resulting in a 24×24 matrix of decoding accuracy. If information is stored in a static or stable pattern, then a classifier trained on one TR should successfully be able to decode information on nonadjacent TRs within the trial. Instead, if information is stored dynamically in temporally varying patterns of activity, then a classifier created from data from one TR should not be able to successfully decode information about the relevant stimulus category from another part of the trial (Stokes et al., 2013; Meyers et al., 2008). As our interest was in the temporal properties of sensory representations, our analysis focused on data from the EC ROI. Decoding accuracy was above chance along the diagonal of the matrix, when the classifier was trained and tested on data from the same part of the trial, but was reduced when the classifier was trained and tested on data from different TRs (Figure 5A and B). To formally test whether patterns were stable throughout the trial, we framed our question in terms of model selection. For each training TR, our measure of interest was the difference between the decoding accuracy from the model tested on data from the same TR (the on-diagonal element of a given row of the decoding accuracy matrix) and the average decoding accuracy from the other 23 models (the average of the off-diagonal elements of the same row of the decoding accuracy matrix). If the on-diagonal element outperformed the average of the off-diagonal elements, we took this as evidence that the pattern of information on the training TR was not sustained across the trial. We then compared the proportion of participants for whom the on-diagonal model outperformed the average of the off-diagonal models against the binomial distribution $B(49, 0.5)$ for each TR. The diagonal model significantly outperformed the average off-diagonal model at all 24 TRs (Figure 5C, all p s < .005). Critically, the use of cross-validation to evaluate our models on independent sets of data precluded the possibility that our results were the result of a single stable pattern plus noise and allowed us to conclude that patterns containing WM information shifted over the course of the trial.

DISCUSSION

Our results demonstrate that (i) EC retains sensory WM representations whereas LPFC encodes category representations that are nonsensory in nature, (ii) WM information is stored in patterns of activity that are distributed over voxels with a broad range of selectivity, (iii) WM storage is independent of the magnitude of population delay period activity, and (iv) patterns of activity encoding WM representations vary over the course of main-

tenance. Along with other work describing features of WM that are incompatible with the canonical view of WM, our findings highlight the need for a reevaluation of the neural instantiation of WM. These findings also emphasize the utility of multivariate decoding analyses of fMRI data in the study of WM.

Contrasting the Roles of EC and LPFC in WM

Our results show that EC retains sensory WM representations whereas LPFC retains nonsensory information. These findings are consistent with growing evidence that visual WM representations are stored in visual cortex (Ester, Anderson, Serences, & Awh, 2013; Christophel et al., 2012; Riggall & Postle, 2012; Silvanto & Cattaneo, 2010; Harrison & Tong, 2009; Serences et al., 2009) as well as studies highlighting the role of LPFC in forming and maintaining categorical representations and representations of important task variables (Rigotti et al., 2013; Meyers et al., 2008; Freedman et al., 2001, 2003). The key advance of the present work is that we were able to contrast the nature of the information stored in LPFC and EC within the same task, thus clarifying the respective roles of these two regions. Critically, our results provide a potential alternative explanation for previous work indicating that sensory representations are stored in LPFC; patterns of activity associated with specific stimuli in previous work may encode categorical or rule information associated with that stimulus and not the sensory properties themselves. It is important to note that our conclusions do not rely on a comparison of decoding accuracy across regions, which could yield spurious differences arising from vascular or other differences across ROIs that might obscure informative patterns of activity. Instead, we used misclassification rates to distinguish between the nature of patterns in two ROIs that demonstrated successful decoding, allowing us to conclude that EC stores sensory information about WM items.

How do we explain discrepancies between our findings and other work that was unable to decode WM information in LPFC (Christophel et al., 2012; Riggall & Postle, 2012)? Previous fMRI studies that were unable to decode WM information from LPFC were decoding stimulus identity (e.g., one of several directions of motion), whereas our study decoded stimulus category while participants maintained stimulus identity in WM. Although studies decoding stimulus category have the disadvantage of not being able to identify stimulus-specific patterns of activity, given LPFC's preference for category boundaries (i.e., learned abstract distinctions) over item similarity (i.e., sensory features; Freedman et al., 2003), it is possible that the nature of our task facilitated decoding in LPFC. In line with this notion, a recent fMRI decoding study found information about maintained visual items in visual cortex and information about maintained visual categories in LPFC (Lee et al., 2013). Although the authors interpret this dissociation as a distinction between visual and verbal WM, in light of the present results, we suggest that these findings

can be interpreted as a distinction between sensory and categorical representations. Although the present work focuses on these regions in isolation, WM likely requires coordinated activity between these regions and others, including parietal cortex and BG. Further study is required to understand the individual and collective function of these regions.

Delay Period Activity and WM Storage

Previous decoding analyses have demonstrated successful decoding of the contents of WM in the absence of delay period activity (Linden et al., 2011; Serences et al., 2009); however, these studies did not rule out the possibility that subpopulations of voxels within their ROIs may have exhibited delay period activity and contributed disproportionately to their decoding success. One study removed all voxels with significant delay period activity and still observed information about WM items (Riggall & Postle, 2012); however, these results do not exclude the possibility that voxels with greater magnitude delay period activity may contribute more information to a classifier. Previous work also used arbitrary significance thresholds to define delay active voxels, which may obscure the contributions of just below threshold activity. By dividing our ROIs into strata based on the magnitude of delay period activity in each voxel, we were able to demonstrate a more convincing dissociation between patterns coding for WM storage and the magnitude of delay period activity. This dissociation was strengthened by our finding that delay magnitude and voxel weights were not positively correlated. In concert with evidence that delay period activity is associated with cognitive operations besides WM (Curtis & Lee, 2010; Meyer, Qi, & Constantinidis, 2007) and successful WM in the absence of delay period activity (Offen, Schluppeck, & Heeger, 2009; Serences et al., 2009), our work suggests an independence between delay period activity and WM storage. How might WM representations be sustained without relying on delay period activity? One possibility is suggested by work showing that information can be sustained over brief intervals via rapid shifts in synaptic weights (Mongillo, Barak, & Tsodyks, 2008; Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008). In such a scenario, neurons that store memory traces serve as matched filters, and stimulus- or category-specific delay activity may be a function of nonspecific input into the system rather than an index of storage per se.

What, then, is the function of elevated sustained delay activity frequently observed during WM? Although our analyses suggest that delay period activity and WM storage are not synonymous, persistent neural throughout WM maintenance is related to WM performance (Cohen et al., 2012; Pessoa et al., 2002) and is thus an important element of WM. The strong association between delay period activity and regions of PFC that carry out complex operations, such as the temporal integration of behav-

iorally relevant goals (Fuster, 2001), suggests that one possible function of delay period activity in LPFC may be to sustain higher-order task and goal representations (Miller & Cohen, 2001). We suggest that stable delay period activity may be one of several possible neural mechanisms for retaining information in an active state.

An important consideration in evaluating these findings is the degree of accordance between data from single-unit recordings in nonhuman primates and multivariate analyses of human fMRI data. The former combines excellent temporal resolution with the ability to observe spiking activity in single neurons, whereas the latter has relatively coarse spatial resolution but has the advantage of broad spatial coverage to examine population codes across wide regions of cortex. Given that fMRI voxels represent the summed activity of hundreds of thousands of neurons as well as the uncertain relationship between neuronal spiking and BOLD activity (Cardoso, Sirotin, Lima, Glushenkova, & Das, 2012; Logothetis, 2008), we cannot rule out the possibility that significant stable delay period spiking activity exists even within voxels demonstrating low levels of sustained BOLD activity. Additionally, one must consider the differences in the tasks employed in human fMRI and monkey electrophysiological studies; the former employs delay periods lasting up to 20 sec, whereas the latter typically has delay periods shorter than 5 sec. Germane to this distinction, spiking models of delay period activity in PFC typically find that item-specific delay period activity decays after several seconds (Hansel & Mato, 2013). Ultimately, methods such as monkey fMRI and electrocortical recordings in humans may help reconcile some of the differences between findings in humans and monkeys, providing a more complete picture of WM.

Dynamic Patterns of Activation in WM

Although the other aspects of WM storage that we investigated are explicit elements of the canonical view, the temporal stability of WM representations is largely an implicit property of WM models. Experimental manipulations that disrupt delay period activity, such as the presentation of distracting items during the delay period, are often used to dissociate regions that participate in storage from regions that perform auxiliary roles in WM (Artchakov et al., 2009; Yoon et al., 2006; Miller, Li, & Desimone, 1991, 1993). The strong implication in this work is that WM representations must persist in a stable form across the period of maintenance. This is in contrast to evidence from psychology suggesting that WM representations undergo changes during the period of maintenance, as evidenced by different levels of susceptibility to intrusion (Oberauer, 2001), as well as evidence that stimulus features can be encoded via dynamic population codes during perception (Crowe, Averbach, & Chafee, 2010; Mazor & Laurent, 2005).

The temporal properties of the neural correlates of WM have not been well studied; however, extant electrophysiological evidence from rats (Baeg et al., 2003) and monkeys (Meyers et al., 2008) indicates that population coding of WM representations can involve spatially and temporally varying patterns of activity. These empirical findings are supported by theoretical work indicating that population dynamics can support the encoding of stable representations (Druckmann & Chklovskii, 2012). A recent noteworthy study used a similar temporal cross-generalization decoding approach to investigate WM representations in monkey IPFC and found that information was encoded in time-varying patterns of activity (Stokes et al., 2013). Interestingly, the authors observed that the patterns of activity coding for WM items were more stable during the delay period relative to other parts of the trial.

In contrast to Stokes and colleagues, we found that informative patterns of activity were not stable at any point during the trial. Additionally, our finding of dynamic population coding in EC is inconsistent with previous fMRI decoding results demonstrating that WM information is contained in stable patterns of visual cortical activity (Riggall & Postle, 2012; Harrison & Tong, 2009; Serences et al., 2009). An intriguing possibility is that these discrepancies may be explained by WM load. In our study, WM load varied between two and four items, whereas most previous studies did not tax WM load to the same degree. This possibility receives tentative support from recent findings by Emrich and colleagues, who performed a decoding analysis of fMRI data during the maintenance of multiple directions of motion (Emrich, Riggall, LaRocque, & Postle, 2013). Although their analyses did not explicitly focus on temporal cross-generalization, decoding of direction of motion did not appear to generalize across the entire trial, particularly when load was high. This finding is particularly striking when compared to previous results from the same group showing robust temporal generalization with a WM load of one (Riggall & Postle, 2012). Further work will be necessary to explicitly investigate the relationship between WM load and temporal dynamics of population coding.

Acknowledgments

The authors thank M. Todd for several helpful discussions about the details of the decoding analyses. This work was supported by NIH grant MH63901 to M. D.

Reprint requests should be sent to Kartik K. Sreenivasan, Division of Science and Mathematics, New York University Abu Dhabi, Room 401, 19 Washington Square North, New York, NY 10011, or via e-mail: kartik.sreenivasan@nyu.edu.

REFERENCES

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, *21*, 373–383.

- Artchakov, D., Tikhonravov, D., Ma, Y., Neuvonen, T., Linnankoski, I., & Carlson, S. (2009). Distracters impair and create working memory-related neuronal activity in the prefrontal cortex. *Cerebral Cortex*, *19*, 2680–2689.
- Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., & Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, *40*, 177–188.
- Cardoso, M. M. B., Sirotin, Y. B., Lima, B., Glushenkova, E., & Das, A. (2012). The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. *Nature Neuroscience*, *15*, 1298–1306.
- Chen, A. J.-W., Britton, M., Turner, G. R., Vytlačil, J., Thompson, T. W., & D'Esposito, M. (2012). Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. *Frontiers in Human Neuroscience*, *6*, 187.
- Christophel, T. B., Hebart, M. N., & Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, *32*, 12983–12989.
- Cohen, J. R., Sreenivasan, K. K., & D'Esposito, M. (2012). Correspondence between stimulus encoding- and maintenance-related neural processes underlies successful working memory. *Cerebral Cortex*. doi:10.1093/cercor/bhs339.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386*, 608–611.
- Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, *19*, 261–270.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Crowe, D. A., Averbach, B. B., & Chafee, M. V. (2010). Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *Journal of Neuroscience*, *30*, 11640–11653.
- Curtis, C. E., & Lee, D. (2010). Beyond working memory: The role of persistent activity in decision making. *Trends in Cognitive Sciences*, *14*, 216–222.
- Druckmann, S., & Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, *22*, 2095–2103.
- Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, *33*, 6516–6523.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*, 754–761.
- Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *Journal of Neuroscience*, *29*, 15258–15265.
- Ewbank, M. P., Schluppeck, D., & Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *Neuroimage*, *28*, 268–279.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science (New York, NY)*, *291*, 312–316.

- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*, 5235–5246.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*, 331–349.
- Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology*, *36*, 61–78.
- Fuster, J. M. (2001). The prefrontal cortex—An update: Time is of the essence. *Neuron*, *30*, 319–333.
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science (New York, NY)*, *173*, 652–654.
- Fuster, J. M., Bauer, R., & Jervey, J. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research*, *330*, 299–307.
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*, 495–504.
- Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience*, *17*, 507–517.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*, 477–485.
- Han, X., Berg, A. C., Oh, H., Samaras, D., & Leung, H.-C. (2013). Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions. *Neuroimage*, *73*, 8–15.
- Hansel, D., & Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *Journal of Neuroscience*, *33*, 133–149.
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, NY)*, *293*, 2425–2430.
- Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*, 523–534.
- Jha, A. P., Fabian, S. A., & Aguirre, G. K. (2004). The role of prefrontal cortex in resolving distractor interference. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 517–527.
- Jha, A. P., & McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: An event-related functional MRI study. *Journal of Cognitive Neuroscience*, *12*(Suppl. 2), 90–105.
- Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*, 544–552.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kriegeskorte, N. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi:10.3389/neuro.06.004.2008.
- Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology*, *34*, 337–347.
- Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*, 997–999.
- Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral Cortex*, *17*, 2072.
- Linden, D. E. J., Oosterhof, N. N., Klein, C., & Downing, P. E. (2012). Mapping brain activation and information during category-specific visual working memory. *Journal of Neurophysiology*, *107*, 628–639.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*, 869–878.
- Mazor, O., & Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, *48*, 661–673.
- Meyer, T., Qi, X. L., & Constantinidis, C. (2007). Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. *Cerebral Cortex*, *17*(Suppl. 1), i70–i76.
- Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*, 1407–1419.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Miller, E. K., Erickson, C., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*, 5154.
- Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science (New York, NY)*, *254*, 1377–1379.
- Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience*, *13*, 1460.
- Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science (New York, NY)*, *319*, 1543–1546.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.
- Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 948–957.
- Offen, S., Schluppeck, D., & Heeger, D. J. (2009). The role of early visual cortex in visual short-term memory and visual attention. *Vision Research*, *49*, 1352–1362.
- O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*, 580–590.
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, *45*(1 Suppl.), S199–S209.
- Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron*, *35*, 975–987.
- Petrides, M. (2000). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *Journal of Neuroscience*, *20*, 7496.
- Ranganath, C., Cohen, M., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience*, *24*, 3917.
- Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, *32*, 12990–12998.

- Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., et al. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*, 585–590.
- Sakai, K., Rowe, J. B., & Passingham, R. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*, 479–484.
- Schluppeck, D., Curtis, C. E., Glimcher, P. W., & Heeger, D. J. (2006). Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. *Journal of Neuroscience*, *26*, 5098–5108.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207.
- Silvanto, J., & Cattaneo, Z. (2010). Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *Neuroimage*, *50*, 1683–1689.
- Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (in press). Revisiting the role of persistent neural activity in working memory. *Trends in Cognitive Science*. doi:10.1016/j.tics.2013.12.001.
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*, 364–375.
- Sugase-Miyamoto, Y., Liu, Z., Wiener, M., Optican, L., & Richmond, B. (2008). Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Computational Biology*, *4*.
- Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science (New York, NY)*, *260*, 1955–1958.
- Xing, Y., Ledgeway, T., McGraw, P. V., & Schluppeck, D. (2013). Decoding working memory of stimulus contrast in early visual cortex. *Journal of Neuroscience*, *33*, 10301–10311.
- Yoon, J., Curtis, C. E., & D'Esposito, M. (2006). Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *Neuroimage*, *29*, 1117–1126.
- Zarahn, E., Aguirre, G., & D'Esposito, M. (1997). A trial-based experimental design for fMRI. *Neuroimage*, *6*, 122–138.
- Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1999). Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI. *Cognitive Brain Research*, *7*, 255–268.