OXFORD

Cerebral Cortex, 2016; 1–10

doi: 10.1093/cercor/bhw283
Original Article

ORIGINAL ARTICLE

Within-Category Decoding of Information in Different Attentional States in Short-Term Memory

Joshua J. LaRocque[1,2], Adam C. Riggall[3], Stephen M. Emrich[4], and Bradley R. Postle[2,3]

[1]Medical Scientist Training Program, University of Wisconsin-Madison, Madison, WI 53726, USA, [2]Department of Psychiatry, University of Wisconsin-Madison, Madison, WI 53706, USA, [3]Department of Psychology, University of Wisconsin-Madison, Madison, WI 53726, USA, and [4]Department of Psychology, Brock University, St. Catharines, L2S 3A1, Canada

Address correspondence to Joshua J. LaRocque, Department of Psychiatry, 1056 Wisconsin Psychiatric Institute, University of Wisconsin-Madison, 6001 Research Park Blvd, Madison, WI 53719, USA. Email: jjlarocque@gmail.com

Abstract

A long-standing assumption of cognitive neuroscience has been that working memory (WM) is accomplished by sustained, elevated neural activity. More recently, theories of WM have expanded this view by describing different attentional states in WM with differing activation levels. Several studies have used multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) data to study neural activity corresponding to these WM states. Intriguingly, no evidence was found for active neural representations for information held in WM outside the focus of attention ("unattended memory items," UMIs), suggesting that only attended memory items (AMIs) are accompanied by an active trace. However, these results depended on category-level decoding, which lacks sensitivity to neural representations of individual items. Therefore, we employed a WM task in which subjects remembered the directions of motion of two dot arrays, with a retrocue indicating which was relevant for an imminent memory probe (the AMI). This design allowed MVPA decoding of delay-period fMRI signal at the stimulus-item level, affording a more sensitive test of the neural representation of UMIs. Whereas evidence for the AMI was reliably high, evidence for the UMI dropped to baseline, consistent with the notion that different WM attentional states may have qualitatively different mechanisms of retention.

Key words: attention, multivariate pattern analysis, neural representations, short-term memory, working memory

Introduction

Working memory (WM), the ability to transiently remember information no longer present in the environment, is necessary for many everyday behaviors, including carrying on a conversation, following the plot of a film, and playing board games. According to several influential theoretical models (Cowan, 1988; McElree, 1998; Oberauer, 2002), WM can be understood as the attentional selection of representations that underlie perception and long-term memory (LTM). In these frameworks, information may be transiently retained in any of several distinct states that vary in their level of "activation," with the activation level determined by the allocation of attention. Each of these models distinguishes between a capacity-limited state, comprising highly accessible information selected by the focus of attention (FoA), and a state that is characterized by a lower, intermediate level of activation, comprising information that is in WM but not currently in the FoA. We term information in those two states "attended memory items" (AMIs) and "unattended memory items" (UMIs). Depending on the nature of the to-be-remembered stimuli, these state-based models can also be referred to as "activated long-term memory" or "sensorimotor recruitment" models (reviewed in LaRocque, Lewis-Peacock et al., 2014).

Recently, with the advent of multivariate pattern analysis (MVPA) of neuroimaging data, state-based models of WM have garnered considerable empirical support. With regard to "activated LTM," for example, MVPA classifiers trained on functional magnetic resonance imaging (fMRI) data from a retrieval-from-LTM task can successfully decode fMRI data from a WM task using the same stimuli (Lewis-Peacock and Postle, 2008). This could only be possible if the WM task entailed the attentional activation of the same neural representations that had been engaged by the LTM task. With regard to sensory recruitment, several studies have shown that MVPA classifiers trained on sensory-evoked signal can decode delay-period stimulus representations (Harrison and Tong, 2009; Serences et al. 2009; Riggall and Postle 2012). Note, however, that all of these studies address the retention of information that is in the FoA. Our interest in the present report is the physiological state of UMIs, representations that are in WM, but outside the FoA.

With conventional procedures for testing WM, the to-be-remembered item or items are of immediate behavioral relevance, and are therefore likely to be AMIs. Thus, the vast majority of studies of neural activity related to WM have, in effect, confounded the short-term retention of information (i.e., "WM storage") with attention. Consequently, many, if not most, of the results supporting an active-trace account of WM might reflect the FoA, rather than memory retention per se. Indeed, two recent studies from our laboratory have called into question the idea that UMIs are also supported by an active trace (Lewis-Peacock et al. 2012; LaRocque et al. 2013). In these experiments, each trial began with the presentation of two stimuli, each from a different category (the categories were pairs of line segments, pronounceable nonwords, and words). After an initial delay period, a retrocue signaled which of the two memory items would be the target of the first memory probe. Based on previous behavioral studies (Oberauer 2005), we assumed that, during the ensuing delay period, the cued item was an AMI, and the uncued item a UMI. Both items had to be retained during this portion of the trial, because following the first memory probe there was a second retrocue, indicating (with $P = 0.5$) which of the two would be the target of a second memory probe. Many behavioral studies have established the efficacy of this retrocuing technique (e.g., LaRocque et al. 2013; Oberauer 2005). MVPA of neural activity (measured with fMRI, Lewis-Peacock et al. 2012 and electroencephalography, EEG, LaRocque et al. 2013) revealed evidence for the active representation of the categories of both items upon their initial presentation, but only for the category of the AMI after the retrocue indicated which item would next be probed. These results suggest that an active neural representation might only be present for items in WM when they are potentially relevant for an impending behavioral response, and thus putatively held in the FoA. Provocatively, they also suggest that UMIs may not be maintained in the same manner as AMIs.

Although intriguing, the studies that have failed to find evidence for an active neural representation of UMIs (Lewis-Peacock et al. 2012; LaRocque et al. 2013) are subject to an important limitation, in that they decoded stimulus information at the level of the category of the stimuli, but not at the level of the stimulus itself. Indeed, in these studies, the three stimulus categories were constructed expressly to encourage subjects to employ one of three different encoding strategies: a visual sensory code for the line segment category; an auditory sensory code for the nonwords category; or a semantic code for the word category. It was reasoned that this would maximize the ability to identify distinct neural signatures with MVPA, and in this regard the design was successful. Of ultimate interest, however, is how subjects retain stimulus-specific information in order to respond correctly to the memory probe. Decoding category-level information limits the interpretability of the null finding with UMIs, due to its inherently lower sensitivity to within-category, item-specific features that must constitute the representation of any individual UMI. To illustrate this lower sensitivity, consider the evidence that holding an item in the FoA entails the broad attentional activation of attributes shared by many items belonging to the same category (e.g., Polyn et al. 2005), whereas the retention of a UMI, in principle, need only entail the markedly reduced activation of just the minimal number of features needed to represent that item. Under such a scenario, a classifier trained to generalize across items within a category, and to learn only their shared attributes—which was the case for Lewis-Peacock et al. (2012) and LaRocque et al. (2013)—would fail to detect activity representing features specific to a single item. To be concrete, for an item from the "line segments" category of the previous studies, such features could include the orientation of the stimuli—a category-level classifier could not be sensitive to different orientations, because it was trained to learn only features in common to all possible orientations, in particular those features that distinguished "line segments" from the other two categories. A more definitive assessment of the neural bases of the short-term retention of UMIs would draw all stimuli from the same category, and train multivariate pattern classifiers with greater sensitivity, such that they could discriminate activity corresponding to each individual stimulus from a memory set.

We designed the present study to test the hypothesis that item-specific neural representations in WM are maintained in the same way after transitioning from AMI to UMI. We utilized the retrocuing procedure described above, adopting as stimuli dots moving coherently in one of three directions. These stimuli have been successfully decoded with MVPA in several previous studies, including in paradigms with multiple items simultaneously in WM (Emrich et al. 2013), although never in studies that explicitly controlled the FoA.

Methods

Subjects and General Procedure

Eight healthy subjects (3 female; mean age = 24.1, SD = 4.5 years) were recruited from the UW-Madison student population. All were screened for neurological and psychiatric disorders and for their ability to undergo a magnetic resonance imaging session. All subjects performed short-term memory tasks while being scanned with a 3-T MRI scanner. This study was approved by the UW-Madison Institutional Review Board.

The logic of the design was as follows. A one-item delayed-recognition task generated data unambiguously associated with the WM for each of the three critical target directions, which were used to train MVPA classifiers. These classifiers were then used to decode fMRI signal from a second task, in which subjects performed a two-item delayed-recognition task with retrocues that generated AMIs and UMIs. This approach allowed us to estimate classifier evidence for AMIs and UMIs using MVPA, the most direct measure of the level of activity of neural representations.

fMRI Data Acquisition and Preprocessing

Whole brain images were acquired with the 3 T MRI scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. High-resolution T1-weighted images were acquired for all subjects with an FSPGR sequence (8.132 ms time repetition (TR), 3.18 ms time echo (TE), 12° flip angle, 156 axial slices, 256 × 256 in-plane, 1.0 mm isotropic). Blood oxygen level-dependent (BOLD)-sensitive data were acquired using a gradient-echo, echoplanar sequence (2 s TR, 25 ms TE) within a 64 × 64 matrix (39 sagittal slices, 3.5 mm isotropic). Ten MRI acquisition runs were collected for each subject, alternating between 11-min blocks of a two-item delayed-recognition task with retrocues and 7-min blocks of a one-item delayed-recognition task: four of the latter and six of the former were collected.

All fMRI data processing was performed in AFNI (Cox, 1996). Subjects' functional scans were re-aligned to the final volume of the last functional run, then to the anatomical scan collected for each subject. The processing steps were slice-timing correction, detrending, conversion to percent signal change, and spatial smoothing with a 4-mm FWHM Gaussian kernel.

Tasks

Prior to the main experimental session, subjects performed a brief practice session to become familiar with the behavioral tasks. Subjects then underwent fMRI while performing two different delayed-recognition tasks in alternating blocks: a one-item delayed-recognition task and a two-item delayed-recognition task with retrocues. Both tasks required subjects to briefly remember the direction of motion of dots moving within a circular aperture.

In both tasks, the directions of motion on the majority of trials took one of three values: 73°, 193°, or 313°. The remaining trials had dot arrays moving in pseudo-randomly selected directions. These trials were included to reduce the likelihood of subjects realizing that the same three directions re-occurred across trials; this strategy was also used previously (Emrich et al. 2013). In the pre-experiment practice sessions, all stimuli were drawn randomly from a uniform distribution. In post-experiment debriefing, all subjects denied awareness that most of the stimuli were drawn from a set of three directions of motion.

For the duration of the experiment, the difficulty of the task was adjusted to keep behavioral performance at ~75% accuracy across subjects. This was accomplished by adjusting the degree of mismatch between to-be-rejected memory probes and the memory target, using a dynamic staircasing procedure (Levitt, 1971).

One-Item Delayed-Recognition Task

Each trial began with the presentation of an array of coherently moving dots (sample, 1 s), followed by an 8-s delay period (Fig. 1A). At the end of the delay period, a second array of moving dots appeared (probe, 2 s). Subjects were required to compare the direction of motion of the probe array with the remembered direction of the sample, indicating whether the directions were the same or not by pressing one of two buttons. Subjects received immediate feedback, with the fixation cross color changing to green following correct responses and red

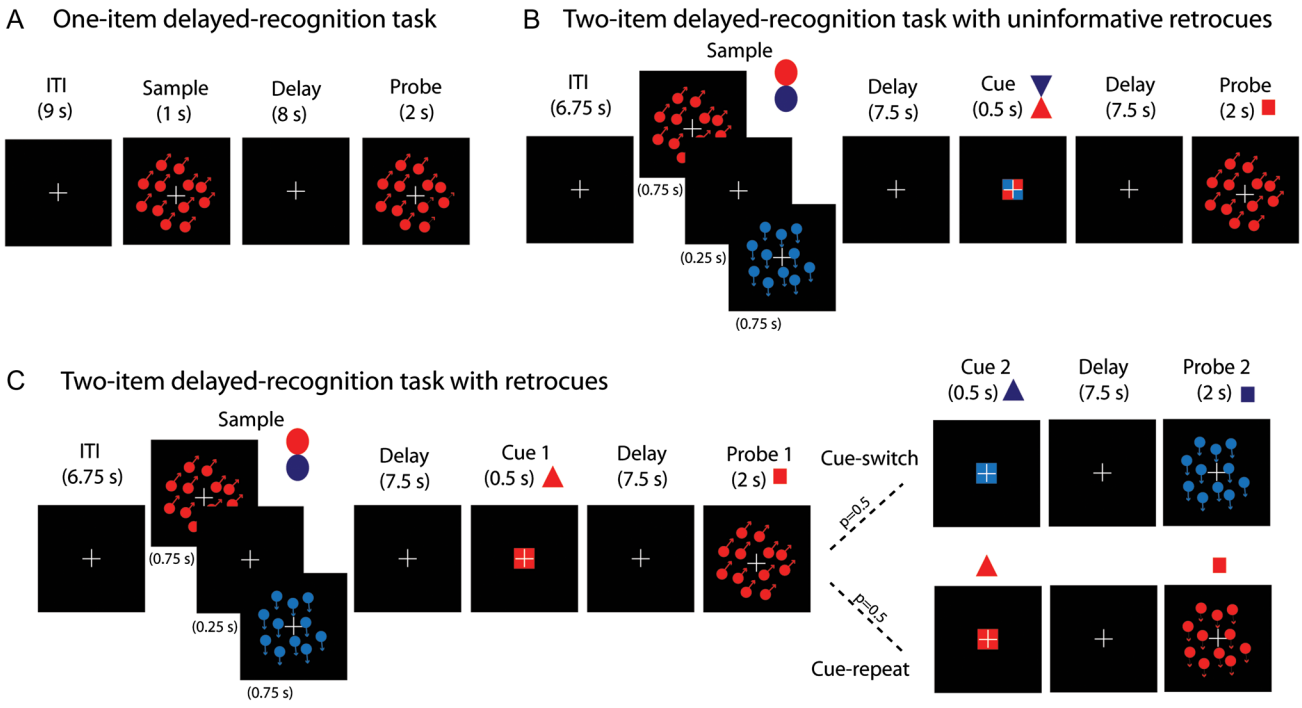

Figure 1. Delayed-recognition tasks for motion direction. (*A*) The one-item delayed-recognition task was used to generate date for classifier training. The data from the last 6 s of the delay period, accounting for a 4-s hemodynamic lag, were used to train the classifiers. Subjects responded to the memory probe with a button press to indicate a match or a nonmatch. (*B*) The two-item delayed-recognition task with an uninformative retrocue was included as an additional control condition for the two-item task with retrocues. In the uninformative-cue trials, subjects saw a retrocue that contained both red and blue, meaning that both items had to be prioritized in preparation for the memory probe. (*C*) In the two-item delayed-recognition task with retrocues, the color of centrally presented squares (the retrocues) indicated which of the two memory items would be relevant for the upcoming probe. The item prioritized by the retrocue is assumed to be in the FoA, whereas the uncued item is assumed to be unattended.

following incorrect responses. Subjects performed four blocks of this task; each block comprised 22 trials over 7 min. Trials were balanced so that stimuli of both colors (red and blue) appeared equally often, and so that there were 24 trials each of the three dot movement directions plus 16 trials with pseudo-random movement directions.

Two-Item Delayed-Recognition Task with Retrocues

Each trial began with the serial presentation of two arrays of moving dots (0.5 s each, with 0.5 s of fixation in between), one red and one blue (the order of the colors was randomized—Fig. 2). After an initial delay period (8 s), a red or blue square appeared over the fixation cross (cue 1, 0.5 s), indicating by its color which of the memory items would be the target of the first memory probe. After another delay period (7.5 s), a memory probe (another array of coherently moving dots, matched in color to the cue) appeared to which subjects responded in the same manner as the one-item delayed-recognition task. After feedback, a second cue (cue 2, 0.5 s) appeared—on half of the trials it indicated the same item as the first cue (cue-repeat trials), and on the remaining trials it indicated the initially uncued item (cue-switch trials). The presence of cue-switch trials guaranteed that subjects could not simply forget the item that was initially uncued, because there was a 50% likelihood that this item would be the target of the second probe. Feedback was provided after each response. Randomly intermixed within the blocks of the two-item task with retrocues were trials in which the first retrocue was uninformative about the target of the memory probe, in that it was half red and half blue (Fig. 2); on these trials, subjects had to wait until they saw the color of the probe to know which memory item was the target. These uninformative-cue trials ended after the offset of the first probe. In all other ways, including timing and appearance of the stimuli, delay period length, and response period, these trials were identical to the two-item delayed-recognition task with retrocues. Altogether, there were 126 two-item trials, with the following distribution: 72 were two-cue trials with canonical motion directions; 12 were two-cue trials with random motion directions; 36 were uninformative-retrocue trials with canonical motion directions; and 6 were uninformative-cue trials with random motion directions.

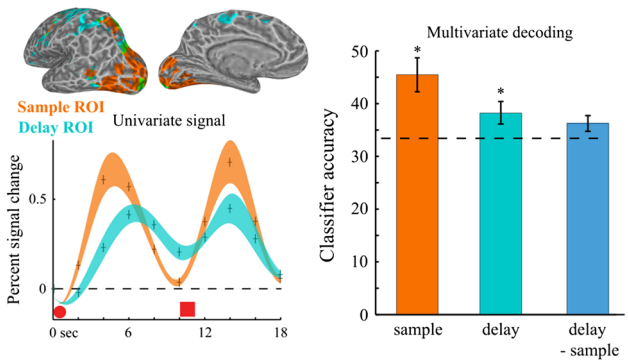

Figure 2. Univariate and multivariate analysis of data from the one-item delayed-recognition task. Pictured in the top left are ROIs (from a representative subject), constructed by taking voxels with the most significant phasic response to the sample (orange) or with the most significant sustained, elevated delay-period signal (cyan). Areas in green show voxels that appear in both ROIs. The aggregated BOLD signal from these ROIs is illustrated in the bottom left, with sample onset at 0 s and the probe onset at 10 s. The width of the ribbons represents mean ± SEM, with spline interpolation (third-degree polynomial) across the group-averaged results to create continuous curves. On the right are plotted the MVPA results obtained from k-fold cross-validation decoding of the direction of motion from delay-period activity in these regions.

Analysis of Behavioral Data

Even though subjects' overall performance was kept near 75% accuracy due to the dynamic staircasing procedure, performance could vary between different response types. We compared reaction times and accuracies between different trial types using two-tailed, paired *t*-tests, primarily to ensure that subjects' behavior accorded with our expectations (e.g., that accuracy was higher on the second probe of cue-repeat trials, after feedback had been provided).

Generation of ROIs

Functional ROI Generation

We performed a general linear model (GLM) analysis for each subject on the data from the one-item delayed-recognition task in order to identify brain regions of interest (ROIs) for the MVPA. Each epoch of the task, including the sample, delay, and probe, was modeled, along with covariates to control for motion and block-specific effects. The sample and probe were modeled as 2 s boxcars and the delay as a boxcar of 8 s. All were convolved with a canonical hemodynamic response function. Each of these independent regressors was entered into a modified GLM for analysis using AFNI. We were particularly interested in voxels showing significant activation to the sample and significant sustained, elevated activation during the delay period. For each subject, we extracted the 2000 voxels with the highest positive *t*-statistic associated with the sample and delay regressors. These voxels composed the "sample" and "delay" ROIs used for MVPA. We also computed a "delay minus sample" mask for each subject, in which we extracted the voxels that showed the maximal positive difference between the t-statistics for the delay and sample regressors from the subject-specific GLM—this procedure had the effect of selecting for delay-specific voxels that did not exhibit transient sample-evoked signal. This approach to ROI generation has been used in prior studies in our laboratory (Riggall and Postle 2012; Emrich et al. 2013) has the advantage of accounting for individual differences in task-relevant neural activity. Additionally, because the GLM used for ROI generation did not model the different stimulus types, the generation of the ROIs is effectively independent from the subsequent classification of trials by stimulus type. Thus, our approach does not represent "double-dipping," which can be defined as selecting voxels based on properties that will be used for a subsequent statistical test, and so the success or failure of decoding analyses can be validly assessed.

Anatomical ROI Generation

Anatomical ROIs were generated using an automated parcellation method from FreeSurfer. Briefly, a surface mesh model was reconstructed for each subject. Each subject's surface was then auto-parcellated based on the folding pattern of the gyri and sulci. We generated ROIs corresponding to occipital cortex regions V1, V2, and MT+ in this manner. We also hand-drew ROIs for the intraparietal sulcus, which also encompassed much of the bordering superior and inferior parietal lobules, and the prefrontal cortex, encompassing all of Brodmann areas 9 and 46 and including parts of Brodmann areas 10, 11, and 47.

Multivariate Pattern Analyses of fMRI Data

MVPA was performed in MATLAB using the Princeton MVPA toolbox (http://code.google.com/p/princeton-mvpa-toolbox). The classification algorithm used for this analysis was logistic regression, with a penalty term of 10. The classification was performed in several functionally and anatomically defined ROIs. No other preclassification feature selection was performed.

For all classification analyses, we used only those trials with directions of motion drawn from the three canonical directions (73, 193, 313), and discarded the first trial of each run. We did not throw out trials in which subjects gave an incorrect response, because our staircase procedure forced performance to stay near the 75% mark, thereby permitting us to assume that the same processes were engaged, at a comparable level, on each trial; and to do so for both trial types. All neural data were z-scored across trials, within runs, before MVPA.

MVPA Validation and Training—One-Item Delayed-Recognition Task

First, the classification procedure was validated using k-fold cross-validation on the data from the one-item delayed-recognition task. Preprocessed fMRI data from the last 6 s (three volumes) of each 8-s delay period, after accounting for a 4-s hemodynamic lag, were used for the analysis. Our analysis scheme considered each functional volume (acquired over a 2-s TR) as a separate training exemplar, so that every trial yielded three exemplars. Each exemplar had associated with it an array of features corresponding to BOLD signal in voxels in the ROI used. The k-fold cross-validation scheme ($k = 68$, due to tossing the first trial of each run) trained a classifier on data from 67 trials and then used this classifier to test the excluded three volumes from the one withheld trial. This process was repeated until every trial had been held out for testing. The statistical significance of classifier accuracy was evaluated by performing a one-sample, one-tailed *t*-test comparing accuracy to chance performance (33%). Data from one subject for which the cross-validation classification accuracy was below chance were not included in subsequent analyses.

MVPA Decoding—Two-Item Delayed-Recognition Task with Retrocues

The classifiers trained on data from the one-item delayed-recognition task were then applied to the data from the two-item delayed-recognition task with retrocues, producing a measure of classifier evidence for each stimulus at each time point (note that classifier evidence can be construed as an estimate of the similarity between patterns of activity in a training data set and test data set). Classifier evidence was averaged separately across cue-switch and cue-repeat trials for the initially cued, initially uncued, and not present stimulus. In order to assess the significance of the critical delay-period evidence, we used the mean of the BOLD signal acquired during during two TRs, in the middle 4 s of the delay period, accounting for a 4-s hemodynamic lag. Excluding the first and last volumes of each delay period minimized any influence of the evoked responses to the cues and probes, which were not signals of interest in this analysis—however, including the data from these TRs did not alter the pattern of results. We first computed an omnibus ANOVA, incorporating data from each of the three delay periods, both main trial types (cue-repeat and cue-switch), and each of the stimulus types, sorted trial-by-trial on the basis of whether the stimulus was selected by the first cue, present on the trial but not selected by the first cue, or not present on the trial (thus, the three item-level classifier evidence estimates obtained from each trial were each sorted as initially cued, initially uncued, and not present).

We also performed a follow-up analysis to address the possibility that UMIs may be encoded by different patterns of brain activity than AMIs, including that they may be stored in different brain areas altogether, by running a leave-one out cross-validation analysis on the two-item data with a roving searchlight. We restricted this analysis to the critical postcue delay period, utilizing three functional volumes from each trial, corresponding to the last 6 s of the first postcue delay period (trial time 11.25 –17.25 s). Because of the presence of two stimuli on each trial, we performed separate analyses for the presence of each stimulus for the cued and uncued cases. To illustrate the scheme, for the cued case for motion in the direction of 73°, the 24 trials with this stimulus present as the first-cued memory item were labeled as "group A," regardless of the identity of the second memory item; the 24 trials without this stimulus, that is, trials with dot-motion directions of 193° and 313°, were labeled as "group B." In this way, groups A and B differ by the presence of the critical stimulus 73°. This is then repeated for each of the three stimuli, and for the uncued condition as well (wherein group A is defined as those trials with the critical stimulus present but not selected by the first retrocue). Using this method of labeling the two-item data, we performed a two-step classification using the searchlight method of Kriegeskorte et al. (2006) for feature selection combined with a leave-one-out classification design. Briefly, for each voxel we constructed a spherical ROI with a radius of four voxels, constrained by the whole brain mask. We then trained a classifier for each spherical ROI using the first four fMRI runs, and then tested this classifier on the fifth fMRI run, with the resultant accuracy assigned to the voxel at the center of the spherical searchlight ROI. From these initial searchlight results, we selected the voxels with the 200 highest decoding accuracies to compose a searchlight-defined ROI, and retrained a new classifier on the same four fMRI runs using only these voxels. Finally, we tested this classifier on the sixth, as-yet-unused fMRI run, allowing a test of the generalization of the information identified by the searchlight procedure. This analysis was iterated six times, so that each of the six runs was utilized once as the final testing run. This method is substantially similar to that described in the documentation for the Princeton MVPA Toolbox (https://github.com/PrincetonUniversity/princeton-mvpa-toolbox/wiki).

Bayesian Statistical Analysis

Because this experiment was designed to test for the presence or absence of evidence for neural representations, we calculated Bayes factors for the critical comparisons of evidence between memory items and evidence for an absent stimulus (the latter of which served as a baseline) to better evaluate our results. The Bayes factor can be understood as the ratio of the likelihood of the alternate hypothesis compared with the null hypothesis. In order to calculate a Bayes factor, it is necessary to provide a prior probability distribution for the alternate hypothesis—to do this, we used a previously published and validated calculator (Dienes, 2014). To constrain this distribution, we utilized the decoding results from the one-item task—specifically, the difference in evidence between the correct and the two incorrect stimulus types on each trial—to estimate a maximum plausible value above baseline for the evidence of the AMI and UMI, reasoning that one would not expect the

evidence above baseline in the two-item task to exceed the differences found in the simpler, one-item task used to train the classifiers. Because specifying any particular value or shape for the distribution of expected evidence differences would require some guesswork, we chose the simplest possible prior, a uniform distribution extending from zero difference to the maximum plausible difference. Alternatives, such as a half-Gaussian distribution centered at the maximum plausible value or a Gaussian centered at the midpoint, were also considered, but ultimately rejected because they might excessively bias the results—but, Bayes factors calculated using those prior distributions agreed substantively with those of the uniform distribution.

Results

Behavioral Results

In the one-item delayed-recognition task, the adaptive staircase procedure manipulation kept the performance at 73.6% (SD = 5.8%). In the two-item delayed-recognition task, the overall performance across all probes was 73.0%, SD = 4.6% (among the seven subjects whose neural data from the two-item, delayed-recognition task were analyzed). Subjects were more accurate on the second probe of cue-repeat trials (M = 81 %, SD = 6.3%) than on the second probe of cue-switch trials (M = 64.9%, SD = 8.4%; $t(6) = 5.7$, $P = 0.001$)—this may be due to feedback after the first probe, as performance improved from the first (M = 71.0%, SD = 6.5%) probe to the second probe on cue-repeat trials ($t(6) = 3.6$, $P = 0.01$). Because a staircasing procedure continuously adjusted the difficulty of the task to maintain accuracy near 75%, these behavioral results can only be taken to indicate that subjects were engaged with the task and were performing it correctly.

Classification Results—One-Item Delayed-Recognition Task

The first step of our analysis entailed the k-fold cross-validation classification of trials from the one-item delayed-recognition task on the basis of the direction of motion of the memory sample. This analysis was performed using multivariate patterns of delay-period BOLD signal from two different functionally-defined ROIs, defined uniquely for each subject. The "sample" ROI comprised the 2000 voxels with the most significant sample-evoked activity, and the "delay" ROI comprised the 2000 voxels with the most significant sustained, elevated delay-period activity. The sample ROI primarily included voxels in visual cortex and area MT+, though some voxels were in parietal and frontal regions; the delay ROI partially overlapped with the sample-evoked ROI, but included much more extensive portions of parietal and frontal cortex (Fig. 2). In the sample ROI, decoding was well above chance (M = 45.3%, SD = 8.9%, $P = 0.006$; Fig. 2); in the delay ROI, decoding was less successful, but still significantly above chance (M = 38.2%, SD = 5.2%, P = 0.02; Fig. 2). When the delay ROI was restricted only to voxels showing selective sustained delay-period activity, that did not show a transient sample-evoked response, decoding trended toward being significantly different from chance (M = 36.9%, SD = 5.1%; $P = 0.07$). Though comparing classifier accuracies between regions was not the primary aim of the study, we also performed classification in a series of anatomical ROIs, selected because of their potential importance for WM (Fig. 3). Classification was well above chance in visual areas V1, V2, and area MT+ (M = 39.6%, 48.6%, and 39.6%, respectively), but closer

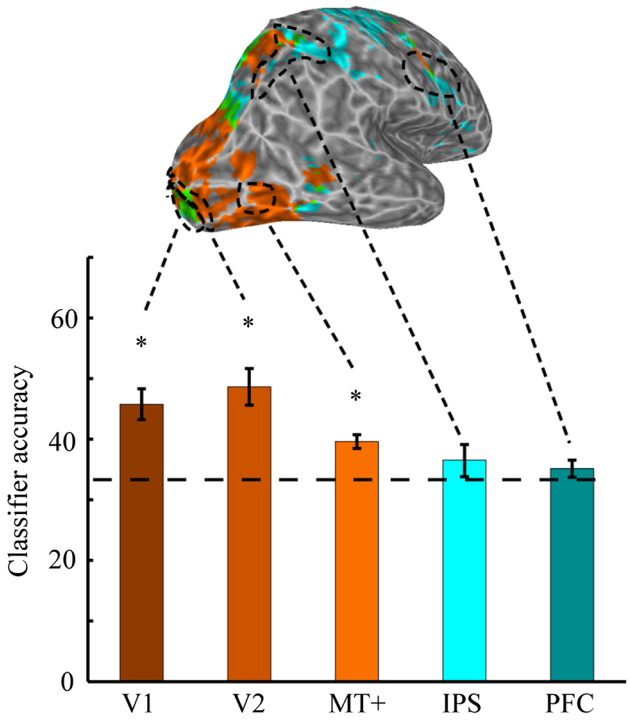

Figure 3. Multivariate analysis of one-item delayed-recognition task data in anatomical ROIs. Plotted are the MVPA results obtained from k-fold cross-validation decoding of the direction of motion from delay-period activity in the anatomically defined ROIs.

to chance levels in prefrontal cortex and parietal cortex (M = 35.4% and 34.2%, respectively). Because the sample ROI yielded excellent cross-validation accuracy and our aim was to utilize a decoding algorithm sensitive to sensorimotor representations, we used this ROI in subsequent decoding analysis of data from the two-item delayed-recognition task with retrocues.

Classification Results—Two-Item Delayed-Recognition Task with Retrocues

In order to extract estimates of the level of activation of item-level representations in different states in short-term memory, we applied classifiers trained on the one-item delayed-recognition task to data from the two-item delayed-recognition task with retrocues. We only analyzed those trials in which the directions of motion of the dots were drawn from the three canonical directions (73°, 193°, and 313°). The evidence obtained for each stimulus was collapsed with respect to whether that item was present (initially cued or initially uncued) or not present on each trial (Fig. 4). In both cue-repeat and cue-switch trials, the evidence for the initially cued item remained elevated following the presentation of the first retrocue—note that the cue was simply a colored square, and therefore could not, by itself, explain the elevated evidence for one direction of motion over another. In cue-repeat trials, the same pattern was observed after the onset of the second retrocue. In cue-switch trials, evidence for the previously irrelevant direction of motion was reinstated above baseline, and evidence for the previously relevant direction of motion dropped to baseline. Note that prior to the onset of the stimuli, all three evidence traces were relatively high. This is due to the influence of the previous trial's probe, which, having ~1/3 probability of targeting any of the three canonical stimuli, had the effect of

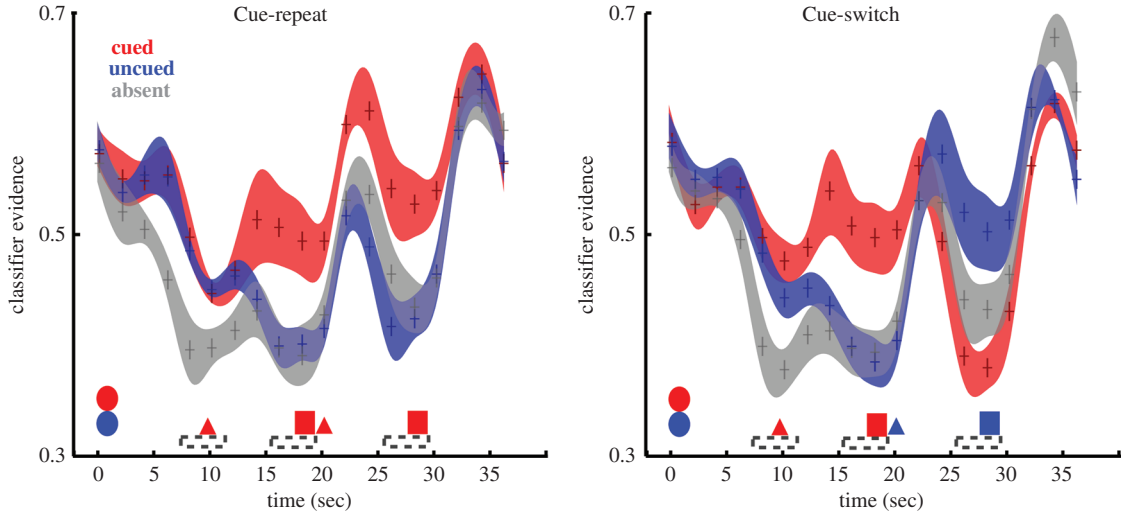

Figure 4. Multivariate analysis of two-item delayed-recognition task with retrocues. Mean classifier evidence values are plotted separately for cue-repeat (left) and cue-switch trials (right). The three canonical directions of motion were collapsed across all trials into new categories defined by the first cue: cued (red) is the direction indicated by the first cue, uncued (blue) is the direction of the other moving-dot array which is not selected by the first cue, and absent (gray) is the direction of motion not present on that trial. Time is represented on the horizontal axis. The critical delay-period intervals used for statistical analysis are indicated by the gray-dashed lines. The geometric symbols above the timeline represent events in the trial, with stimulus presentation (circles) from 0 to 1.75 s, the first cue (triangle) at 9.75 s, the first probe (square) at 17.25 s, the second cue (triangle) at 19.25 s, and the second probe (square) at 26.75 s. The width of the ribbons represents mean ± SEM, with spline interpolation (third-degree polynomial) across the group-averaged results to create continuous curves. The actual mean evidence values are also plotted for each trace, centered at the midpoint of the TR.

equivalently increasing each of the three evidence. We performed a 3 × 2 × 3 repeated measures ANOVA with the factors of delay period (first/second/third), trial type (cue-repeat/cue-switch), and stimulus type (initially cued/initially uncued/not present). There was a significant three-way interaction ($F_{4,24}$ = 4.25, P = 0.01), reflecting the reversal in evidence values between initially cued and initially uncued stimuli between delay 2 and delay 3 that occurred only on cue-switch trials. Because procedure was to evaluate for active neural representations by testing for above-baseline evidence, we calculated Bayes factors for the critical comparisons of delay-period evidences to determine which of two alternate hypotheses was more likely to be true: the null hypothesis that item-level classifier evidence was not higher than baseline, or the alternative hypothesis that the evidence was above baseline for UMIs. For the crucial delay period after the first retrocue, the Bayes factors for the cued stimuli were 4.14 and 10.15 for cue-repeat and cue-switch trials, respectively; these numbers signify how much more likely it is that the evidence is greater than baseline rather than the null hypothesis that it is not. For the uncued stimuli in the first postcue delay period, the Bayes factors for cue-repeat (0.31) and cue-switch (0.21) trials indicate that the null hypothesis is approximately 3.2 and 4.8 times more likely than the alternative hypothesis for the two trial types, respectively. The t-statistics and P-values for all pairwise comparisons of delay-period evidence can be found in Supplementary Table 1. We also performed a set post hoc power analyses in order to illustrate our statistical power under several different assumed effect sizes. To create these effect sizes, we used the difference between the AMI and baseline classifier evidence and the pooled standard deviation over these two measures. Post hoc, we were 94.9% powered to have detected the AMI in the cue-stay trials and 97.9% powered on cue-switch trials. To have detected a representation for a UMI with half the evidence level of the AMI (relative to the baseline amount of evidence for the absent stimulus), we were 75.8% powered on cue-stay trials and 76.1% powered on cue-switch trials. If the UMI had only one-fourth the relative level of evidence of an AMI, then our power dropped to 32.2% for cue-stay trials and 32.3% for cue-switch trials.

As an additional control analysis, we also examined the uninformative-cue trials using the classifiers trained on delay-period activity from the one-item task in the sample-defined ROI (Fig. 5). The postcue delay period from the uninformative-cue trials provides a highly relevant additional control condition. During that delay period, two items must be held in memory, equally prioritized in anticipation of a probe, at the exact same postsample delay as the critical delay period analyzed for the UMIs above. A Bayesian analysis of the evidence from the evidence from this delay period yields Bayes factors of 4.45 and 1.58 for the both the to-be-probed and not-to-be-probed AMIs, respectively, indicating how many times more likely it is that the MVPA evidence for these two items are above baseline rather than at baseline level. This result shows that our classification analysis can show above-baseline evidence for two memory items after a retrocue, when both remain AMIs, at the same time postsample that evidence for UMIs was found to be at baseline.

Finally, because we were unable to find above-baseline evidence for the UMI in the sample-defined ROI, we also performed a supplemental whole brain searchlight classification analysis, in which we trained the classifiers only on the two-item task (see Methods for a description of the exact classification scheme). This analysis failed to successfully decode the AMI (M = 48.9% [chance = 50%], $t(7)$ = 0.54, one-tailed P = 0.6) or the UMI (M = 50.5%, $t(7)$ = 0.26, one-tailed P = 0.4).

Discussion

In this study, we tested the hypothesis that active neural representations are present for items in short-term memory outside the FoA (UMIs) in the same way that these representations are present for memory items in the FoA (AMIs). This hypothesis is suggested by several models of short-term memory that posit

similar representations, but with intermediate levels of activation, for items in short-term memory that are not selected by the FoA (Cowan, 1988; McElree, 1998; Oberauer, 2002). Two previous studies (Lewis-Peacock et al. 2012; LaRocque et al. 2013) have failed to support this hypothesis, in that they failed to find evidence for elevated neural activation of UMIs. Because the MVPA methods employed in those studies featured only category-level specificity, however, their interpretation is equivocal. The present study, in contrast, employed a design and analysis with a level of specificity that afforded the detection of, and discrimination between, individual stimuli. Using a Bayesian analysis, we were able to show that UMI evidence being at baseline levels is the more likely hypothesis to account for our results. This absence of evidence for active neural representation of UMIs thus represents an important extension of the previous studies' results. We will first discuss the interpretability and limitations of this result before turning to its theoretical implications.

Before proceeding to interpret our results, it is important to acknowledge and discuss their limitations, especially those related to the localization, magnitude, and quality of any neural signal related to UMIs. We showed that such a signal was not evident in subject-specific functional ROIs, defined on the basis of sample-evoked activity. We chose this ROI because it was possible to decode stimulus identity from delay-period activity in this ROI with a comparatively high degree of accuracy (Fig. 2). We cannot rule out, however, the possibility that a memory item could have been retained by a qualitatively different, yet still active, representation when it was not selected by a cue. One concrete example of an analogous phenomenon is from the Emrich et al. (2013) study that used very similar stimuli. In this study, a classifier successfully trained to discriminate stimuli from sample-evoked activity failed to decode stimulus identity from late in the delay period, and the converse was true for a classifier trained on data from late in the delay period. This suggests that the neural representations of our moving-dot stimuli are different when they are being viewed than when they are being held in WM (see also Albers et al. 2013). Another example comes from Lee et al. (2013), who could decode WM representations of visually presented objects from posterior visual areas when subjects were preparing to make a fine-grained perceptual judgment about the memory probe, but not when they were preparing to make a categorical judgment; the converse was true for the prefrontal cortex. (Note, however, that neither of the factors that account for the recoding of stimulus representations in these studies apply directly to the interpretation of the present study.)

A second possibility that we cannot rule out is that the UMI-related neural signal might still be present in the ROI we analyzed, but at a level too small to be detected by the present study. This limitation is mitigated by the Bayesian analyses we employed, which suggested that the null hypothesis of no above-baseline UMI-related signal is actually more likely than the alternative of above-baseline signal. Finally, a third possible limitation of our results arises from the nature of our classifier design. By training on data from the one-item task, we are able to generate classifiers that unequivocally distinguished the retention of single, attended stimulus. These classifiers successfully, simultaneously decoded both memory items during the precue delay period (Fig. 4), one AMI during the postcue delay periods (Fig. 4), and two items during the post-uninformative-cue delay period (Fig. 5). However, it is possible that neural activity related to a UMI might qualitatively differ from that required for remembering a single memory item, in

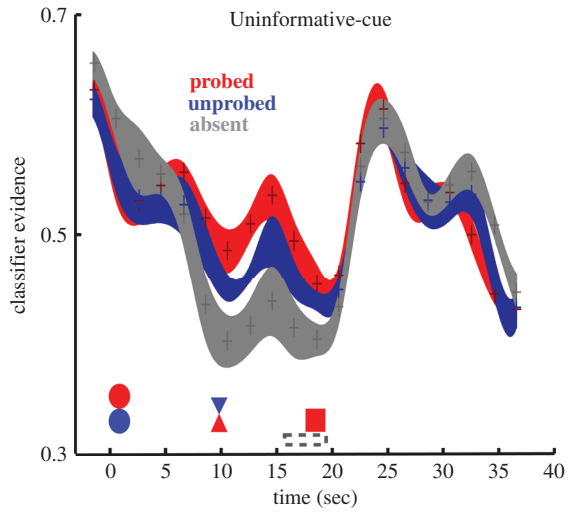

Figure 5. Multivariate analysis of two-item delayed-recognition task with uninformative retrocues. Mean classifier evidence values are plotted for the uninformative-cue trials, using the same plotting conventions as Figure 4 with the following exception. Here, the three canonical directions of motion were collapsed across all trials into new categories defined by which item was probed on each trial: the direction that was target of the memory probe (red), the direction that was present but not probed (blue), and absent (gray), the direction of motion not present on that trial. Time is represented on the horizontal axis, with stimulus presentation (circles) from 0 to 1.75 s, the uninformative cue (two triangles) at 9.75 s, and the probe (square) at 17.25 s. As in Figure 4, the width of the ribbons represents mean ± SEM, with spline interpolation (third-degree polynomial) across the group-averaged results to create continuous curves. The actual mean evidence values are also plotted for each trace, centered at the midpoint of the TR.

which case our classifiers trained on the one-item task would be insensitive to UMIs. Though this study was not designed to test this hypothesis, we attempted to test the possibility of a within-trial recoding of the UMI-related signal into a novel active representation by training and testing classifiers on the two-item task using a whole brain searchlight analysis. Unfortunately, this approach failed to successfully decode the AMI or the UMI. This failure could be due to the limited number of trials available for this analysis (see Methods—only two-thirds of those used for the principal analysis), or the increased noise inherent in training a classifier to be sensitive to one of two stimuli that are present on each trial (as seen in Emrich et al. 2013, classifier performance declines monotonically with load).

The amount of caution we must take in interpreting our result is somewhat mitigated by the considerable sensitivity of MVPA. Unlike univariate analyses, in which regional signal intensity changes are averaged, MVPA makes no assumptions about homogeneity of function in a given brain region (in that all voxels in a region are not assumed to equally contribute to a representation), nor does it require that concurrent changes in signal intensity in individual voxels share the same sign. Indeed, it has been shown that MVPA can successfully decode at an item level the contents of WM in regions that show no elevated delay-period activity (Riggall and Postle 2012; Emrich et al. 2013); furthermore, this method has been shown to be successful at decoding the category of memory items from a data set from which all "category selective" voxels, as identified by GLM, have been removed (Lewis-Peacock and Postle, 2012). Furthermore, MVPA has been shown to be remarkably sensitive to patterns of neural activity reflecting very subtle cognitive distinctions, including: which of two overlapping gratings

subjects are attending (Kamitani and Tong, 2005); the mnemonic precision of memory for directions of motion (Emrich et al. 2013); the orientation of laterally presented stimulus from ipsilateral visual cortex (Ester et al. 2009); and information about visually masked stimuli in the absence of subjective awareness (Haynes and Rees, 2005; Sterzer et al. 2008). In light of this demonstrably high sensitivity of MVPA, and in light of the ability, in this study, to decode both memory items before the onset of the first retrocue, we feel it is permissible to interpret the lack of evidence for active neural representations of UMIs that are similar to those of AMIs. Furthermore, by calculating Bayes factors, which allow the estimation of the likelihood of competing hypotheses being true, we were able to show that the null hypothesis is more likely to be true than the alternative of above-baseline evidence for UMIs.

The results of the present study are interesting to compare with those of Emrich et al. (2013). In that study, subjects also performed a WM task for directions of motion, with one, two, or three moving-dot arrays serially presented on each trial. Both classifier sensitivity to the remembered direction and behavioral estimates of mnemonic precision decreased with increasing memory load, suggesting that the stimulus information decodable from activity in visual cortex is a correlate of behavioral precision. In the present work, however, we found a binary distinction, in that classifier evidence was present only for AMIs, and not for UMIs. This naturally raises the question: in the Emrich et al. (2013) study, were all of the memory items on the load-three trials AMIs? This interpretation invokes what Oberauer has termed the "broad focus" of attention (Oberauer and Hein 2012). In this interpretation, multiple memory items must compete for limited attentional resources, so that each is maintained with equal precision for the duration of the delay period. Another alternative, consonant with an FoA limited to one item (McElree 1998), is that each item was sequentially selected and deselected by the FoA in a "refreshing"-like process (Higgins and Johnson 2009). The lack of temporal specificity inherent in the BOLD signal could make such a process manifest as a constant, intermediate level of decoding for all of the memory items. A third possibility is that on each load-three trial, subjects may have arbitrarily chosen one of the three items to focus on as an AMI, so that "on average," across trials, classification performance for all items appeared to be graded as a function of load. It is even possible that subjects may have adopted each of these strategies on different trials. To tease apart the issue of how multiple items in memory are selected by attention will be a difficult task, perhaps requiring set-size manipulations in the context of a retrocuing paradigm to allow more precise experimental control of the allocation of attention.

An intriguing line of discussion emerging from these results concerns the distinction between WM and LTM. One could pose the question: should a UMI be considered to be "in" WM or LTM? This question is especially relevant because we argue that ongoing elevated neural signal, the "active trace" long held to be a neural marker of WM (Hebb 1949), may not required for UMIs. To resolve this issue, it is helpful to return to the actual behavior being studied; that is, the transient retention, over an interval of ~20–30 s, of the direction in which an array of dots was moving. To our mind, this is incontrovertibly short-term memory in a behavioral sense. It may be the case that mechanisms traditionally associated with LTM, for example, synaptic plasticity, play a role in this behavior. Indeed, we recently performed an experiment in which subjects were given a surprise memory test for items (pictures of everyday objects such as a telephone or a book) seen 1 week earlier in the context of a delayed-recognition task with retrocues (LaRocque, Eichenbaum et al. 2015). There was no improvement in LTM for items which had been UMIs—however, LTM performance was reliably above chance for all item types, illustrating that long-term encoding does occur in the context of a short-term recognition task with retrocues. The behavior of recognizing items after a brief delay, whether they have been continuously attended or not, can be accurately described with the term WM; the results of the present study suggest that this term should not be understood to imply the neurophysiological mechanism of an "active neural trace."

The present results, along with those of previous category-level decoding studies, offer a strong case that UMIs are not maintained with active neural representations—at least, not an active representation of the same sort observed for AMIs, that is detectable in patterns of BOLD or EEG signal. As we have noted previously (LaRocque, Lewis-Peacock et al. 2014), the present design does not allow us to test how UMIs might be retained, if not by an active mechanism detectable in fMRI or EEG signals. However, we can speculate that a weight-based mechanism comprising a network of transiently potentiated synapses (Sugase-Miyamoto et al. 2008; Erickson, Maramara, and Lisman, 2009; Barak and Tsodyks, 2014) would be a physiologically plausible retention mechanism that is consistent with our findings. In this scenario, the pattern of neural activity established by encoding of a memory item in the FoA would vanish after the FoA shifts; left behind, however, would be a network of synaptic weights reconfigured by the recent activity. In this way, when attention reselects a previously UMI, the return of activity to the network of transiently potentiated synapses would permit the reinstatement of the original pattern of neural activity. A perturbation-based approach, such as transcranial magnetic stimulation, could potentially unmask such a latent trace.

In conclusion, the present study adds important additional evidence to the question of whether an active sensorimotor representation is required for WM, by showing that item-level MVPA reveals no evidence for such a representation when items are retained in WM, but outside the FoA.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by the National Institute of Mental Health at the National Institutes of Health (grant numbers MH064498 and MH095984 to B.R.P.). Additional support (to J.J.L.) came from National Research Service Awards (grant numbers T32 GM008692 to the Medical Scientist Training Program and T32 EB011434 to the Clinical Neuroengineering Training Program).

Notes

Conflict of Interest: None declared.

References

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP. 2013. Shared representations for working memory and mental imagery in early visual cortex. Curr Biol. 23(15):1427–1431. .

Barak O, Tsodyks M. 2014. Working models of working memory. Curr Opin Neurobiol. 25:20–24.

Cowan N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychol Bull. 104 (2):163–191. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3054993.

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

Dienes Z. 2014. Using Bayes to get the most out of non-significant results. Front Psychol. 5:1–17.

Emrich SM, Riggall AC, LaRocque JJ, Postle BR. 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci. 33(15):6516–6523.

Erickson MA, Maramara LA, Lisman J. 2009. A single brief burst induces GluR1-dependent associative short-term potentiation: a potential mechanism for short-term memory. J Cogn Neurosci. 22(11):2530–2540.

Ester EF, Serences JT, Awh E. 2009. Spatially global representations in human primary visual cortex during working memory maintenance. J Neurosci . 29(48):15258–15265.

Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. Nature. 458(7238): 632–635.

Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8(5):686–691.

Hebb DO. 1949. The Organization of Behavior: A Neuropsychological Theory. New York: Wiley. Retrieved from http://books.google.com/books?hl=en&lr=&id=gUtwMochAI8C&pgis=1

Higgins JA, Johnson MK. 2009. The consequence of refreshing for access to nonselected items in young and older adults. Mem Cognit. 37(2):164–174.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8(5):679–685.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103(10): 3863–3868.

LaRocque JJ, Eichenbaum AS, Starrett MJ, Rose NS, Emrich SM, Postle BR. 2015. The short- and long-term fates of memory items retained outside the focus of attention. Mem Cognit. 43(3):453–468.

LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. 2013. Decoding attended information in short-term memory: an EEG study. J Cogn Neurosci. 25:127–142.

LaRocque JJ, Lewis-Peacock JA, Postle BR. 2014. Multiple neural states of representation in short-term memory? It's a matter of attention. Front Hum Neurosci. 8:5.

Lee S-H, Kravitz DJ, Baker CI. 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat Neurosci. 16:997–999.

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

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. 2012. Neural evidence for a distinction between short-term memory and the focus of attention. J Cogn Neurosci. 24(1):61–79.

Lewis-Peacock JA, Postle BR. 2008. Temporary activation of long-term memory supports working memory. J Neurosci. 28(35): 8765–8771.

Lewis-Peacock JA, Postle BR. 2012. Decoding the internal focus of attention. Neuropsychologia. 50(4):470–478.

McElree B. 1998. Attended and non-attended states in working memory: accessing categorized structures. J Mem Lang. 38 (2):225–252.

Oberauer K. 2002. Access to information in working memory: exploring the focus of attention. J Exp Psychol Learn Mem Cogn. 28(3):411–421.

Oberauer K. 2005. Control of the contents of working memory–a comparison of two paradigms and two age groups. J Exp Psychol Learn Mem Cogn. 31(4):714–728.

Oberauer K, Hein L. 2012. Attention to Information in working memory. Cur Dir Psychol Sci. 21(3):164–169.

Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. Science. 310(5756):1963–1966. 10.1126/science.1117645.

Riggall A, Postle BR. 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci. 32(38): 12990–12998. Retrieved from http://www.jneurosci.org/content/32/38/12990.short.

Serences JT, Ester EF, Vogel EK, Awh E. 2009. Stimulus-specific delay activity in human primary visual cortex. Psychol Sci. 20 (2):207–214.

Sterzer P, Haynes J-D, Rees G. 2008. Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. J Vis. 8(15):1–12. Retrieved from http://www.journalofvision.org/content/8/15/10.short.

Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, Richmond BJ. 2008. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. PLoS Comput Biol. 4(5):e1000073.