

## Neural Evidence for the Flexible Control of Mental Representations

Jarrold A. Lewis-Peacock<sup>1</sup>, Andrew T. Drysdale<sup>2</sup> and Bradley R. Postle<sup>3</sup>

<sup>1</sup>Department of Psychology and Neuroscience, University of Texas at Austin, Austin, TX 78712, USA, <sup>2</sup>Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, NY 10065, USA and <sup>3</sup>Department of Psychology and Psychiatry, University of Wisconsin–Madison, Madison, WI 53706, USA

Address correspondence to Jarrold Lewis-Peacock, 108 E. Dean Keeton St., Austin, TX 78712, USA. Email: jalewpea@utexas.edu

**This study was designed to explore neural evidence for the simultaneous engagement of multiple mental codes while retaining a visual object in short-term memory (STM) and, if successful, to explore the neural bases of strategic prioritization among these codes. We used multivariate pattern analysis of fMRI data to track patterns of brain activity associated with three common mental codes: visual, verbal, and semantic. When participants did not know which dimension of a sample stimulus would be tested, patterns of brain activity during the memory delay indicated that a visual representation was quickly augmented with both verbal and semantic re-representations of the stimulus. The verbal code emerged as most highly activated, consistent with a canonical visual-to-phonological recoding operation in STM. If participants knew which dimension of a sample stimulus would be tested, brain activity patterns were biased toward the probe-relevant stimulus dimension. Interestingly, probe-irrelevant neural states persisted at an intermediate level of activation when they were potentially relevant later in the trial, but dropped to baseline when cued to be irrelevant. These results reveal the neural dynamics underlying the creation and retention of mental codes, and they illustrate the flexible control that humans can exert over these representations.**

**Keywords:** encoding, fMRI, MVPA, short-term memory, working memory

### Introduction

One of the central concepts of cognitive psychology is that of the mental code—the hypothetical format in which information is represented in the brain. The existence of distinct mental codes is inferred from evidence of selective interference, in which, for example, short-term memory (STM) for one type of information is disproportionately disrupted by a concurrent task involving the same type of information, versus when the concurrent task involves information from another domain (e.g., [Baddeley and Hitch 1974](#)). This is a domain within the study of cognition in which research on STM and working memory makes contact with the broader notion of “thinking” ([Johnson-Laird 1995](#); [Jonides 1995](#)).

The flexibility with which mental codes are engaged has been observed in STM tasks in which participants strategically recode information during the delay period from the format presented at sample to one best suited for judging the impending memory probe ([Tversky 1973](#)). However, the default, perhaps obligatory ([Simons 1996](#); [Postle et al. 2005](#); [Postle and Hamidi 2007](#)), tendency of humans is to recode information into a phonologically based verbal form ([Shulman 1971](#)), a phenomenon referred to as “the very lifeblood of the thought processes” ([Miller 1956](#)). One illustration is the phonological similarity effect ([Conrad and Hull 1964](#)), in which STM for visually presented words, letters, or pictures suffers if the names for items on the list rhyme. The dissimilarity advantage

is abolished, however, when participants are required to engage in concurrent articulation ([Baddeley and Hitch 1974](#); cf. [Camos et al. 2011](#)). The inference is that the visual-to-phonological recoding of stimuli is blocked by concurrent verbal processing.

To date, most support for models of mental coding has been limited to inferences drawn from behavioral results, because neural correlates of hypothetical mental codes are notoriously difficult to measure directly. This has begun to change, however, with the advent of multivariate pattern analysis (MVPA) of neuroimaging datasets ([Haxby et al. 2001](#); [Kamitani and Tong 2005](#); [Haynes and Rees 2006](#); [Norman et al. 2006](#); [Pereira et al. 2009](#)). Thus, for example, distributed patterns of functional magnetic resonance imaging (fMRI) signal across face-selective voxels in the inferior temporooccipital cortex are diagnostic of whether participants are preferentially attending to the race or the gender of a face that they are viewing (e.g., [Chiu et al. 2011](#)), as is signal from inferior temporooccipital cortex or prefrontal cortex, depending on whether participants are attending to fine-grained perceptual features or category membership of a visual object, respectively ([Lee et al. 2013](#)). Such studies support the idea that one can measure neural states that correspond to the mental codes hypothesized by cognitive psychology ([Lewis-Peacock and Postle 2012](#)).

The work reviewed up to this point might lead to the idea that information can only be represented in one mental code at a time. However, demonstrations of a release from proactive interference, when performance improves if experimental context is changed, provide evidence that information is recoded and retained, in parallel, in as many representational formats as are afforded by the to-be-remembered information (the “multiple encoding” hypothesis; [Wickens 1973](#)). Indeed, many cognitive theories explicitly model the multidimensional nature of mnemonic representations. The principal goals of the present study, therefore, were to seek evidence for multiple encoding at the level of neural representation and, if successful, to explore the neural bases of strategic prioritization among multiple mental codes being maintained in STM.

The present work was performed in the context of several prior studies from our group that have employed variants of the procedure of first presenting one-or-more to-be-remembered stimuli (target(s)), then during the ensuing delay period, presenting a retrocue that indicates which target—or which dimension of the target—will be relevant for the impending memory probe. In experiments in which two targets are presented, we have consistently found that MVPA evidence for the target that was not selected by the retrocue drops to baseline levels, despite the fact that participants know that there is a 50% likelihood that this not-cued target will later be cued by a second retrocue, and therefore needed in order to evaluate the second of two serially occurring memory probes. Further,

behavioral performance confirms that initially not-cued target information is nonetheless retained in STM. These findings therefore suggest that active neural representation of an item, as indexed by MVPA evidence, may not be necessary for its retention in STM (Lewis-Peacock et al. 2012; Lewis-Peacock and Postle 2012; LaRocque et al. 2013). Thus, an additional goal of the present research, assuming successful MVPA decoding of multiple dimensions of a single stimulus held in STM, was to determine what would be the dynamics of MVPA evidence for un-prioritized dimensions of a multiply encoded stimulus. Note that, despite the intuition that it may be difficult to think about one dimension of a stimulus to the exclusion of others, we have preliminary evidence that this may be possible: When participants are first presented with a field of dots moving in one direction, then informed via a retrocue that memory for the speed of motion will be tested, MVPA decoding of the direction of motion falls to chance levels (Riggall and Postle 2012). The present study would also permit us to examine this phenomenon more rigorously.

## Materials and Methods

### Participants

Twelve participants (all right-handed; three men; ages 19–28) were recruited from the undergraduate and medical campuses of the University of Wisconsin–Madison. None reported any medical, neurological, or psychiatric illness, and all gave informed consent.

### Design

The experiment proceeded in three phases, with the logic that the first two would generate fMRI data in which participants were naive to the goals of the experiment and thus any demand characteristics influencing their thought processes would be minimized. Data from these first two phases would be decoded with pattern classifiers trained on data from the third, in which participants were instructed how and what to think. This analysis procedure (modeled on Lewis-Peacock et al. 2012) enabled the assessment of the distinct neural states (presumed to underlie distinct mental codes) recruited to encode and retain information in STM.

### Stimulus Materials

Public domain images of familiar objects were downloaded from Google Images (<http://images.google.com>). Colored images of a single object on a white background were preferred, but many images not fitting this criterion were also selected and subsequently modified using image processing software. Nineteen categories were identified to guide the collection of stimuli: airport, baseball, bathroom, beach, bedroom, bowling, car, cinema, classroom, doctor, football, grocery, gym, kitchen, living room, office, park, restaurant, and tools. At least six items were selected for each category yielding over 120 total stimuli (Fig. 1). Image processing software was used to enhance the contrast of the foreground object, to remove background items or color, and to resize all images to 400 × 400 pixels with 72 pixels-per-inch.

Words were nouns, verbs, and adjectives selected from an online psycholinguistic database ([http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa\\_mrc.htm](http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm)) with concreteness, imageability, and Brown verbal frequency within one standard deviation of the mean of the entire database. Pseudowords consisted of single-syllable, pronounceable letter strings. Intended pronunciation of the pseudowords was based on standard English (i.e., a string ending with the letter “e” indicated a long vowel sound and a string ending with a double consonant indicated a short vowel sound). No compound vowels (e.g., “ou”) were used. Line stimuli consisted of a pair of line segments, each line tilted between 10° and 170° (excluding 90), at intervals of 10°, away from vertical. Tilt angles of 0°, 90°, and 180° were avoided to discourage participants from recoding the stimuli into categorical codes (e.g., “vertical”).



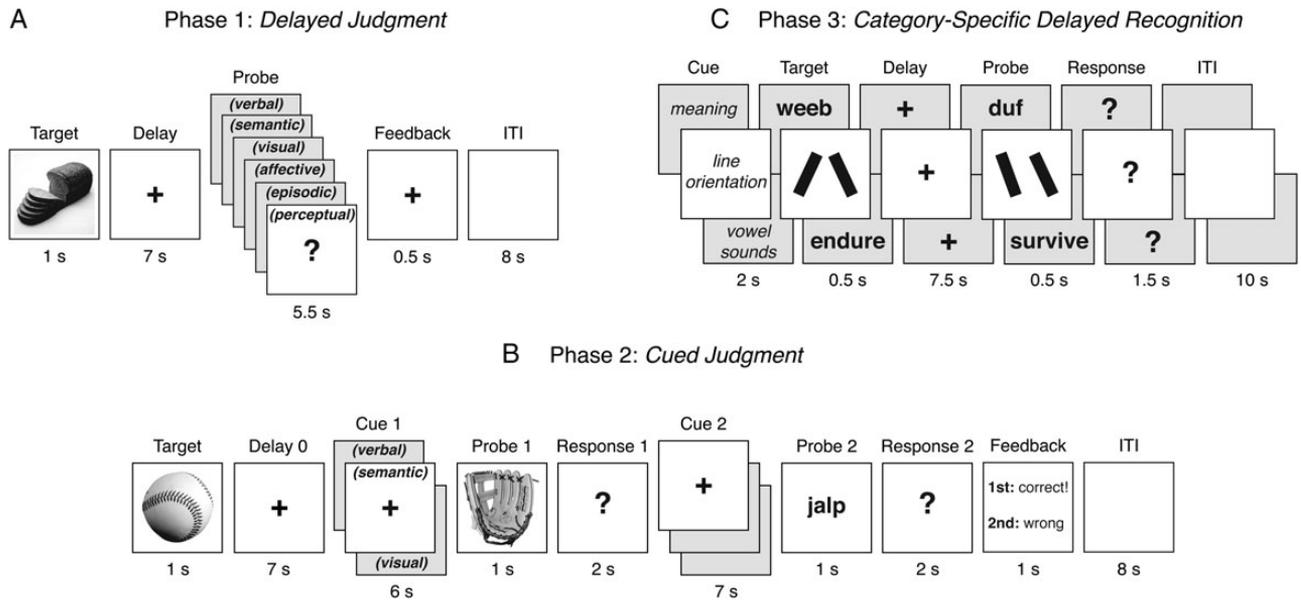
**Figure 1.** Object stimuli. Pairs of object stimuli used as targets in Phases 1 and 2 of the experiment are shown from 18 different schematic categories (e.g., a loaf of bread and a carton of eggs were representatives of the “food” category).

### Task Procedures

The study consisted of three phases that were all completed in a single scanning session lasting 2 h (Fig. 2). Participants received instructions and performed practice trials for the Phase 1 (“Delayed Judgment”) task outside the scanner before the experiment began (30 min). Details of Phases 2 and 3 were not discussed, but participants were informed that they would learn two additional tasks inside the scanner. After completing the Delayed-Judgment task, instructions and practice trials for the Phase 2 (“Cued Judgment”) task were administered inside the scanner followed by the experimental trials for that phase. The instructions, practice trials, and experimental trials for the Phase 3 (“Category-Specific Delayed Recognition”) task followed inside the scanner. Task procedures for each phase will now be described in detail.

### Phase 1: Delayed Judgment

Participants performed 54 trials of a Delayed-Judgment task, which required the short-term retention of a picture of a familiar object followed by a memory probe requiring judgment of the item based on a randomly selected criterion (Fig. 2A). The logic of the design was that participants would not be able to anticipate what stimulus dimension would be probed, and so strategic recoding would not be advantageous. Each trial included a target presentation (1 s), a delay period (7 s), a probe presentation (5.5 s—on trials that also included a probe stimulus, the question was shown for 3 s followed by the stimulus for 2.5 s), a response feedback (0.5 s), and a blank screen (8 s), which preceded the next trial. Target stimuli were randomly drawn (without replacement) from the full set of object stimuli. All probes included a sentence presented in the form of a question that specified the criterion required for judgment of the memory item on the current trial. Some questions were followed by an image, word, or a pronounceable string of letters to be used for a comparison with the memory item. Judgment criteria were based on six different categories of stimulus characteristics, with nine total trials drawn from each category. Probe questions were pseudo-randomly selected from one of four possibilities created for each of six categories: affective (e.g., “How threatened are you by the item?”), episodic (e.g., “Have you seen the item recently?”), perceptual (e.g., “Is this new item the same size as the original item?”), phonological (e.g., “How many syllables are in the name of the item?”), semantic (e.g., “Is the item made of wood?”), and visual (e.g., “Is this new item a vertically reflected copy of the original item?”). The probe



**Figure 2.** Task procedures. (A) Phase 1: participants' memory of object stimuli was tested after a brief delay. (B) Phase 2: participants' memory of object stimuli was again tested, but during two separate delay periods in each trial, they were cued as to which dimension of the stimulus would be tested. (C) Phase 3: participants performed short-term recognition of a word, a pronounceable pseudoword, or line segments. They received explicit instructions to encode stimuli primarily in the intended representational format for each trial type ("semantic" for words, "verbal" for pseudowords, and "visual" for line segments). Phase 3 was performed at the end of the experiment to prevent carryover effects that could influence task performance in the previous two phases.

varied randomly across trials, which prevented participants from anticipating which stimulus dimension (e.g., its shape, color, name, etc.) would be interrogated at the end of each trial. Participants responded with a button press on a four-button response controller. During the practice session, participants learned the appropriate mappings of responses to the four buttons: for True/False probes, 1 = True and 4 = False; for probes requiring an assessment along a continuum, 1 = High, 2 = Medium-High, 3 = Medium-Low, and 4 = Low; for probes requiring a quantity judgment, each button represented its corresponding quantity (e.g., button 1 = "1 item", etc.).

### Phase 2: Cued Judgment

Participants performed 54 trials of a Delayed-Judgment task with retro-cueing modeled on a modified Sternberg task (Oberauer 2005; Lewis-Peacock et al. 2012) (Fig. 2B). The first half of each trial consisted of a target presentation (1 s), a delay period (7 s), a cue (6 s), a probe stimulus (1 s), and a response period (2 s). The second half consisted of a second cue, delay, probe, and response sequence (without a representation of the memory item), followed by response feedback (1 s) and a blank screen (8 s), which preceded the next trial. Target stimuli were randomly drawn (without replacement) from the full set of object stimuli. Cues indicated whether the verbal, semantic, or visual characteristics of the item in memory were relevant for an upcoming probe. Cues were conveyed by changing the color of the central fixation cross from white (on black background) to cyan, red, or yellow, respectively. On a random half of trials, the category cued as relevant for the first decision was also cued as relevant for the second decision ("Repeat" trials). The other half of trials ("Switch" trials) required a switch to one of the two previously irrelevant categories. On these trials, the domain was chosen randomly, but equally often, from the two alternatives. Importantly, this multiple-cue procedure created a scenario in which all three representational domains (verbal, visual, and semantic) were "potentially relevant" for the second decision in every trial.

Comparisons between the probe and the memory item varied for each category. For verbal cues, participants evaluated whether the vowel sound in a single-syllable pseudoword letter string also appeared anywhere in the name of the object held in memory. For semantic cues, participants evaluated whether a probe object (another picture drawn from the object stimulus set) would commonly appear near the target object in real-world situations. For visual cues,

participants evaluated whether the outline of a silhouetted image was identical to the outline of the object held in memory. Silhouette images were black-foreground/white-background versions of the original object pictures. Trials were configured such that there was a probability of 0.5 that the probe stimulus satisfied the criterion. Foils (to-be-rejected probes) for the three categories were single-syllable pseudowords with a non-matching vowel sound, images of familiar objects drawn from a different category than the target, and modified silhouette images of the target in which a small portion of the original silhouette had been removed.

### Phase 3: Category-Specific Delayed Recognition

Participants performed delayed recognition of a stimulus drawn from one of three categories—pronounceable pseudowords, real words, and line segments (Fig. 2C). This task is identical to the one used in Lewis-Peacock et al. (2012), Experiment 2, Phase 1. Memory probes required a domain-specific comparison for each category, creating a situation in which the most relevant dimension of the stimulus was verbal, semantic, or visual, respectively. Criteria for comparing the probe to the memory item were different for each stimulus category. A synonym judgment was required for words, a rhyme judgment of vowel sounds was required for pseudowords, and a visual orientation judgment was required for line segments. The comparison criteria described here were modeled after a rich literature highlighting dissociations between verbal and visual processes in STM (e.g., Baddeley and Hitch 1974), as well as more recent studies that have further dissociated verbal memory into semantic and phonological components (e.g., Crosson et al. 1999; Romani and Martin 1999; Haarmann and Usher 2001; Martin et al. 2003; Shivde and Thompson-Schill 2004). Foils (to-be-rejected probes) for the three categories were conceptually unrelated words, single-syllable pseudowords with a non-matching vowel sound, and line segments in which one of the segments differed in orientation by at least 30°. Participants performed 72 trials, with 24 trials drawn from each category. Each trial consisted of a category cue (2 s), a target presentation (0.5 s), a delay period (7.5 s), a probe presentation (0.5 s), a response period (1.5 s), and a blank screen (10 s) that preceded the next trial. Participants indicated with a Yes/No button press whether the probe stimulus matched the memory item according to a category-specific criterion. Trials were configured such that there was a probability of 0.5 that the probe stimulus satisfied the

criterion. The stimuli and task demands were designed to encourage domain-specific encoding in a primary dimension for each trial. That is, we attempted to elicit the short-term retention of information in a semantic (i.e., conceptual) form on trials that required a synonym judgment, in a verbal/phonological form on trials that required a rhyme judgment, and in a visuospatial form on trials that required an orientation judgment. Words were presented in white (on black background) to indicate that the stimulus was to be primarily encoded based on its semantic characteristics. Pseudowords were presented in cyan to indicate that the stimulus was to be primarily encoded based on its phonological characteristics. Line segments were always presented in white (on black background). Participants received explicit instructions to encode stimuli primarily in the intended representational format for each stimulus. This Phase 3 task was performed at the end of the experiment to avoid biasing participants into creating verbal, semantic, and visual representations of stimuli during earlier phases of the experiment.

### Data Collection and Preprocessing

All experiments were implemented with E-Prime software version 2.0 (Psychology Software Tools), and an Avotec goggle system (Avotec, Inc.) was used to display visual stimuli inside the scanner. Whole-brain images were acquired with a 3-T scanner (GE Signa VH.I). For all participants, we acquired high-resolution T1-weighted images ( $1 \times 1 \times 1$  mm). We used a gradient-echo, echo-planar sequence with ramp sampling (flip angle =  $60^\circ$ , echo time = 25 ms, FOV = 24 cm, and time repetition = 2000 ms) to acquire data sensitive to the blood oxygen level-dependent (BOLD) signal within a  $64 \times 64$  matrix (40 axial slices coplanar with the T1 acquisition,  $3.75 \times 3.75 \times 4$  mm). All task runs were preceded by 20 s of dummy pulses to achieve a steady state of tissue magnetization. Three blocks of the Phase 1 task were obtained, each consisting of 18 trials (3 trials per response category) lasting 6 min 56 s, for a total of 20 min 48 s in functional scans. Next, instructions were presented visually, and six practice trials were completed for the Phase 2 task (5 min). Any confusion regarding task instructions was resolved verbally via the scanner intercom system before continuing. Six blocks of the Phase 2 task were obtained, each consisting of nine trials lasting 5 min 44 s, for a total of 34 min 24 s in functional scans. Next, instructions and six practice trials were presented for the Phase 3 task (5 min). Again, any confusion regarding task instructions was resolved verbally via the scanner intercom system before continuing. Finally, four blocks of the Phase 3 task were obtained, each consisting of 18 trials (6 trials per stimulus category) lasting 6 min 56 s, for a total of 27 min 44 s in functional scans. Preprocessing of the functional data was done with the AFNI software package (Cox 1996) using the following preprocessing steps, in order: correction for slice time acquisition and rigid-body realignment to the first volume from the experimental task with 3dvolreg, removal of signal spikes with 3dDespike, removal of the mean from each voxel and linear and quadratic trends from within each run with 3dDetrend. Note that neither was spatial smoothing imposed nor were the data spatially transformed into a common atlas space prior to hypothesis testing. Rather, the data from each participant were analyzed in that participant's un-smoothed, native space. For classification analyses, a feature selection analysis of variance (ANOVA) was applied to the preprocessed images from the Phase 3 task to select those voxels whose activity varied significantly ( $P < 0.05$ ) between the four conditions (three stimulus categories + the inter-trial interval). The mean number of voxels passing feature selection was 12,972 (SD = 1463). Voxels from these masks served as input nodes to the pattern classifier for hypothesis testing in the independent data from Phase 1 and Phase 2.

### Multivariate Pattern Analysis

All fMRI pattern classification analyses (see Norman et al. 2006; Pereira et al. 2009 for reviews) were performed using the Princeton MVPA Toolbox in Matlab (downloadable from <http://www.pni.princeton.edu/mvpa>), using L2-penalized logistic regression. The L2 regularization term biases the algorithm to find a solution that minimizes the sum of the squared feature weights, thus reducing the likelihood of overfitting the training data. Logistic regression uses a parameter ( $\lambda$ ) that determines the impact of the regularization term. To set the

penalty  $\lambda$ , we explored how changing the penalty affected our ability to classify the Phase 3 data (using the cross-validation procedure described later). We used a value of  $\lambda = 50$  for all of our classifier analyses (this is the same value used in our prior, related study: Lewis-Peacock et al. 2012). As a final preprocessing step, all functional data were z-scored (separately for each run) prior to pattern analysis. Three data points from each trial of the Phase 3 task corresponding to the final 6 s of the delay period, at intervals of 2 s, were used to train the pattern classifiers. The classifier was trained to distinguish patterns of delay-period brain activity corresponding to the short-term retention of information encoded primarily in a verbal, semantic, or visual form. For a reference or baseline category, the classifier was also trained on data sampled from the 10-s inter-trial interval (baseline or "rest" period activity). A unique classifier was created for each participant and applied only to that participant's data. Regressors for the training data were shifted forward by 4 s to account for hemodynamic lag of the BOLD signal. We evaluated classification accuracy by using the method of k-fold cross-validation, i.e., training on three blocks of trials and testing on the novel fourth block. The blocks used for training were then rotated, and a new block of data was tested until all trials had been classified (note that feature selection was done separately, and only on the training blocks, for each iteration.) For each epoch of fMRI data, the classifier produced an estimate (from 0 to 1) of the extent to which the brain activity matched the pattern of activity corresponding to the four categories on which it had been trained. We refer to this estimate as "classifier evidence." Prediction accuracy was calculated as the proportion of fMRI epochs in which the classifier's strongest evidence corresponded to the correct category (e.g., the "visual" category for all delay-period epochs in trials with line segment stimuli). Finally, the pattern classifier was re-trained on all four blocks of Phase 3 data and applied to the independent data from Phases 1 and 2. The classifier assessed the extent to which domain-specific patterns of brain activity (learned from the Phase 3 data) could be identified during the delay period of Delayed-Judgment and Cued-Judgment tasks, respectively. Preprocessed fMRI data at intervals of 2 s were classified from every trial of these two tasks. Evidence for simultaneous STM retention of a stimulus in multiple representational formats would come from the classifier's identification of multiple domains of information during the memory delay of the Phase 1 Delayed-Judgment task. Evidence for flexible coding in STM would be indicated by the classifier's differentiation of brain activity for task-relevant and task-irrelevant stimulus dimensions during the memory delays of the Phase 2 Cued-Judgment task.

### Brain Mapping

A traditional mass-univariate analysis based on the GLM was performed on the Phase 3 data using AFNI's "3dDeconvolve." All trial events were modeled with boxcar regressors of different lengths: cue (1 s), target (0.5 s), delay (7.5 s), probe (2 s), and feedback (0.5 s). A third-order polynomial was used for the null hypothesis, and all basis functions for trial events were normalized to have amplitude of one. The GLM activation maps for each participant were transformed into standardized space with voxel dimensions of  $4 \text{ mm}^3$  using AFNI's @auto\_t1rc and then blurred with a full-width half-max of 8 mm using 3dmerge. Group data were analyzed using 3dttest++, which performed *t*-tests (with respect to baseline) of delay-period regressors from verbal, semantic, and visual trials, respectively. A clustering algorithm (NN level = 1; 20 voxels) was used to restrict selection of voxels to those with at least some degree of spatial contiguity. Three thresholded ( $P < 0.01$ , uncorrected) sets of voxels (including positive and negative activations, with respect to baseline) were extracted for each trial category. Finally, the activation results were mapped onto an inflated anatomical version of the N27 brain dataset (Holmes et al. 1998) using AFNI's surfacing mapping utility.

To assess the relative importance of different brain areas to the classification of the stimulus categories in our MVPA analyses, we determined, from a classifier trained using all brain voxels separately for each participant, which voxels were important for (correctly) identifying patterns of brain activity corresponding to each of the trained categories. We applied a modified version of the voxel importance formula from McDuff et al. (2009):  $\text{imp}_{ij} = 1000 \times w_{ij} \times \text{avg}_{ij}$ , where  $w_{ij}$  is the weight

between input unit  $i$  and output unit  $j$ , and  $avg_{ij}$  is the average activity of input  $i$  during the short-term retention of category  $j$ . Positive importance was assigned to a voxel whose average activity was positive (indicating that it was more active than usual), negative importance was assigned to a voxel whose average activity was negative (indicating that it was less active than usual), and voxels where the sign of  $w_{ij}$  differed from the sign of  $avg_{ij}$  (indicating a net negative contribution of that voxel to detecting that task state) were assigned an importance value of zero. Importance maps for the three categories were calculated separately for each participant, transformed into standardized space, averaged across all participants with *3dmerge*, thresholded at an absolute value of importance of 0.10, and mapped onto an inflated brain (as described earlier for the GLM voxel activation maps).

## Results

### Behavioral Results

#### Phase 1: Delayed Judgment

Responses on affective and episodic trials were subjective and therefore not scoreable (e.g., “Have you seen this item recently in the real world?” in the episodic condition; see Methods for more details). However, response times (RTs) indicated that participants were complying with the instructions for all trial types. Behavioral accuracy across the remaining four trial categories was 88% (SEM = 2%). The accuracies per condition, listed in descending order, were semantic (95%, SEM = 3%), perceptual (92%, SEM = 2%), visual (89%, SEM = 3%), and verbal (77%, SEM = 4%). The overall ANOVA on accuracy based on trial condition was significant ( $F_{5,66} = 356.7$ ,  $P < 0.0001$ ). Excluding the affective and episodic trials, follow-up pairwise comparisons revealed that participants were more accurate on semantic, perceptual, and visual trials compared with verbal trials ( $P < 0.0001$ ,  $P < 0.0001$ , and  $P < 0.05$ , respectively; Bonferroni corrected). The mean RT on the Phase 1 task was 1803 ms (SEM = 71 ms). The RTs per condition, listed from fastest to slowest responses, were visual (1210 ms, SEM = 36 ms), perceptual (1302 ms, SEM = 43 ms), semantic (1432 ms, SEM = 73 ms), episodic (2134 ms, SEM = 121 ms), affective (2160 ms, SEM = 117 ms), and verbal (2580 ms, SEM = 122 ms). The overall ANOVA on RTs, as a function of trial condition, was significant ( $F_{5,66} = 36.6$ ,  $P < 0.0001$ ), and follow-up pairwise comparisons revealed significant differences between all pairs except affective-episodic, perceptual-semantic, perceptual-visual, and visual-semantic (all  $P$ s < 0.05). Together, these results indicate that participants were performing well on the task and were complying with the task instructions.

#### Phase 2: Cued Judgment

The mean accuracy and RT for the first probe in the Phase 2 task were 78% (SEM = 2%) and 1286 ms (SEM = 33 ms), respectively. The mean accuracy and RT for the second probe were 67% (SEM = 2%) and 1236 ms (SEM = 32), respectively. One-way ANOVAs of accuracy and RTs based on trial type (Switch and Repeat) revealed a significant difference on probe2 accuracy between the trial types (Repeat: 74% (SEM = 3%); Switch: 64% (SEM = 2%);  $F_{1,106} = 4.4$ ,  $P < 0.05$ ). All other one-way ANOVA results were insignificant.

A two-way ANOVA of accuracy based on trial type (Switch and Repeat) and cue1 type (verbal, semantic, and visual) revealed a significant main effect of cue1 type ( $F_{2,102} = 5.1$ ,  $P < 0.001$ ), and a significant interaction between trial type and cue1 type ( $F_{2,102} = 3.9$ ,  $P < 0.05$ ). Follow-up pairwise

comparisons revealed that participants were more accurate ( $P < 0.05$ , Bonferroni corrected) on the first probe for semantic trials (85%, SEM = 3%) compared with verbal trials (73%, SEM = 3%). A two-way ANOVA of RT based on trial type and cue1 type revealed no significant main effects or interaction (all  $P$ s > 0.05). A two-way ANOVA of accuracy based on trial type and cue2 type revealed significant main effects for trial type ( $F_{1,102} = 4.8$ ,  $P < 0.05$ ) and for cue2 type ( $F_{2,102} = 4.0$ ,  $P < 0.05$ ), but the interaction was not significant. Follow-up pairwise comparisons revealed that participants were more accurate ( $P < 0.05$ , Bonferroni corrected) on the second probe for semantic trials (79%, SEM = 3%) compared with verbal trials (62%, SEM = 4%). A two-way ANOVA on RT based on trial type (Switch and Repeat) and cue1 type (verbal, semantic, and visual) revealed no significant main effects or interaction (all  $P$ s > 0.05). A two-way ANOVA on RT based on trial type and cue2 type revealed no significant main effects or interaction (all  $P$ s > 0.05).

Performance on this retrocuing task was relatively worse compared with our previous task (Lewis-Peacock et al. 2012, Experiment 2, Phase 2), which used two unique memory items from different categories on each trial (the stimuli were words, pseudowords, and line segments, identical to the stimuli used in Phase 3 in the present study). In this previous task, the mean accuracy and RT across both probes in each trial were 91% (SEM = 1%) and 936 ms (SEM = 10 ms), respectively; participants in that task were significantly more accurate ( $F_{1,19} = 58.5$ ,  $P < 0.0001$ ) and faster to respond to the memory probes ( $F_{1,19} = 11.3$ ,  $P < 0.001$ ) than the participants in the present study. This reduction in performance may be explained by an increase in cognitive demands of the present task compared with the previous task. Here, participants were interrogated, after a brief delay, on one of “three” possible features of an object (verbal, semantic, or visual) during each memory probe, whereas in the previous task, participants were required to make a delayed judgment on one of “two” possible features, and each of these features belonged to a different memory item (e.g., on a word/pseudoword trial, a semantic probe would only apply to the word and a verbal probe would only apply to the pseudoword). It appears from the present results that making delayed judgments about one of three possible features of a single item is more difficult than making similar judgments about one of two different items being held in memory. We will return to this point when interpreting the neural results for this study.

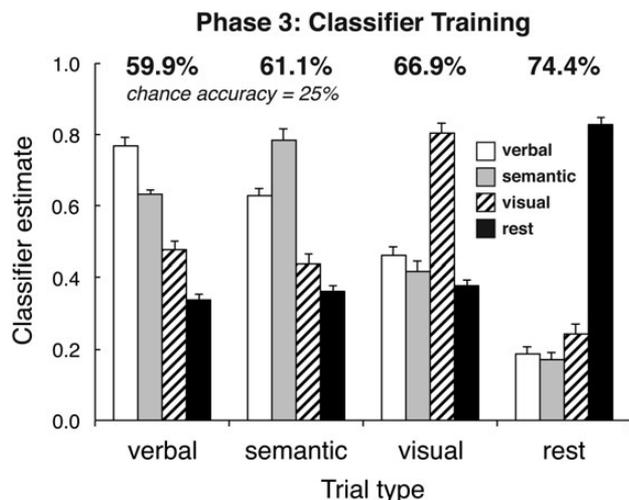
#### Phase 3: Category-Specific Delayed Recognition

Behavioral accuracy on the Phase 3 task was 94% (SEM = 1%). An ANOVA on accuracy for the three types of trials (words, pseudowords, and line segments) revealed a significant overall effect ( $F_{2,33} = 6.01$ ,  $P < 0.001$ ). Follow-up pairwise comparisons revealed that participants were more accurate on word trials (98%, SEM = 1%) compared with pseudoword trials (90%, SEM = 2%;  $P < 0.005$ , Bonferroni corrected); no other comparison was statistically significant. The mean RT for this task was 865 ms (SEM = 21 ms). An ANOVA on RTs for the three types of trials revealed a significant overall effect ( $F_{2,33} = 4.27$ ,  $P < 0.05$ ). Follow-up pairwise comparisons revealed that participants were faster to respond on visual trials (794 ms, SEM = 29 ms) compared with word trials (935 ms, SEM = 40 ms),  $P < 0.05$ , Bonferroni corrected; no other comparison was statistically significant. Performance on this task was comparable ( $P$ s > 0.15 for both accuracy and RT) with that from another group of

participants who performed the same task in our previous experiment (Lewis-Peacock et al. 2012, Experiment 2, Phase 1).

### fMRI Classifier Training (Phase 3)

For all participants, cross-validation classification analyses of brain data from the Phase 3 category-specific delayed-recognition task demonstrated that brain activity from the delay period was reliably classified as matching the stimulus dimension being tested in the trial (Fig. 3; all  $P$ s < 0.001 based on chance-level accuracy of 25%). This indicated that the classifier successfully differentiated visual from verbal (Baddeley 1986), from



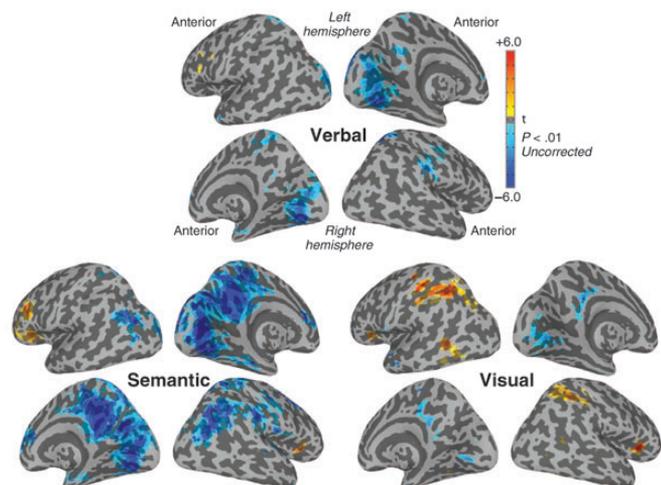
**Figure 3.** Classifier training performance. On each trial in Phase 3, the classifier generated an evidence score (ranging from 0 to 1) reflecting the extent to which the pattern of activity on that trial resembled each of the four categories to which it had been trained (verbal, semantic, visual, rest). The bar graphs display the group-mean classifier estimates as a function of trial type (“verbal”: pseudoword trials; “semantic”: word trials; “visual”: line segment trials; “rest”: inter-trial intervals). Error bars are SEM. The percentages at the top of the plots indicate the group-mean classifier accuracy for each trial type (chance was 25%). The classifiers performed above chance for all categories ( $P < 0.001$ ).

semantic (Martin et al. 1994; Haarmann and Usher 2001; Martin et al. 2003; Martin and He 2004; Shivde and Thompson-Schill 2004; Cameron et al. 2005) STM, and all three from resting activity as classified from the inter-trial interval, and could therefore be used to reliably decode brain activity from the other phases. Delay-period voxel activation maps, derived from mass-univariate analyses, show distinct activation patterns for each category (Fig. 4A). In addition, the pattern classifiers trained on the delay-period data from this task were analyzed to estimate the extent to which each voxel contributed to the classifier’s identification of each of the four categories (including “rest” activity from the inter-trial interval). This analysis confirmed that multiple, distributed brain regions contributed to the classification of each category (Fig. 4B). These results replicate our previous findings in which a different group of participants performed the same task (Lewis-Peacock and Postle 2012). We next proceeded to decode data from Phase 1 and Phase 2 with the classifiers so trained on the Phase 3 data.

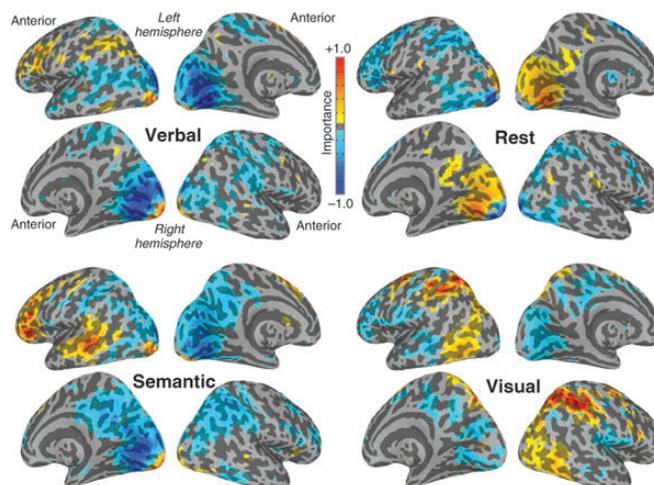
### fMRI Classifier Decoding of Delayed Judgments (Phase 1)

This was the first task that participants performed, and it therefore reflects performance when participants were naïve to our interest in mental coding formats. Additionally, the use of six randomly occurring probe types was intended to discourage the strategic adoption of an explicit coding strategy during the memory interval. Group-averaged classification results identified a progression of representational formats supporting STM in the Phase 1 task (Fig. 5). Continuous decoding of brain activity from the beginning of each trial indicated that, on average, an initial visual representation of the object stimulus was quickly augmented with both verbal and semantic recordings of the stimulus. All three domains of representation remained active throughout the delay period, with classifier evidence for the verbal code quantitatively surpassing those for the visual and semantic codes by the end of the delay period (at  $t = 8$  s). These results indicate that when participants were unable to anticipate how a memory item would be tested,

#### A Delay-period activation maps

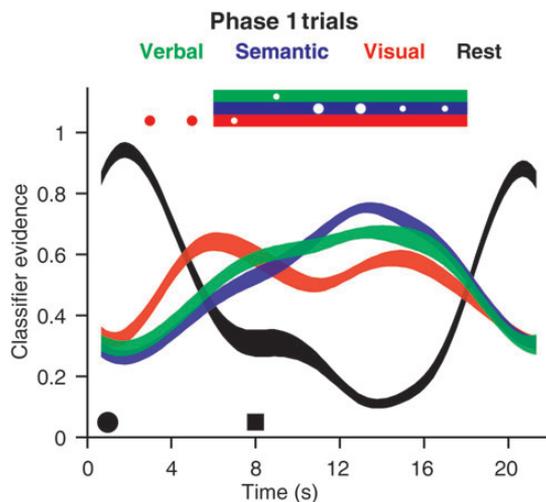


#### B Delay-period classifier importance maps



**Figure 4.** Brain maps. (A) Univariate delay-period voxel activation maps based on group data ( $P < 0.01$ , uncorrected) are shown for verbal, semantic, and visual trials from Phase 3. (B) Multivariate delay-period classifier importance maps, averaged across participants (absolute value importance threshold = 0.10), are shown for verbal, semantic, and visual trials, and rest activity from the inter-trial intervals of Phase 3.

they encoded and retained multiple dimensions of stimulus-related information in STM, eventually favoring a verbal recoding of the visual stimulus. This pattern is consistent with what might be expected from the visual-to-phonological recoding operation reviewed in the Introduction. An additional insight into this operation, afforded here by the increased sensitivity of MVPA applied to fMRI data (Norman et al. 2006), is that verbal recoding seems not to entail the “replacement” of an item’s initial visual representation with a verbal one, but, instead, the “supplementation” of a visual representation with a verbal and semantic recoding of the memory item. In response to the memory probe, estimates for all three domains



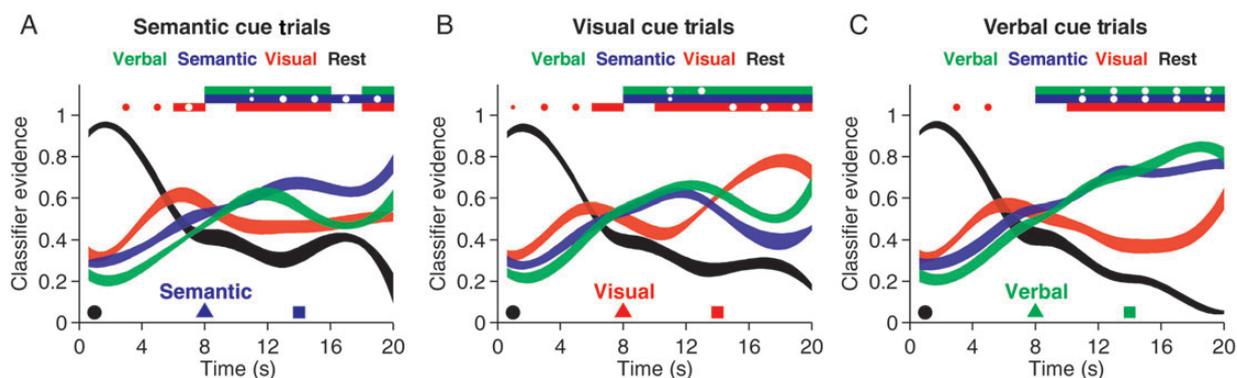
**Figure 5.** Default encoding in STM. The plot shows group-averaged, trial-averaged classifier evidence time courses for verbal, semantic, visual, and rest classifier evidence during the Phase 1 task. Data for each category of activity are shown as ribbons, interpolated across the 11 discrete fMRI data points for each trial, whose thicknesses indicate  $\pm 1$  SEM. The horizontal axis shows the trial timeline and indicates the onset of the target (black circle, 0 s) and probe (black square, 8 s). Statistical comparisons between classifier estimates focused on the within-subject difference between a given category’s estimate and the estimate of rest (baseline) activity. Statistical significance ( $P < 0.05$ ) is indicated by horizontal lines at the top of each graph colored accordingly for each category. White circles inside one of these horizontal lines (or colored circles outside the lines) indicate 2-s intervals for which the estimate of that stimulus category was reliably higher than the “average estimates” of the other two stimulus-related categories (small circle:  $P < 0.05$ ; large circle;  $P < 0.0045$ , Bonferroni corrected).

of representation rose sharply, with the active representation of semantic information quantitatively greater than the representation of visual and verbal information. Note that this pattern results from collapsing across questions probing each of six different stimulus dimensions. Further, our methods do not permit us to sort out the extent to which probe-evoked activity reflected memory-related processes versus the perceptual processing of the memory probe.

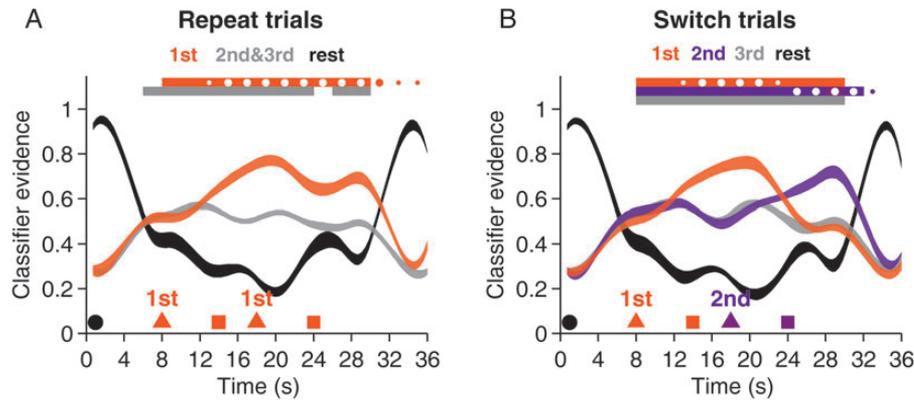
#### fMRI Classifier Decoding of Delayed Judgments (Phase 2)

The first of two analyses performed on the Phase 2 Cued-Judgment task focused on the first portion of the trial, to isolate the effect of the initial retrocue that informed participants whether the impending probe would require a judgment about the semantic, visual, or verbal properties of the target stimulus (Fig. 6). Across all trial types, the initial target-evoked response emphasized the visual properties of the stimulus, findings that replicate the results from Phase 1. With the onset of the retrocue, however, brain activity reconfigured itself to emphasize the cued dimension. For example, after a retrocue indicating that the semantic properties of the target would be tested, classifier estimates of semantic coding increased sharply, whereas estimates of both verbal and visual coding dropped to baseline (Fig. 6A). Similarly, following a retrocue indicating that the visual properties of the target would be tested, classifier estimates of visual coding increased sharply, and estimates of verbal and semantic coding declined (although these estimates did not fall to baseline; Fig. 6B). Finally, for trials on which the cue indicated that the verbal properties of the target would be tested, classifier estimates of verbal coding increased sharply (Fig. 6C). Note that in this latter case, classifier estimates of semantic coding also increased at the same rate. The implications of this for the phonological recoding hypothesis will be taken up in the Discussion section.

The aggregate effect of retrocuing, collapsed across the three stimulus dimensions, can be seen in the second analysis of these data, as illustrated in Figure 7. Inspection of the first half of both Repeat and Switch trials (i.e., from  $t = 0$  to 18 s) indicates that, upon delivery of the retrocue, there was a transient increase in MVPA evidence for all stimulus dimensions, but that whereas evidence for the neural representation of the cued dimension continued to climb throughout the remainder of the delay period, neural evidence for the two uncued



**Figure 6.** Strategic recoding in STM. In Phase 2, participants were cued as to which stimulus dimension would be tested. Group-averaged data are shown separately for the first half of trials in which the semantic (A), visual (B), or verbal (C) stimulus dimension was cued as relevant for the first memory probe. The horizontal axes show the trial timeline and indicate the onset of the target (black circle, 0 s), the first cue (colored triangle, 8 s), and the first probe (colored square, 14 s). The color and label of the cue indicates the stimulus dimension tested by the first memory probe. Data for each category of activity are shown as ribbons, interpolated across the 11 discrete fMRI data points for each trial, whose thicknesses indicate  $\pm 1$  SEM. Statistical comparisons between the classifier estimates are indicated at the top of the graphs using the conventions explained in Fig. 5.



**Figure 7.** Flexible coding in STM. Data for trials in which the second cue was the same as the first cue (Repeat trials, *A*) or different from the first cue (Switch trials, *B*) are shown, collapsed across all trial types and all participants. All graphical conventions are as described in Fig. 5, with a few modifications. Data from the entire trial are shown with relabeled estimates from each trial. Repeat trials include estimates for the first cued dimension (first, orange), the average of the two uncued dimensions (second and third, grey), and the rest activity from the inter-trial intervals (rest, black). Switch trials include estimates for the first cued dimension (first, orange), the second cued dimension (second, purple), the uncued third dimension (third, grey), and the rest activity from the inter-trial intervals (rest, black). The horizontal axes show the trial timeline and indicate the onset of the target (black circle, 0 s), the first and second cues (colored triangles, 8 s and 18 s), and the first and second probes (colored squares, 14 s and 24 s). Statistical significance ( $P < 0.05$ ) compared with baseline is indicated by horizontal lines at the top of each graph colored accordingly for each category. White circles inside one of these horizontal lines (or colored circles outside the lines) indicate 2-s intervals for which the estimate of that stimulus category was reliably higher than the “average estimates” of the other two stimulus-related categories (small circle:  $P < 0.05$ ; large circle;  $P < 0.0023$ , Bonferroni corrected).

dimensions returned to an intermediate level, statistically above baseline, but statistically below the cued dimension (the implication of this finding for models of multiple levels of activation within STM will also be considered in the Discussion.)

In addition to the preceding observations, another reason for performing this second analysis was to obtain enough statistical power to evaluate the effect of “switch” cues—the 50% of second retrocues that indicated that the second memory probe would probe a different stimulus dimension than had the first probe on that trial. The effect of switch cues is seen in the second half of the trial-averaged data presented in Figure 7*B* (i.e., from  $t = 18$  to 36 s): switch cues prompted a change in neural state such that neural evidence for the no-longer-relevant stimulus dimension dropped to baseline, and evidence for the newly relevant stimulus dimension remained elevated for the remainder of the delay period, at a level statistically greater than those of the two uncued stimulus dimensions and of baseline. This result sits in stark contrast to Repeat trials, in which the second cue prompted the sustained elevation of the previously cued stimulus dimension for the remainder of the trial. Together, these results replicate and extend our previous findings regarding the selective effects of retrocues on the informational content of delay-period activity (Lewis-Peacock et al. 2012; Lewis-Peacock and Postle 2012, LaRocque et al. 2013), with important differences concerning the neural fate of the uncued memory items (stimulus features, in this case), the implications of which will be addressed in the following section.

## Discussion

The idea of the “mental code” has a venerable history in cognitive psychology. Behavioral research on STM in the 1960s and 1970s provided evidence for an important principle of the coding of mental representations, and two corollaries. The principle of multiple encoding is that we hold information “in mind” in as many representational formats as are afforded by the stimulus. That is, mental representations are multidimensional. One corollary, phonological recoding, is that visually presented information is obligatorily recoded into a phonological/articulatory

code. This confers many advantages for STM, one being the increase in STM capacity from the roughly  $3 \pm 1$  associated with visual STM (Cowan 2001; Anderson et al. 2011) to  $7 \pm 2$  that is enabled by chunking and covert articulatory rehearsal (Baddeley 1986; Acheson and MacDonald 2009). A second advantage to phonological recoding is that it maintains the information that is in STM in a format that is more readily converted to verbal output. A second corollary of multiple encoding is strategic recoding: when we know the dimension of the mental representation (i.e., the mental code) that we will need to interrogate for an upcoming decision, we can prioritize that dimension (Tversky 1973). In recent years, conceptual and technical advances in neuroimaging have allowed the investigation of the neural bases of these principles. Conceptually, the discovery of a default mode of activity in the brain (Raichle et al. 2001), which predominates during the inter-trial interval of challenging cognitive tasks, has prompted renewed interest in the dynamics of uncontrolled thought. Technically, the adoption of multivariate techniques has afforded levels of measurement sensitivity and selectivity that now support investigations of the neural bases of mental codes, including their temporal dynamics.

### Neural Evidence for Multiple Mental Codes, and Their Strategic Control

In the present study, we leveraged multivariate pattern classification of fMRI data to decode the moment-to-moment information content of brain activity during tests of STM. Our results confirmed that a progression of representational formats is involved in the encoding and short-term retention of information about visually presented objects. Phase 1 revealed that following the presentation of a to-be-remembered object stimulus, classification of brain activity patterns indicated that visual, verbal, and semantic dimensions of the stimulus were encoded into STM. This is a result predicted by the multiple encoding hypothesis (Wickens 1973) and consistent with cognitive models describing STM as the re-activation of long-term memory representations (Cowan 1995; Oberauer 2002; Lewis-Peacock and Postle 2008). Phase 2 indicated that participants could exert strategic control over these codes, dynamically reconfiguring them to align with the

impending probe. In so doing, it corroborates decades-old behavioral evidence for the use of strategic recoding of information in STM (Tversky 1973) and bolsters recent multivariate evidence for the retuning and expansion of the semantic processing of visual information for attended stimuli compared with unattended stimuli while participants view natural movies (Cukur et al. 2013).

### ***The Consequences of Prioritizing one Stimulus Dimension in STM***

One important question for this study was how the attentional prioritization of one dimension of a stimulus (via retrocuing) would influence the neural representation of uncued dimensions. Previous MVPA of fMRI and EEG data indicate that classifier evidence for the initially uncued stimulus from a two-item memory set drops to baseline, but that its classifier evidence will return to an elevated level if, later in the trial, the second retrocue indicates that it will be relevant for evaluating the final memory probe (Lewis-Peacock et al. 2012; LaRocque et al. 2013). This has led us to speculate that classifier evidence for the active neural representation of a stimulus may more closely reflect that item's status in the focus of attention rather than its retention in STM (LaRocque et al. 2014; Postle, in press). The results from Phase 2 of the present study, however, may seem to contradict this idea. This is because, prior to the first memory probe, classifier evidence for the uncued stimulus dimensions following the first retrocue dropped to a level that, although lower than that of the attended stimulus dimension, was nonetheless "higher" than that of the empirically determined baseline ("rest" activity from the inter-trial interval). This sustained activation of uncued (and thus task-irrelevant) information about the memory item may have reduced the overall availability of attentional resources which, in turn, may have contributed to the reduction in performance on this task compared with our previous work (Lewis-Peacock et al. 2012). In this previous task, participants responded more quickly and accurately to the probes on individual memory items, and we found that the uncued, irrelevant memory items became neurally deactivated following the cue of the relevant item. The pattern following the second retrocue in Phase 2, however, was consistent with our previous results, in that classifier evidence for uncued stimulus dimensions dropped to baseline (Fig. 7B). Thus, the present findings suggest a divergence of the effects of the first retrocue on neural activation as indexed by MVPA classifier evidence, depending on whether it is cuing one from among two target stimuli or one dimension from among the many that are intrinsic to a single-target stimulus. To account for this discrepancy, we offer an account that is grounded in research on object- and feature-based attention.

One well-established property of object-based attention is that, in a multi-item scene, selection of one stimulus confers an attentional advantage to that stimulus, and a commensurate disadvantage to unselected items, the well-established principle of biased competition (Desimone and Duncan 1995). At the neural level, this manifests as a boost in the efficacy of the neural representation of the selected item and a weakening of that of the unselected item, as indexed, for example, in extracellular recordings (e.g., Chelazzi et al. 2001) and in multivariate analyses of EEG data (Garcia et al. 2013). Thus, it may be that the loss of classifier evidence for an uncued memory item in two-item arrays reflects that item's "loss" of the competition

for attentional prioritization. When attention selects one feature of an object, however, the situation is more nuanced. Evidence from experimental psychology indicates that attentional selection of one feature of an object automatically "spreads" to encompass the remaining elements of that object, such that they also benefit from attentional prioritization relative to other unselected objects (Duncan 1984; Vecera et al. 2000; Driver 2001). At the neural level, however, there is evidence for competition among features belonging to the same stimulus, when only one is prioritized for attention (reviewed by Maunsell and Treue 2006). Thus, it may be that the "intermediate" neural activation status that we observed for uncued stimulus dimensions after the first retrocue (i.e., classifier evidence that was statistically below that of the cued dimension, yet still above baseline) reflects the counteracting effects of the putative "spread" of attention to the entirety of an object versus biased competition at the level of feature representation.

Regardless of whether or not this "object- and feature-based" account of the present results can be supported by future work, the effects of the "second" retrocue in Phase 2 seem more straightforward, indicating that participants are able to fully suppress the neural representation of uncued dimensions of a remembered stimulus when they know that these dimensions are no longer relevant for performance on the trial. Thus, contrary to the intuition of some, it may be possible to focus attention on one dimension of a stimulus to the exclusion of others.

### ***Evidence for Phonological Recoding?***

On the question of phonological recoding, the present findings are more equivocal. Phase 1 may offer the purest data with which to address this question, because it was carried out when participants were least likely to be biased to adopt any particular strategy—they had not yet received any explicit instructions about how to mentally represent the stimuli, and they could not know on any given trial which dimension of the target stimulus was going to be probed. With this in mind, the Phase 1 data do reveal a pattern that is consistent with a phonological recoding account: following the predominantly visual target-evoked response, phonological and semantic patterns of activity emerged, supplementing the existing visual representation, and all three codes persisted across the duration of the delay period (Fig. 5). This pattern was not evident, however, in the verbally cued trials from Phase 2 (Fig. 6C). For these, although the retrocue did, indeed, trigger a sharp rise in evidence for a verbal representation, it also triggered an equally sharp rise in evidence for a semantic representation. One possible explanation for the discrepancy between these aspects of the Phase 1 and Phase 2 studies is that the latter may reflect a greater influence of volitional control over mental representation in STM, because trials of all three types were interleaved during this phase, and, on each trial, participants were, in effect, "told how to think". Nonetheless, until the results from Phase 1 are replicated or refuted in a future study, speculation about implication of these results for phonological recoding must remain qualified.

In summary, the present results reveal the neural dynamics underlying the creation and regulation of a mental representation and illustrate the flexibility with which the brain can recode information in response to environmental exigencies. Further, these analyses illustrate a novel way in which

measurements of the brain can be exploited to gain insight about the structure of the mind.

## Funding

This research was funded by the National Institutes of Mental Health: R01 MH064498 (BP) and F31 MH085444 (JL-P).

## References

- Acheson DJ, MacDonald MC. 2009. Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychol Bull.* 135(1):50–68.
- Anderson DE, Vogel EK, Awh E. 2011. Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *J Neurosci.* 31(3):1128–1138.
- Baddeley AD. 1986. *Working Memory*. London: Oxford University Press.
- Baddeley AD, Hitch GJ. 1974. Working memory. In: Bower GH, editor. *The Psychology of Learning and Motivation*, vol. 8. New York: Academic Press. p. 47–89.
- Cameron KA, Haarmann HJ, Grafman J, Ruchkin DS. 2005. Long-term memory is the representational basis for semantic verbal short-term memory. *Psychophysiology.* 42(6):643–653.
- Camos V, Mora G, Oberauer K. 2011. Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Mem Cognit.* 39(2):231–244.
- Chelazzi L, Miller EK, Duncan J, Desimone R. 2001. Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb Cortex.* 11(8):761.
- Chiu Y-C, Esterman M, Han Y, Rosen H, Yantis S. 2011. Decoding task-based attentional modulation during face categorization. *J Cognit Neurosci.* 23(5):1198–1204.
- Conrad R, Hull AJ. 1964. Information, acoustic confusion and memory span. *Br J Psychol.* 55:429–432.
- Cowan N. 1995. *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.
- Cowan N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci.* 24(3): 87–185.
- Cox RW. 1996. Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.* 29:162–173.
- Crosson B, Rao SM, Woodley SJ, Rosen AC, Bobholz JA, Mayer A, Stein EA. 1999. Mapping of semantic, phonological, and orthographic verbal working memory in normal adults with functional magnetic resonance imaging. *Neuropsychology.* 13(2):171–187.
- Cukur T, Nishimoto S, Huth AG, Gallant JL. 2013. Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci.* 16(6):763–770.
- Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Ann Rev Neurosci.* 18(1):193–222.
- Driver J. 2001. A selective review of selective attention research from the past century. *Br J Psychol.* 92(1):53–78.
- Duncan J. 1984. Selective attention and the organization of visual information. *J Experim Psychol General.* 113(4):501–517.
- Garcia JO, Srinivasan R, Serences JT. 2013. Near-real-time feature-selective modulations in human cortex. *Cur Biol.* 23(6):515–522.
- Haarmann H, Usher M. 2001. Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bull Rev.* 8(3):568–578.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 293(5539):2425–2430.
- Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci.* 7(7):523–534.
- Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. 1998. Enhancement of magnetic resonance images using registration for signal averaging. *Comput Assist Tomography.* 22(2):324–333.
- Johnson-Laird PN. 1995. Mental models, deductive reasoning, and the brain. In: Gazzaniga MS, editor. *The Cognitive Neurosciences*. Cambridge, MA: MIT Press. p. 999–1008.
- Jonides JJ. 1995. Working memory and thinking. In: Smith EE, Osherson DN, editors. *Invitation to Cognitive Science: Thinking*, 2 ed, Vol. 3. Cambridge, MA: MIT Press. p. 215–265.
- Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci.* 8(5):679–685.
- 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(1):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 SH, Kravitz DJ, Baker CI. 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci.* 16(8):997–999.
- 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 Cognit 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.
- Martin RC, He T. 2004. Semantic short-term memory and its role in sentence processing: a replication. *Brain Lang.* 89(1):76–82.
- Martin RC, Shelton JR, Yaffee LS. 1994. Language processing and working memory: neuropsychological evidence for separate phonological and semantic capacities. *J Mem Lang.* 33(1):83–111.
- Martin RC, Wu D, Freedman M, Jackson EF, Lesch M. 2003. An event-related fmri investigation of phonological versus semantic short-term memory. *J Neurolinguistics.* 16(4–5):341–360.
- Maunsell JH, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci.* 29(6):317–322.
- McDuff SG, Frankel HC, Norman KA. 2009. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *J Neurosci.* 29(2):508–516.
- Miller GA. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev.* 63:81–97.
- Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: Multi-voxel pattern analysis of fmri data. *Trends Cognit Sci.* 10(9):424–430.
- Oberauer K. 2002. Access to information in working memory: exploring the focus of attention. *J Experim Psychol Learn Mem Cognit.* 28:411–421.
- Oberauer K. 2005. Control of the contents of working memory—a comparison of two paradigms and two age groups. *J Experim Psychol Learn Mem Cognit.* 31(4):714–728.
- Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fmri: a tutorial overview. *NeuroImage.* 45(1 Suppl):S199–S209.
- Postle BR. (in press). Activation and information in working memory research. In: Duarte A, Barense M, Addis DR, editors. *The Wiley-Blackwell Handbook on the Cognitive Neuroscience of Memory* (Wiley-Blackwell), 897–901.
- Postle BR, D'Esposito M, Corkin S. 2005. Effects of verbal and non-verbal interference on spatial and object visual working memory. *Mem Cognit.* 33(2):203–212.
- Postle BR, Hamidi M. 2007. Nonvisual codes and nonvisual brain areas support visual working memory. *Cereb Cortex.* 17(9):2151–2162.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. *Proc Natl Acad Sci.* 98(2):676–682.
- Riggall AC, Postle BR. 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J Neurosci.* 32(38):12990–8.
- Romani C, Martin R. 1999. A deficit in the short-term retention of lexical-semantic information: Forgetting words but remembering a story. *J Experim Psychol General.* 128(1):56–77.
- Shivde GS, Thompson-Schill SL. 2004. Dissociating semantic and phonological maintenance using fmri. *Cognit Affect Behav Neurosci.* 4:10–19.
- Shulman HG. 1971. Similarity effects in short-term memory. *Psychol Bull.* 75(6):399–415.

Simons DJ. 1996. In sight, out of mind: when object representations fail. *Psychol Sci.* 7(5):301–305.

Tversky B. 1973. Encoding processes in recognition and recall. *Cognit Psychol.* 5(3):275–287.

Vecera SP, Behrmann M, McGoldrick J. 2000. Selective attention to the parts of an object. *Psychonom Bull Rev.* 7(2):301–308.

Wickens DD. 1973. Some characteristics of word encoding. *Mem Cognit.* 1(4):485–490.