Decoding Attended Information in Short-term Memory: An EEG Study

Joshua J. LaRocque[1], **Jarrod A. Lewis-Peacock**[2], **Andrew T. Drysdale**[3], **Klaus Oberauer**[4], and **Bradley R. Postle**[3,5]

[1]Neuroscience Training Program and Medical Scientist Training Program, University of Wisconsin-Madison, USA [2]Princeton Neuroscience Institute and Department of Psychology, Princeton University, USA [3]Department of Psychiatry, University of Wisconsin-Madison, USA [4]Department of Psychology, University of Zurich, Switzerland [5]Department of Psychology, University of Wisconsin-Madison, USA

Abstract

For decades, it has been assumed that sustained, elevated neural activity – the so-called active trace – is the neural correlate of the short-term retention of information. However, a recent functional magnetic resonance imaging (fMRI) study has suggested that this activity may be more related to attention than retention. Specifically, a multivariate pattern analysis (MVPA) failed to find evidence that information that was outside the focus of attention, but nonetheless in short-term memory (STM), was retained in an active state. Here, we replicate and extend this finding by querying the neural signatures of attended vs. unattended information within STM with electroencephalography (EEG), a method sensitive to oscillatory neural activity. We demonstrate that in the delay-period EEG activity, there is information only about memory items which are also in the focus of attention. Information about items outside the focus of attention is not present. This result converges with the fMRI findings to suggest that, contrary to conventional wisdom, an active memory trace may be unnecessary for the short-term retention of information.

Introduction

The short-term retention of information no longer present in the environment, the core feature of short-term memory (STM) and working memory (WM), is central to much of human behavior. Decades of theoretical (Baddeley & Hitch, 1974; Hebb, 1949) and physiological (Curtis & D'Esposito, 2003; Fuster, 1973; Vogel, McCollough, & Machizawa, 2005) accounts have posited that short-term retention is accomplished by sustained, elevated neuronal activity (i.e., an active trace). It has recently been noted, however, that the vast majority of physiological studies of STM and WM have confounded memory with attention (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012). That is, in most tests of STM that afford the isolation of delay-period activity, the to-be-remembered information is also the most behaviorally relevant information, and is therefore likely to also be the focus of attention. This raises the possibility that sustained activity may relate more directly to selective attention than to STM, per se.

One class of models that explicitly considers the relationship between attention and memory is those that explain STM and WM as emerging from an interaction between attention and long-term memory (Cowan, 1988; Oberauer, 2002). These "activated long-term memory" models posit multiple states of activation, including, variously, a capacity-limited focus of attention, a region of direct access, and a broader pool of temporarily activated representations, all nested within the immense network of latently stored LTM. These models dissociate the retention of information – which can be accomplished in any of the

activated states of long-term memory – from attention to information – which is a capacity-limited resource that can be applied only to a small subset of highly activated representations. A recent study (Lewis-Peacock et al., 2012) explored the neural correlates of these different hypothesized states of activation by acquiring functional magnetic resonance imaging (fMRI) data while subjects performed a delayed-recognition task that allowed for the dissociation of attended memory items (AMI) from unattended memory items (UMI). In order to assay the neural signal underlying these two states within STM, multivariate pattern analysis (MVPA) was used to decode the information content of the delay-period activity. MVPA is a highly sensitive analytical technique which permits information to be extracted from high dimensional data sets such as fMRI (Pereira, Mitchell, & Botvinick, 2009; Polyn, Natu, Cohen, & Norman, 2005). The authors trained pattern classifiers to distinguish the category of remembered items in a single-item delayed-recognition task. They then applied these classifiers to the delay period of a two-item delayed-recognition task. On each trial, the two items being remembered were drawn from different categories, which permitted separate readouts of category-specific activity corresponding to each item. The item that was to be probed first was indicated by a retrocue; this served as an experimental manipulation of attention because it allowed participants to focus their attention on the relevant memory item in anticipation of the probe. MVPA of the delay-period activity only found evidence that the cued item was held in an active state. Evidence for an active representation of the UMI returned to baseline levels, although it could quickly be reactivated if it was cued during the second half of the trial. Thus, despite the apparent loss of sustained activity, the UMI was nonetheless remembered after a brief delay. These results demonstrated that the retention of information in STM may not require an active trace. Unlike previous neuroimaging studies that have used a similar approach (Lepsien & Nobre, 2007), the authors applied an information-based analysis that afforded direct inference about the state of activation of representations being retained during the delay period. This is a critical difference, because traditional univariate analyses that focus on changes in signal intensity provide information about the level of activity at an anatomical area (either in the brain or on the scalp), but not about the representational content of that activity. To induce the latter from the former can be highly problematic, as has been shown theoretically (Poldrack, 2006), and empirically (Lewis-Peacock & Postle, 2012). (Specifically, Lewis-Peacock and Postle (2012) have demonstrated that activity in voxels that have been determined to be "category specific" by conventional definition can carry information about which of two other categories is in the focus of attention. Thus, signal intensity-based analysis of delay-period activity can yield misleading conclusions about the activation state of attended and unattended items in STM.)

The implications of the Lewis-Peacock et al. (2012) findings are potentially very broad-reaching, in that they may call for a reinterpretation of decades' worth of studies of STM and WM. One important caveat, however, is that these new findings were derived solely from fMRI data. The possibility exists that UMIs are retained in an active state to which the blood oxygen level dependent (BOLD) signal of fMRI is not sensitive. To address this concern, the present study was designed to replicate the critical features of Lewis-Peacock et al. (2012), with the exception that we measured neural activity with the electroencephalogram (EEG), rather than with fMRI. EEG is sensitive to oscillatory dynamics in large populations of neurons, a signal that may reflect storage-related processes during tests of STM (Jensen & Tesche, 2002; Johnson, Sutterer, Acheson, Lewis-Peacock, & Postle, 2011; Palva, Kulashekhar, Hämäläinen, & Palva, 2011). Furthermore, EEG is more temporally precise than fMRI, which permitted more nuanced interpretations of the time course of delay-period activity. We applied MVPA to delay-period spectral data in order to test for category-specific patterns of neural activity. Although MVPA has not been as widely used in EEG as in fMRI, several recent studies have been successful in its application to EEG (Newman & Norman, 2010; Simanova, van Gerven, Oostenveld, & Hagoort, 2010) as well as MEG

(Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010). As was the case with Lewis-Peacock et al. (2012), in the present study we set out to test the hypothesis that UMIs are maintained at a level of activation that is intermediate between that of items that are in the focus of attention and baseline, a finding that would be consistent with the prediction of theoretical models (Cowan, 1988; Oberauer, 2002) and with previous fMRI studies (Lepsien & Nobre, 2007; Nee & Jonides, 2008). We did so with the foreknowledge, however, that our study may fail to find such evidence for an intermediate level of activation. In the event of a failure to find support for the primary hypothesis, we also planned a second analysis that would address a specific possible concern about our procedure: Training classifiers on data from single-item retention intervals before testing them on two-item retention intervals relies on the assumption that the neural code used to represent information in the former will be the same in the latter. Thus, to confirm that a null finding in the primary analysis could not be attributed to a failure of this assumption, a second analysis would be conducted which used pattern classifiers that were trained and tested solely on the two-item retention intervals, using a k-fold cross-validation approach.

A second goal of our experiments was to further explore the dynamics and neural bases of the unloading of information from the focus of attention. The work presented here draws on prior work showing that retrocuing is an effective way to experimentally manipulate the effective memory load. Taking advantage of the fact that reaction times (RTs) are positively related to memory load (Sternberg, 1969), Oberauer and colleagues have shown with lists of words (Oberauer, 2001), numbers (Oberauer, 2002), and mixed-domain lists (Oberauer & Gothe, 2006), that retrocuing a subset of the memoranda produces a decrease in RT. The magnitude of this decrease, however, is dependent on the latency between the retrocue and the memory probe – the cue-probe interval (CPI; referred to in other publications as the cue-stimulus interval; CSI). At short CPIs (100-500 ms), response times reflected a memory load consisting of the number of items in both the cued and uncued memory lists (i.e., the number of items presented at the beginning of the trial). However, at longer CPIs (1-2 s), they reflected a load consisting only of the number of items in the cued list. These results were even observed when the uncued memory items were potentially relevant for a second memory probe that could occur subsequently in a two-step variant of the task (Oberauer, 2005). The interpretation of these findings was that, upon receiving the retrocue, participants initiated the removal of the uncued memory items from the capacity-limited focus of attention and that, because this process takes time, short CPI trials probed memory before the uncued items had been unloaded. This logic has been used to estimate that the unloading process takes approximately 1 s per item to complete. The items were "unloaded" into a higher-capacity secondary storage layer in STM (e.g., "the region of direct access", Oberauer, 2002; "activated long-term memory", Cowan, 1995) in which their maintenance would not interfere with ongoing processing. Here, we sought to replicate these findings in a series of behavioral experiments using the stimuli designed for our EEG experiment, systematically varying the CPI to investigate the unloading process. The first experiment used a task with stimuli drawn from the same category on each trial, in order to establish a baseline estimate of unloading time for these stimuli. The second experiment differed from the first only in that the two memory sets were drawn from different categories on each trial, in order to examine cross-category effects of the unloading process. The third experiment, in addition to the trials with two memory sets, had trials with only one memory set, permitting us to assess the effect of remembering an uncued memory set. Then, because EEG permits excellent temporal resolution of neural activity, we sought to corroborate our behavioral estimate with an independent neural estimate of the unloading process. This neural estimate was derived by assessing the trajectory of MVPA evidence for the uncued item (i.e., the UMI) in the delay-period EEG signal. We hypothesized that the neural deactivation of the UMI underlies the unloading effect, and therefore we predicted that the deactivation time estimate should correspond to the unloading time estimate from the behavioral experiments.

Any discrepancy between these estimates would indicate that other neural processes, in addition to UMI deactivation, might underlie the behavioral unloading effect.

Methods

Overall design

Experiment 1 involved two behavioral tasks (phases 1 and 2), which were administered in a single session with concurrent EEG recording. The Phase 1 task – a single-item delayed-recognition task – permitted us to train a classifier to distinguish category-specific patterns of delay-period EEG data corresponding to the active retention of a single memory item. The Phase 2 task – a two-item, two-step delayed-recognition task with retrocues – permitted us to dissociate the active retention of attended vs. unattended memory items in STM. To interpret the EEG data from the Phase 2 task, we applied patterns classifiers that were trained either on single-item retention intervals (Phase 1, all trials) or two-item retention intervals (Phase 2, independent trials via k-fold cross validation).

Experiment 2 involved behavioral data collection only. We administered three different delayed-recognition tasks with two sets of stimuli, while systematically varying the time between the retrocue and probe to derive an estimate of the time course of the unloading process. In Experiment 2a, two sets of stimuli were presented from the same category; each stimulus set could be either high (multiple items) or low (single item) load. After the stimuli were presented, a retrocue indicated which of the two sets would be the target of the first memory probe (this set contained the AMIs) and which was irrelevant for the first probe (this set contained the UMIs). The primary measure of interest was the RT to this first probe and a determination of the amount of time during which it depended on the size of the UMI set. Experiment 2b was nearly identical, except that on each trial, stimuli from one set were drawn from a different category than stimuli from the other set. This task most closely resembled the Experiment 1, Phase 2 EEG task. Finally, to assess whether the unloading of UMIs exerts a residual performance cost, behavioral Experiment 2c compared RTs on trials that required the removal of a UMI set vs. trials that did not (i.e., trials that consisted of only a single set of AMIs).

Experiment 1 participants

Eighteen participants (12 female – average age 22 years) were recruited from UW-Madison and surrounding areas. Participants were screened for medical, neurological and psychiatric diagnoses that were exclusion factors for participation. All participants gave informed consent.

Experiment 1, Phase 1 task

Participants performed a one-item delayed-recognition task (Fig. 1A) modeled closely on the one used in Phase 1 of Experiment 2 in the previous fMRI study (Lewis-Peacock et al., 2012). Each trial was preceded by a 2 s presentation of the instruction to "Blink now." Next, a fixation cross appeared for 2 s, followed by a category cue (2 s), target stimulus (0.5 s), delay period (5 s), probe stimulus (0.5 s), response period (1.5 s), and feedback (1 s). The category cue indicated the category of the item to be remembered on that trial: either a pair of variously oriented line segments, a pronounceable one-syllable pseudoword, or an English word. To prevent confusion between the pseudoword and word stimuli, pseudowords were presented in cyan and words in white. At the end of the delay period, subjects responded to the probe stimulus with a numeric keypad, with "1" indicating a match and "4" indicating a non-match. Subjects had a total of 2.5 s to respond, and feedback was provided on a trial-by-trial basis by changing the color of the fixation cross to green for correct responses, red for incorrect responses, and yellow for late responses. The criterion

for a match differed for the three categories. For line segments, the probe stimulus had to exactly match the orientation of the target. For pseudowords, the vowel sound in the probe had to match the vowel sound of the target. For words, the probe word had to be synonymous with the target word. These criteria were designed to elicit different encoding strategies, and therefore different domains of active retention during the delay period, for the three categories (visual, phonological, and semantic STM, respectively). Subjects performed four blocks of 18 trials (72 total trials; 24 from each category) of this task.

Experiment 1, Phase 2 task

The task in Phase 2 (Fig. 1B) was modeled on the task used in Phase 2 of Experiment 2 in Lewis-Peacock et al. (2012). Two items from different categories (but always from the three categories used in the Phase 1 task) were presented as targets on each trial. One target item appeared on the top and one on the bottom of the screen. After an initial delay period (5 s), two inward-facing red arrows appeared at the top or bottom position (each with p = 0.5) to cue which item would be tested by the first memory probe. After a second delay period (5 s) and the first probe, subjects responded and received feedback in the same manner as the Phase 1 task. After the feedback, a second cue appeared -- on half 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 it was equiprobable that this item would be the target of the second probe. Feedback was provided after each response as in the Phase 1 task. Subjects performed eight blocks of nine trials each, and trials were counterbalanced for stimulus category (pairwise combinations of visual, phonological, and semantic stimuli), stimulus location (top vs. bottom), first cue location (top vs. bottom), and trial type (cue repeat vs. cue switch).

EEG data collection

Both tasks were implemented using EPrime 2.0 software projected on an LCD monitor situated approximately 24 inches in front of the participant. Response times and accuracies were collected from the behavioral tasks. No data were filtered based on response accuracy because of the subjective nature of the domain-specific stimulus comparisons required (e.g., if a subject interpreted the vowel sound of a pseudo-word differently than we intended, that subject may have responded "incorrectly" to a probe based on our phonological interpretation, although they may have been accurately performing the task by maintaining phonologically-based representations of the task stimuli).

EEG data were recorded using a 257-electrode net with an EGI amplifier and Netstation acquisition software. The impedance of each electrode was kept below 75 kOhms, and the sampling rate was 500 Hz.

EEG preprocessing

The data were processed offline using the EEGLAB (Delorme & Makeig, 2004) and Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) toolboxes in MATLAB (Mathworks, Natick, MA, USA). First, 72 channels along the face, ears and neck were removed. These channels are most susceptible to noise due to poor contact and non-neural signals due to muscle and eye movements. Next, the data from Phase 1 and Phase 2 were separated and trials containing excessive noise or artifact were discarded. The total number of trials discarded was quite small (mean less than one, maximum of 3). The signal was then band-pass filtered between 1 and 55 Hz using EEGLAB. Next the data from each phase were epoched, thus discarding the blink-contaminated ITIs. Independent component analysis was performed separately for Phase 1 and Phase 2 data using the EEGLAB Infomax algorithm. A conservative rejection threshold was employed, and by inspecting the

topography, time series and power spectrum of each component, those which predominantly captured eye movements, blinks, muscle artifact, or residual electrical noise were removed (McMenamin et al., 2010). An average of 81 components was removed for each phase for each subject. Finally, the signal from all channels was average referenced and a Morlet wavelet transform was performed on these data using the Fieldtrip toolbox. A wavelet at every integer frequency from 2-20 Hz and every other integer from 22-50 Hz was used with a fixed, Hanning-tapered window of 0.5 s. This transform resulted in spectral power values at each of 34 frequencies and 185 channels, sampled every 0.5 s, for each trial. The spectral time series was smoothed by averaging each value with the two preceding and two subsequent time points, such that each time point reflected the average EEG data from a 2.5 s window of activity. This temporal smoothing procedure was necessary to minimize the noise of the dynamic EEG signal. These data became the features used for all subsequent pattern classification analyses.

MVPA training – Phase 1

Multivariate pattern analysis was performed in MATLAB using the EEG Analysis Toolbox (code.google.com/p/eeg-analysis-toolbox/) together with the Princeton MVPA toolbox (code.google.com/p/princeton-mvpa-toolbox). The classification algorithm used for this analysis was L2-regularized logistic regression, with a penalty term of 1. L2 regularization penalizes large feature weights, preventing any one feature from having a disproportionate effect on the classification. Higher and lower penalty values were tested on a separate group of 4 pilot subjects not included in this analysis. A regularization penalty of 1 produced the most reliable classification in that group, and so was used for the remainder of the analyses. No pre-classification feature selection was performed.

First, the classification procedure was validated using k-fold cross-validation on the Phase 1 data. EEG data from the middle 4 s of the delay period from each trial was used. The first (0 – 0.5 s) and last (4.5 – 5.0 s) time points were excluded in order to minimize the effect of stimulus-evoked activity (from the target and probe, respectively) on the temporally smoothed data. This still permitted some influence of the stimulus and probe to enter into the classification, because the temporal smoothing meant that each data point was an unweighted average of a 2.5 s window; therefore, the data point centered at 0.75 s averaged over data from -0.5 s to 2 s (thus including the stimulus), and the data point at centered at 4.25 s was an average of data from 3 s to 5.5 s (thus including the probe). We included these points in order to maximize the available data points for classification, and because in a pilot data set classification accuracy was improved by retaining them (although classification was still successful if they were excluded). Our analysis scheme considered each 0.5 s time point as a separate training exemplar, so that every trial yielded eight exemplars. Each exemplar was described by a feature matrix of 34 frequencies by 185 channels (i.e., 6,290 unique features). Each feature was z-scored across all trials and time points. The k-fold cross-validation scheme (k = 72) trained a classifier on data from 71 trials and then used this classifier to test the one withheld trial. This process was repeated until every trial had been held out for testing. Statistical significance of classifier accuracy was evaluated by performing a one-sample t-test comparing accuracy to chance performance (33%). Statistical significance of classifier evidence was evaluated by pairwise comparisons of delay-period evidence values using one-tailed paired t-tests for each trial type. Two subjects for whom the cross-validation classification accuracy was below chance were not included in subsequent analyses.

MVPA decoding – Phase 2 (trained on Phase 1)

The classifiers trained in Phase 1 were then applied to the data from Phase 2. The classifier outputs a measure of classifier evidence for each category at every time point. Classifier

evidence was averaged separately across cue switch and cue repeat trials for the initially cued, initially uncued, and not present categories. Before statistically evaluating the results, evidence values were averaged across the middle 3 s of each delay period. Excluding the first and last second of each delay period removed any influence of the evoked responses to the cues and probes, which were not signals of interest in this analysis. Statistical significance of evidence levels was assessed with one-tailed paired t-tests, using the sixteen included subjects as independent measures. Using the absent category (the category not present on a given trial) as a baseline, we compared classifier evidences from the cued and uncued categories to the classifier evidence for the absent category.

MVPA decoding – Phase 2 (trained on Phase 2)

We also performed classification analyses using data exclusively from Phase 2. In this analysis, we focused only on the visual category, because the Phase 1 classifiers were most sensitive to identifying visual STM. We analyzed cued information by attempting to discriminate the trials in which a visual category stimulus was present and cued from those in which it was absent, and we analyzed uncued information by attempting to discriminate the trials in which the visual category was present and uncued from those in which it was absent. To illustrate, for the classification of visual cued trials, we relabeled the EEG data from these trials (visual-semantic and visual-phonological trials) as "visual-cued," and all trials in which the visual category was completely absent (semantic-phonological and phonological-semantic trials) as "visual-absent." Trials in which the visual category was present but uncued were set aside for this analysis. Next, leave-one-out cross-validation classification of "visual-cued" vs. "visual-absent" was performed – the output of this analysis was an evidence value (between zero and one) for "visual-cued" and "visual-absent" for each trial. Successful classification would entail "visual-cued" trials having higher classifier evidence for "visual-cued" than for "visual-absent." We performed the analysis within a time window comprising two adjacent time points (0.5 s each; 1 s total); we then shifted the window along the time axis, performing k-fold cross validation within each window, until the entire time span of interest had been decoded. This meant that classification at a given time window relied only on data from that window, which permitted us to be sensitive to signals which varied temporally throughout the delay period. For example, if the neural signal related to the UMIs was recoded into a different but still active representation during the post-cue delay period, this sliding-window analysis would permit the different active representations to be classified separately (i.e., based on separate training data), and therefore classification would be successful throughout the delay period. This approach contrasts with training and testing over the entire delay period, as was done with the Phase 1 k-fold cross-validation. That approach entailed the implicit assumption that the signal from different parts of the delay period in the training trials would be similar. The entire sliding-window, k-fold cross-validation procedure was repeated for trials in which the visual category was uncued, but this time, trials in which the visual stimulus was cued were set aside. To measure the classifiers' ability to distinguish trials with visual stimuli ("visual-cued" and "visual-uncued") from trials with no visual stimuli ("visual-absent"), we used the area under the receiver operating characteristic curve (AUC) as a metric (Fawcett, 2006; Newman & Norman, 2010) . An AUC of greater than 0.5 indicates sensitivity to the category of interest. Statistical significance of the sensitivity was assessed with one-tailed, one-sample t-tests of AUC values averaged across the middle 3 s of each delay period vs. the zero-sensitivity value of 0.5, using the 18 subjects as independent measures.

Time course estimate of UMI deactivation

We sought to characterize, in the EEG classification data, the time course of the removal of memory items from the focus of attention. To do this, we focused on the first retrocue, because this cue indicated that memory items were to be unloaded from attention but still

retained as memory items. We combined data from the cue-repeat and cue-switch trials to produce the best estimate of unloading time in this first post-cue delay period; this procedure is valid because until the second cue, the two trial types are indistinguishable from one another. A paired, one-tailed t-test comparing the uncued evidence to the absent evidence was performed at each time point; the point at which the uncued evidence was no longer significantly greater than baseline (at $p < 0.05$) served as our estimate of the time point at which the neural signal related to the UMI was no longer evident in the EEG data.

Experiment 2 – subjects

For Experiment 2a, 16 subjects (right-handed; native English speakers; 13 female; ages 18-27) were recruited from the undergraduate and medical campuses of the University of Wisconsin-Madison. An additional 22 subjects (right-handed; native English speakers; 13 female; ages 18-21) were recruited for Experiment 2b, and separate group of 21 subjects (right-handed; native English speakers; 17 female; ages 18-20) were recruited for Experiment 2c. None reported any medical, neurological, or psychiatric illness, and all gave informed consent.

Behavioral experiments – task procedures

Experiment 2a: Same-category stimulus unloading—The first behavioral task was a delayed-recognition task with two sets of stimuli and with retrocues similar to the Phase 2 task of Experiment 1 (appendix Fig. 1A). The rationale was to systematically vary the timing between the retrocues and the recognition probe, in order to determine when the set size of the uncued memory set stopped influencing RT to the probe. Stimuli were drawn from the same three categories (line segments, words, pseudowords) as the Experiment 1 tasks. Each trial began with the visual presentation (for 1.3 s) of two sets of stimuli drawn from the same stimulus category - one set appeared on the top half of the screen and the other on the bottom half. Each set comprised one item (low load) or multiple items (2 words/ pseudowords or 3 line segments; high load), and set sizes were varied orthogonally for the two sets. After 0.7 s of a blank screen, a retrocue, two inward-facing red arrows, appeared at the top or bottom position (each with $p = 0.5$) to indicate which of the two memory sets would be tested by the first memory probe. The cue appeared for 0.1, 2, or 4 s, and was followed immediately by a recognition probe. Subjects responded with a Yes/No button press based on the relevant category-specific comparison. Feedback was provided, followed by a second retrocue that selected, with equal probability, one of the two memory sets as relevant for the second recognition probe. The duration of this second cue also varied between 0.1, 2, and 4 s, orthogonal to the duration of the first cue. Subjects performed 432 trials (144 trials for each stimulus category) in two separate sessions. Each session began with 18 practice trials (six for each category) followed by 216 experimental trials. Trials were arranged into 27 blocks of eight trials each (nine blocks per stimulus category). Within each block, all possible set size combinations (e.g., high-high, high-low, low-high, and low-low) occurred equally often in random order, and the memory set location selected by the first and second cues was counter-balanced. Each block represented one combination of first cue duration and second cue duration, and blocks were presented in one of four pseudo-random orders for each subject. All blocks of one stimulus category were completed before advancing to the next category. The second session was completed between one day and two weeks after completion of the first.

Experiment 2b: Mixed-category stimulus unloading—Trials were identical to those of Experiment 2a, except that the two sets of memory items for each trial were selected from two different categories (appendix Fig. 1B). The rationale of having the two sets come from different categories was to test the hypothesis that UMIs of a different category than the AMIs would still exert a load effect on RT. Within each set, all stimuli were of the same

category. Therefore, the category of the memory set that was selected by the retrocues determined the category of the recognition probes. All trials of one category combination (e.g., words and pseudo-words) were completed before advancing to the next combination.

Experiment 2c: Residual effects of stimulus unloading—Trials were configured as in Experiment 2a, except for three modifications: (1) only nouns were used as stimuli, (2) recognition probes were simple match/non-match comparisons rather than synonym judgments; (3) memory items were presented either as two sets of two (4 items in total) or as one set of two items (appendix Fig. 1C). The timing of cues and probes was identical for both conditions, but there were no UMIs to unload for the trials in which only one memory set was presented. This design allowed us to compare trials in which UMIs had to be retained to trials without any UMIs, permitting us to study the behavioral effects of retaining UMIs. Subjects performed 128 trials, divided into four blocks with 32 trials each, and each block began with 12 practice trials. Within each block, an equal number of trials appeared in each condition (4 stimuli vs. 2 stimuli), and this variable was crossed with three durations of the retrocues (0.1, 2, or 4 s). Trials from the 6 conditions of this design were presented in random order. The experiment was completed in one session lasting 50 minutes.

Behavioral experiments – data collection—All experiments were implemented with E-Prime software version 2.0 (Psychology Software Tools). Response times were collected for both responses in the behavioral experiments, although we present results only for responses to the first probe in each trial (replicating the analyses of Oberauer, 2001). Data were identified as outliers and removed if a response time was at least three standard deviations above or below the mean response time for a given stimulus category (or category combination in behavioral Experiment 2b). Trials in which no response was given were also removed (1.4%, 3.0%, and 1.4 % of trials for the three experiments). By the same logic used with the EEG experiment, no trials were removed from the analysis based on incorrect responses. The percentage of trials used for hypothesis testing in the three behavioral experiments was 98.0%, 96.7%, and 97.3%, respectively.

Results

Experiment 1 behavioral results

For the Phase 1 task, overall performance of the delayed-recognition task was 94.8% correct. A one-way repeated-measures ANOVA indicated that accuracy did not differ significantly between stimulus types ($F(2,17) = 1.84$; accuracy $\pm$ SEM 94.2 $\pm$ 1%, 93.7 $\pm$ 1%, and 96.5 $\pm$ 1% for visual, phonological, and semantic trials, respectively). For Phase 2, participants performed well above chance (overall performance collapsing across first and second probes, 89.6%). In a $2\times2\times6$ repeated measures ANOVA with the factors of probe type (first, second), trial type (cue repeat, cue switch), and stimulus type (each of the six possible pairwise combinations), main effects of probe type ($F(1,17) = 11.89$, $p = 0.003$), trial type ($F(1,17) = 5.2$, $p = 0.036$) and stimulus type ($F(5,85) = 5.31$, $p = 0.0003$) were all present. Participants were significantly better at responding to the first probe (91.5%, SEM = 1%) than the second (87.7%, SEM = 1%) and better at cue-repeat trials (90.9%, SEM = 1%) than cue-switch trials (88.3%, SEM = 1%). Finally, participants' performance varied with the stimulus type: Pairwise comparisons with Bonferroni correction indicated that participants performed worse at visual-phonological trials (when visual stimuli were probed first: 83.8%, SEM = 2%) compared with visual-semantic trials, phonological-semantic trials, or semantic-phonological trials (92.0%, 91.6%, and 91.3%, SEMs 1%, 2%, and 1%, respectively). Of the two-way interactions, only probe type by stimulus type was significant ($F(5,85) = 4.88$, $p = 0.0006$). The three-way interaction was non-significant. An identical repeated measures ANOVA for RT on correct trials was performed, and significant main

effects of trial type (F(1,17) = 6.42, p = 0.021) and stimulus type (F(5,85) = 3.36, p = 0.008) were noted. Subjects were faster to respond to repeat trials (first probe, 917 ms, SEM = 42 ms; second probe, 846 ms, SEM = 35 ms) than to switch trials (first probe, 909 ms, SEM = 41 ms; second probe, 933 ms, SEM = 43 ms). Pairwise comparisons with Bonferroni correction indicated that subjects were much faster at visual-semantic trials (853 ms, SEM = 22 ms) than they were at phonological-visual (927 ms, SEM = 28 ms), phonological-semantic (927 ms, SEM = 24 ms), or semantic-phonological (931 ms, SEM = 24 ms). The semantic-phonological RTs were also significantly slower than the visual-phonological responses (862 ms, SEM = 26.3 ms). Of the interactions, probe type by trial type (F(1,17) = 11.9, p = 0.003), probe type by stimulus type (F(5,85) = 3.46, p = 0.0005), and trial type by stimulus type (F(5,85) = 3.87, p = 0.003) were significant.

MVPA results – Phase 1 training

Spectrally transformed EEG data from the delay periods of all trials were used to train subject-specific classifiers to predict the stimulus category of memoranda. Leave-one-trial-out cross validation was used to verify classifier performance, and classification accuracy and evidence were averaged over the entire delay (Fig. 2). The overall accuracy was 45.3%, and classification accuracies within each stimulus category were all significantly above chance (category accuracy ± SEM: visual 55.0 ± 4.2%, phonological 40.4 ± 2.3%, semantic 40.4 ± 2.6%). Classifier performance for two subjects was below chance (31%, 30%), and so these subjects' data were excluded from subsequent analyses that used classifiers trained on Phase 1 data.

In examining the category level performance (Fig. 2A), it is clear that accuracy was much better for visual trials than for phonological and semantic trials. The classifier evidence indicates that this difference in accuracy stemmed from an inability of the classifier to distinguish phonological from semantic trials (Fig. 2B). That is, on both phonological and semantic trials, the classifier evidence for the visual category is reliably low, but there is no distinction between phonological and semantic categories on either trial type. This inability to distinguish between phonological and semantic information would be problematic if these were the only two categories in our design. However, in a three-category design, there are three possible pairwise comparisons (visual vs. phonological, visual vs. semantic, phonological vs. semantic) – two of these three comparisons are still valid with our design. Importantly, because we average across trials with different stimulus types (which are balanced in our design), this loss in power equally affects distinctions between the cued, uncued, and absent (label we use for the stimulus category not present on the given trial) classifier evidences. The net effect of losing one of the three possible comparisons is therefore simply a loss of power, which is mitigated by averaging across all trial types.

MVPA results – Phase 2 decoding (trained on Phase 1)

The classifiers trained on the Phase 1 data were applied in a subject-specific manner to the spectrally transformed EEG data from Phase 2. The classifier evidence for each category was collapsed across stimulus category by averaging together all of the evidence from each of the cued, uncued, and absent categories (the absent category is the one not present on a given trial). Because the trials were balanced with respect to stimulus type, there is no bias of stimulus category – that is, the visual category, for example, was cued, uncued, and absent on an equal number of trials. Cue switch and cue repeat trials were considered separately (Fig. 3). The evidence during the first delay period after the stimulus presentation clearly differentiated the categories which were present from the one which was absent (p < 0.005, t > 3.1 for both categories on cue-switch trials; p < 0.05, t > 1.9 for both categories on cue-repeat trials). This finding validates the assumption that similar neural codes would be used during the Phase 1 and Phase 2 tasks. After the first cue, evidence for the cued item's

category rose and evidence for the uncued item's category fell to baseline. Evidence for the uncued item's category was indistinguishable from the evidence for the absent category (switch trials: p = 0.09, t = 1.4, repeat trials: p = 0.25, t = 0.68), while evidence for the cued item's category was still clearly above baseline (p < 0.005, t > 3.5 for both switch and repeat trials). In the cue repeat trials, the same pattern was evident after the second cue (cued vs. baseline, p = 0.002, t = 3.4; uncued vs. baseline, p = 0.8, t = -1.01). In the cue switch trials, the evidence traces switched places in the second delay period, whereby evidence for the previously uncued category was reinstated at a high level distinct from baseline (p = 0.002, t = 3.4) and the previously cued category dropped to baseline (p = 0.09, t = 1.4).

We also used these data to derive an estimate of the unloading time of UMIs. The amount of time elapsed after the first retrocue before the uncued category's classifier evidence dropped to baseline level is the time that evidence for an active representation of the uncued category persisted in the EEG signal. Although the temporal smoothing of the EEG data (necessary for successful classification) makes it difficult to exactly infer a time course of this process, the time at which a one-tailed, paired t-test failed to distinguish the two categories at a p < 0.05 level was 1.25 s after the cue onset.

MVPA results – Phase 2 decoding (trained on Phase 2)

We also trained a set of classifiers on the Phase 2 data. In this analysis, we attempted to classify trials by the presence or absence of visual information. We limited the analysis to the visual category because of the inability of MVPA to distinguish between phonological and semantic categories in the EEG signal (see Fig. 2). Thus, this analysis provided maximum sensitivity for detecting any signal related to the active retention of the UMI. Trials in which visual information was cued were considered separately from trials in which it was uncued, and classifier sensitivity to the presence of visual information was plotted (AUC values > 0.5 indicate classifier sensitivity). In the initial, pre-cue delay period, the classifier was clearly sensitive to the presence of visual information (p < 0.05, t > 1.8 for both analyses). After the cue, the classifier was still sensitive to the presence of visual information when it was cued (p = 0.002, t = 3.3), but not when it was uncued (p = 0.4, t = 0.21).

Experiment 2 behavioral results

A summary of the accuracies, RTs, and omnibus ANOVAs can be found in Tables 1 and 2 in the appendix. The critical effect for our primary hypothesis was the interaction between cue duration (CPI) and the size of the uncued memory set. Specifically, we hypothesized that at short cue durations there would be a significant difference between RTs for trials with large vs. small uncued memory sets, but that at longer cue durations this effect would dissipate. The ANOVA supports our hypothesis by demonstrating a significant CPI x uncued set size interaction, F(2,14) = 17.4, p < 0.001 for Experiment 2a; F(2,14) = 3.62, p = 0.047 for Experiment 2b (appendix Table 1). Critically, planned t-tests revealed that after a cue duration of 2 s for same-category stimulus sets (exp. 2a) and 4 s for mixed-category stimulus sets (exp. 2b), RTs were not statistically different for trials with large uncued sets vs. trials with small uncued sets (see appendix Fig. 2). That is, after 4 s in both experiments, response times to probes became insensitive to the number of currently irrelevant memory items. Similar results for Experiment 2c are shown (appendix Fig. 2, and Table 3) for each cue duration comparing high-load vs. no-load conditions for the uncued memory set. After 4 s the response time for relevant stimuli was insensitive to whether irrelevant stimuli had also been presented at the beginning of the trial.

Discussion

The present study sought to test the hypothesis that memory items retained outside the focus of attention ("UMIs") are maintained in an activated state that is intermediate between items retained in the focus of attention and baseline. This hypothesis was motivated theoretically by several models of STM that include different levels of activation for memory items in and out of the focus of attention (Cowan, 1988; Oberauer, 2002; Olivers, Peters, Houtkamp, & Roelfsema, 2011). However, in agreement with a previous fMRI study (Lewis-Peacock et al., 2012), the present EEG study found evidence for elevated activity only for items maintained in the focus of attention. No evidence was found for an intermediate level of activity for UMIs. Critically, this was true when decoding the EEG data with classifiers trained on AMI retention from a different task (Phase 1), but also with classifiers trained on UMI retention from the same task (Phase 2). This suggests that an active trace – the sustained, elevated neuronal firing observed during retention intervals and often interpreted as reflecting the mechanism of STM retention – may not be necessary for STM retention, and may rather reflect the focus of attention.

In our Phase 1 analysis, it was not possible to reliably distinguish trials in which phonological stimuli were remembered from those in which semantic information was remembered. This contrasts with the results previously obtained using fMRI data, in which classifier evidence did reliably separate phonological and semantic trials (Lewis-Peacock et al., 2012). Although the inability to separate phonological and semantic stimuli was unexpected, it is not altogether surprising given the similarity of the stimuli used and the likelihood that subjects retained phonological, as well as semantic, representations of semantic stimuli. In surveying the literature, there is simply not much precedent for decoding phonological and semantic information from EEG data. The most relevant study (Simanova et al., 2010) attempted to classify conceptual categories (animals vs. tools) of stimuli presented in three sensory modalities: visual (an illustration), auditory (a spoken word), and orthographic (a written word). The success of their classification varied significantly as a function of how the stimuli were presented. With visually presented pictures, classification was successful for 20/20 subjects, with auditory presentation of the corresponding words, classification was successful for 8/20 (same subjects), and for visual presentation of the same words (i.e. orthographic), classification was only successful for 2/20 subjects. Their experiment differed from the present work in that they decoded within, not between, the sensory modalities. Additionally, they applied MVPA to time-domain event-related potentials, rather than spectrally transformed EEG data. Nonetheless, it is noteworthy that for Simanova et al. (2010), the auditory and orthographic stimuli were not as amenable to classification as the visual stimuli. It may simply be the case that methods of classification thus far attempted are not as sensitive to phonological and semantic information in EEG data.

In the present study, the Phase 2 decoding results from the initial, pre-cue delay period are interesting to consider. During this time, there was no experimental control of the focus of attention, and the decoding analyses indicate that information was present for both items being retained. There are several accounts which may explain these results. First, it is possible that both items were retained equally in the focus of attention for the duration of this delay period. This would require that the capacity of the focus of attention is at least two items, contradicting models which posit a one-item capacity of the focus of attention (McElree, 1998; Oberauer, 2002). A second possibility is that attention was never focused on more than one item at a time, perhaps switching between the two memory items multiple times during the first delay period, but that our trial-averaging procedure obscured this. Unfortunately, with the design of our study it is not presently possible to adjudicate among these candidate explanations. In order to have sufficient evidence to identify the cued items,

averaging over many trials is necessary. This removes the possibility of testing the hypothesis that on a trial-by-trial basis, attention was preferentially (perhaps transiently) allocated to only one of the two memory items.

A potential objection to the interpretation of the primary analysis and the previous fMRI results (Lewis-Peacock et al., 2012) is that the classifiers used were trained on a task (the Phase 1 task) in which subjects needed to remember only one item. Presumably, this item was being retained in the focus of attention (i.e., it was an AMI), raising the possibility that a classifier trained on such data might only be sensitive to information that is both retained and in the focus of attention (other AMIs). If this were the case, then there would still be the possibility of a separate neural signal which corresponds to retained but unattended information (UMIs). Two lines of evidence argue against this possibility. First, in both the present study and Lewis-Peacock et al. (2012), during the initial, precue delay period of the Phase 2 task, evidence was present for the categories of both items in memory. This indicates that MVPA can be sensitive to the category of information even when more than one category is concurrently active. It is an open question as to how attention interacts with multiple memory items that are equally behaviorally relevant. Second, we performed a follow-up analysis using classifiers trained and tested on Phase 2 data in a k-fold cross-validation analysis (Figure 4). In this analysis, classifiers were trained to distinguish trials in which visual information was present vs. those in which it was absent. This strategy allowed us to test for any sort of signal related to the active retention of visual UMIs. Before any cue appeared, the classifier was sensitive to visual information. After the visual information became uncued, however, this sensitivity was lost. This analysis confirmed that there was visual information in the delay-period activity only when visual information was in the focus of attention.

A key feature of the present study was the higher temporal resolution afforded by EEG compared to fMRI. Some of this temporal precision was lost to temporal smoothing, which was necessary for successful classification. However, even with temporally smoothed data, our higher sampling frequency still permitted us to look for more nuanced, time-varying signals in the EEG than was possible with fMRI. Previous work has suggested that the unloading of items from the focus of attention into less activated STM states takes place over a time scale of approximately 1-2.5 s (Oberauer, 2005; Oberauer, 2001). Our own estimates of this unloading process, using stimuli chosen for the EEG experiment, were in reasonable agreement with prior results. RTs became insensitive to the size of the UMI set between 2 and 4 s, depending on whether the memory sets were drawn from the same or from different categories. Interestingly, our mixed-category unloading result (exp. 2b) represents a departure from previous work, where interference between cross-category sets was never observed (Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002) or only observed when items from both sets needed to be available for processing (Oberauer & Gothe, 2006). This finding is consistent with the idea that the active storage of information in STM is supported by a domain-general focus of attention.

Our estimate derived from the EEG data of the time required for the neural representation of a single UMI to fall to baseline was 1.25 s. In order to compare these results with the behavioral estimates of unloading time, it was necessary to correct for the fact that the behavioral estimates were derived from a comparison of RTs in high-load (multiple UMIs) vs. low-load (single UMI) conditions. Such relative estimates reflect the time required to unload the larger UMI sets, and therefore we must correct for the number of items in these sets in order to derive an estimate for unloading a single UMI. In Experiment 2b, the influence on RTs of the UMI set size had disappeared by 4 s. (Note that we did not test CPIs in between 2 and 4 s, thus 4 s is likely an overestimation of the unloading time.) Our high-load UMI set sizes were a mixture of 2 (for words and pseudowords) and 3 (for line

segments) items, so we can approximate a high-load set size of 2.33 items. Dividing 4 s (unloading time for the high-load UMI set) by 2.33 items yields a per-item estimate of approximately 1.7 s, which agrees relatively well with our EEG-derived neural estimate of 1.25 s. The discrepancy between the two numbers is easily accounted for by the high amount of temporal smoothing in our EEG data. Despite the relative agreement of our behavioral and neural estimates of unloading time, it is worth remembering that some behavioral consequences of the retention of UMIs certainly persist longer than their active, decodable neural traces. For example, in a study by Oberauer (2002), intrusion effects (slowed rejection of probe stimuli that had been part of the uncued memory set) were present at all time points tested, up to 5 s after the retrocue. That this rather subtle behavioral effect of unloading persists until long after fMRI or EEG can find evidence for the UMI indicates that it can be problematic to conflate activation as a theoretical construct supported by behavioral measures with activation as measured by neurophysiological recordings. As we discuss below, there are mechanisms of short-term plasticity to which fMRI and EEG are insensitive which could underlie the phenomenon of an inactive, yet retained and behaviorally impactful UMI.

Independent of the behavioral results' relation to the EEG estimate of the unloading time, the results from the three behavioral experiments provide unique insight into the unloading process. Previously, unloading times for lists of words and numbers had been estimated to be approximately 0.33 to 1.0 s per item, based on reaction times to a probe appearing at various intervals after a retrocue (Oberauer, 2001, 2002, 2005). Our estimate of 1.7 s per item is reasonably consistent with these results. Interestingly, the results of our behavioral Experiment 2b represent somewhat of a departure from previous work, in that we report cross-category effects on RT. Specifically, in trials with two memory sets from different categories, in which one of the categories was cued as relevant and the other category was irrelevant, we observed RTs to be dependent on the uncued memory set size. This implies that the resources freed by unloading uncued memory items are not domain specific.

An irony of terminology arises in our suggestion that "activated long-term memory" does not require an active trace. The discrepancy is indeed one of terminology, not substance – the meaning of "activated" within a theoretical model need not correspond with the usage of "active" to characterize increased measured neural activity. The distinction we suggest between the focus of attention and the broader pool of activated LTM leaves open the question of what might differentiate activated LTM from the immense network of latent LTM. In short, how might information (UMIs) be maintained without giving rise to a measureable active trace? This study is unable to address this important question. However, a possible explanation which would be consistent with our results is a passive storage mechanism in the form of a transient, latent network of potentiated synaptic weights. GluR1-dependent short-term potentiation (Erickson, Maramara, & Lisman, 2009) and transient presynaptic increases in calcium ion concentration (Mongillo, Barak, & Tsodyks, 2008) are two physiologically plausible candidate mechanisms which could create such a transient synaptic network.

The present study demonstrates distinct states of retention in STM corresponding to items inside and outside the focus of attention. Only items inside the focus of attention were maintained in a state that could be detected in the delay-period EEG data. This, together with previous fMRI results (Lewis-Peacock et al., 2012), suggests that an active trace of short term retention is only present when attention is also allocated to the retained information.
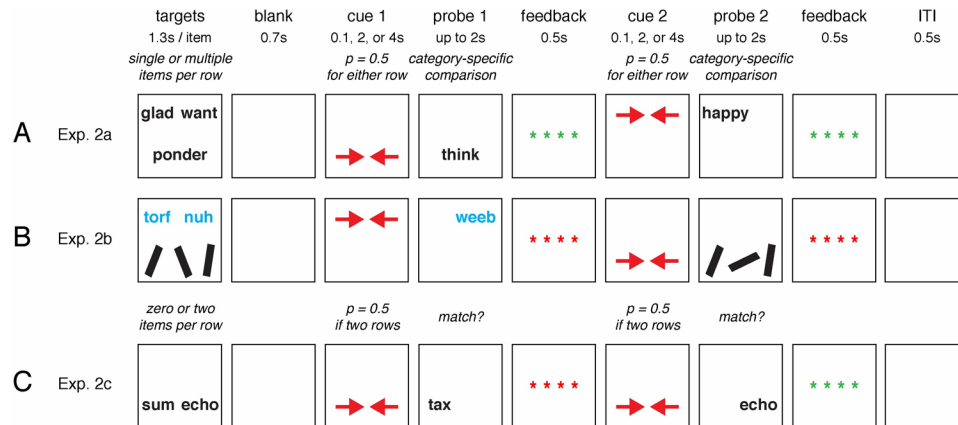
Acknowledgments

Appendix

Figure 1.
Task procedures. (A) Experiment 2a: subjects performed short-term recognition of words, pseudo-words, or line segments. Both sets of stimuli for each trial were drawn from the same category. (B) Experiment 2b: subjects performed the same task but with each stimulus set drawn from a different category. (C) Experiment 2c: subjects performed short-term recognition for trials with either 2 sets of 2 stimuli each (4 total) or 1 set of 2 stimuli.
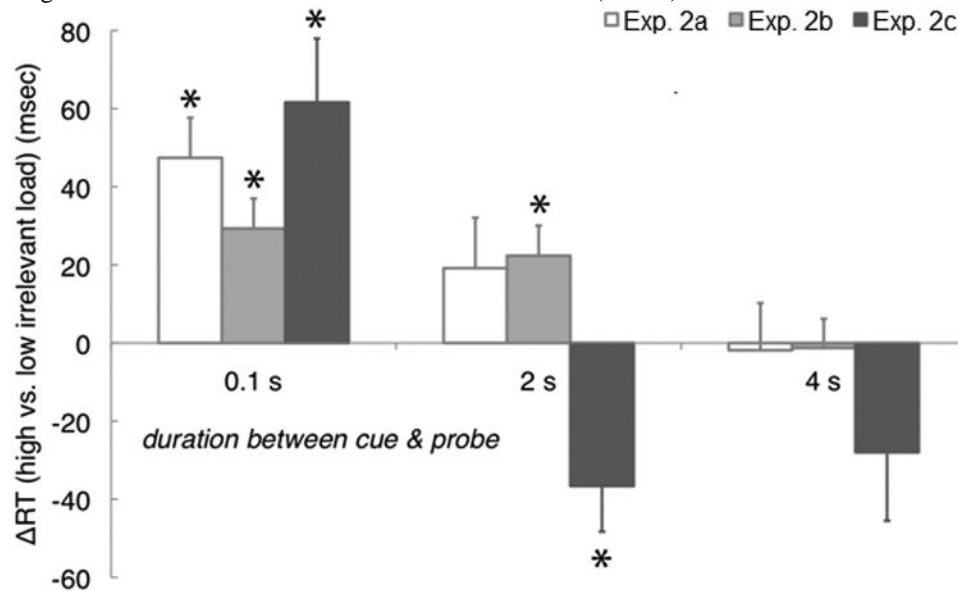

Figure 2.
Unload effects on RT for behavioral experiments. Data are grouped along the horizontal access according to the length of time between the onset of the cue and the onset of the probe (0.1, 2, or 4 s). The vertical axis shows the difference in RTs (from the first recognition probe) between conditions of high- and low-load of uncued (irrelevant) stimuli. For Exp. 1 and 2, the high-load condition consisted of multiple stimuli in an uncued set, and the low-load condition consisted of a single uncued item. For Exp. 3, however, the high-load

condition had two stimuli and the low-load condition had none (i.e., no unloading of UMIs was required).

Table 1

ANOVA results for RTs to the first probe in behavioral Experiments 2a, 2b, and 2c. (CPI = cue-probe interval, the duration between the onset of the cue and the onset of the probe stimulus; stimulus category pair = the pairwise combinations of stimulus category).

A. Experiment 2a

Effect	$F (df_{effect}, df_{error})$	p
Category	102.87 (2, 14)	< .001 *
CPI	45.90 (2,14)	< .001 *
Cued Set Size	154.43 (1, 15)	< .001 *
Uncued Set Size	6.04 (1,15)	.027 *
Stimulus Category × CPI	0.56 (4, 12)	0.696
Stimulus Category × Cued Set Size	2.44 (2, 14)	0.124
CPI × Cued Set Size	30.74 (2, 14)	.000 *
Stimulus Category × CPI × Cued Set Size	6.23 (4, 12)	.006 *
Stimulus Category × Uncued Set Size	4.49 (2, 14)	.031 *
CPI × Uncued Set Size	17.42 (2, 14)	< .001 *
Stimulus Category × CPI × Uncued Set Size	0.45 (4, 12)	0.768
Cued Set Size × Uncued Set Size	6.20 (1, 15)	.025 *
Stimulus Category × Cued Set Size × Uncued Set Size	0.66 (2, 14)	0.531
CPI × Cued Set Size × Uncued Set Size	0.56 (2, 14)	0.581
Stimulus Category × CPI × Cued Set Size × Uncued Set Size	0.46 (4, 12)	0.766

B. Experiment 2b

Effect	$F (df_{effect}, df_{error})$	p
Stimulus Category Pair	69.67 (2, 14)	< .001 *
CPI	5.23 (2, 14)	.016 *
Cued Set Size	228.29 (1, 15)	< .001 *
Uncued Set Size	13.16 (1, 15)	.002 *
Stimulus Category Pair × CPI	6.31 (4, 12)	.003 *
Stimulus Category Pair × Cued Set Size	4.58 (2, 14)	.024 *
CPI × Cued Set Size	5.34 (2, 14)	.014 *
Stimulus Category Pair × CPI × Cued Set Size	0.76 (4, 12)	0.568
Stimulus Category Pair × Uncued Set Size	1.28 (2, 14)	0.3
CPI × Uncued Set Size	3.62 (2, 14)	.047 *
Stimulus Category Pair × CPI × Uncued Set Size	6.90 (4, 12)	.002 *
Cued Set Size × Uncued Set Size	0.49 (1, 15)	0.492
Stimulus Category Pair × Cued Set Size × Uncued Set Size	2.79 (2, 14)	0.086
CPI × Cued Set Size × Uncued Set Size	4.78 (2, 14)	0.021 *
Stimulus Category Pair × CPI × Cued Set Size × Uncued Set Size	1.12 (4, 12)	0.382

C. Experiment 2c

Effect	$F (df_{effect}, df_{error})$	p

A. Experiment 2a		
Effect	**F (df$_{effect}$, df$_{error}$)**	**p**
CPI	0.012 (1, 20)	0.913
Uncued Set Size	1.528 (2, 19)	0.242
CPI × Uncued Set Size	16.086 (2, 19)	< .001 [*]

[*] - p (one-tailed) < 0.05

Table 2

Pair-wise t-tests for RTs to first probe, for behavioral Experiments 2a and 2b. Data are mean RTs for each condition, with standard deviations in parentheses (CPI = cue-probe interval, the duration between the onset of the cue and the onset of the probe stimulus).

CPI	low-load cued	high-load cued	t (paired) cued	low-load uncued	high-load uncued	t (paired) uncued
			Experiment 2a			
0.1 s	1046 (73.59)	1140 (82.44)	[**](-) 6.95	1069 (77.59)	1116 (74.84)	[**](-) 4.61
2 s	906.6 (62.09)	1073 (82.26)	[**](-) 12.5	906 (71.64)	1073 (73.48)	(-) 1.48
4 s	919.8 (70.06)	1081 (67.20)	[**](-) 12.9	1001 (65.68)	999.3 (71.02)	0.153
			Experiment 2b			
0.1 s	1028 (101.6)	1164 (106.2)	[**](-) 12.8	1081 (99.81)	1111 (105.3)	[**](-) 3.82
2 s	969.1 (101.1)	1130 (102.9)	[**](-) 14.8	1038 (103.2)	1061 (97.55)	[**](-) 2.91
4 s	961.7 (103.4)	1136 (112.0)	[**](-) 11.3	1049 (94.74)	1048 (111)	0.172

[*] p < 0.05,
[**] p < 0.01

Table 3

Pair-wise t-tests for RTs to first probe, for behavioral Experiment 2c. Data are mean RTs for each condition, with standard deviations in parentheses (CPI = cue-probe interval, the duration between the onset of the cue and the onset of the probe stimulus).

CPI	2/0	2/2	t(paired)
0.1 s	864.09 (142.37)	925.68 (141.30)	[**](-) 3.76
2 s	936.65 (152.35)	900.01 (154.05)	[**] 3.14
4 s	918.52 (161.53)	890.53 (140.83)	1.60

[*] p < 0.05,
[**] p < 0.01

References

Baddeley, AD.; Hitch, G. Working Memory. In: Gordon, HB., editor. Psychology of Learning and Motivation. Vol. 8. Academic Press; 1974. p. 47-89.

Cocchini G, Logie RH, Della Sala S, MacPherson SE, Baddeley AD. Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. Memory & Cognition. 2002; 30(7):1086–95. [PubMed: 12507373]

Cowan N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological Bulletin. 1988; 104(2):163–191. [PubMed: 3054993]

Cowan, Nelson. Attention and Memory: An Integrated Framework. New York: Oxford University Press; 1995.

Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. Trends in Cognitive Sciences. 2003; 7(9):415–423. [PubMed: 12963473]

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004; 134(1):9–21. [PubMed: 15102499]

Erickson MA, Maramara LA, Lisman J. A single brief burst induces GluR1-dependent associative short-term potentiation: a potential mechanism for short-term memory. Journal of Cognitive Neuroscience. 2009; 22(11):2530–2540. [PubMed: 19925206]

Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27:861–874.

Fuentemilla L, Penny WD, Cashdollar N, Bunzeck N, Duzel E. Theta-coupled periodic replay in working memory. Current Biology. 2010; 20(7):606–612. [PubMed: 20303266]

Fuster JM. Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. Journal of Neurophysiology. 1973; 36(1):61–78. [PubMed: 4196203]

Hebb, DO. The Organization of Behavior: A Neuropsychological Theory. New York: Wiley; 1949.

Jensen O, Tesche CD. Frontal theta activity in humans increases with memory load in a working memory task. European Journal of Neuroscience. 2002; 15:1395–1399. [PubMed: 11994134]

Johnson JS, Sutterer DW, Acheson DJ, Lewis-Peacock Ja, Postle BR. Increased Alpha-Band Power during the Retention of Shapes and Shape-Location Associations in Visual Short-Term Memory. Frontiers in Psychology. 2011; 2(June):128. [PubMed: 21713012]

Lepsien J, Nobre AC. Attentional modulation of object representations in working memory. Cerebral Cortex. 2007; 17(9):2072–83. [PubMed: 17099066]

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. Neural evidence for a distinction between short-term memory and the focus of attention. Journal of Cognitive Neuroscience. 2012; 24(1):61–79. [PubMed: 21955164]

Lewis-Peacock JA, Postle BR. Decoding the internal focus of attention. Neuropsychologia. 2012; 50(4):470–8. [PubMed: 22108440]

McElree B. Attended and Non-Attended States in Working Memory: Accessing Categorized Structures. Journal of Memory and Language. 1998; 252(38):225–252.

McMenamin BW, Shackman AJ, Maxwell JS, Bachhuber DRW, Koppenhaver AM, Greischar LL, Davidson RJ. Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. NeuroImage. 2010; 49(3):2416–32. [PubMed: 19833218]

Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science. 2008; 319(5869):1543–6. [PubMed: 18339943]

Nee DE, Jonides J. Neural correlates of access to short-term memory. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(37):14228–33. [PubMed: 18757724]

Newman EL, Norman Ka. Moderate excitation leads to weakening of perceptual representations. Cerebral Cortex. 2010; 20(11):2760–70. [PubMed: 20181622]

Oberauer K. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27(4):948.

Oberauer K. Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28(3):411–421.

Oberauer K. Control of the contents of working memory--a comparison of two paradigms and two age groups. Journal of experimental psychology Learning, Memory, and Cognition. 2005; 31(4):714–28.

Oberauer K, Gothe K. Dual-task effects in working memory: Interference between two processing tasks, between two memory demands, and between storage and processing. European Journal of Cognitive Psychology. 2006; 18(4):493–519.

Olivers CNL, Peters J, Houtkamp R, Roelfsema PR. Different states in visual working memory: when it guides attention and when it does not. Trends in Cognitive Sciences. 2011; 15(7):327–34. [PubMed: 21665518]

Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience. 2011; 2011:156869. [PubMed: 21253357]

Palva S, Kulashekhar S, Hämäläinen M, Palva JM. Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. The Journal of Neuroscience. 2011; 31(13):5013–25. [PubMed: 21451039]

Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage. 2009; 45(1 Suppl):S199–209. [PubMed: 19070668]

Poldrack, Ra. Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences. 2006; 10(2):59–63. [PubMed: 16406760]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. Science. 2005; 310(5756):1963–1966. [PubMed: 16373577]

Simanova I, van Gerven M, Oostenveld R, Hagoort P. Identifying object categories from event-related EEG: toward decoding of conceptual representations. PloS One. 2010; 5(12)

Sternberg S. Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist. 1969; 57(4):421–457. [PubMed: 5360276]

Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. Nature. 2005; 438(7067):500–3. [PubMed: 16306992]
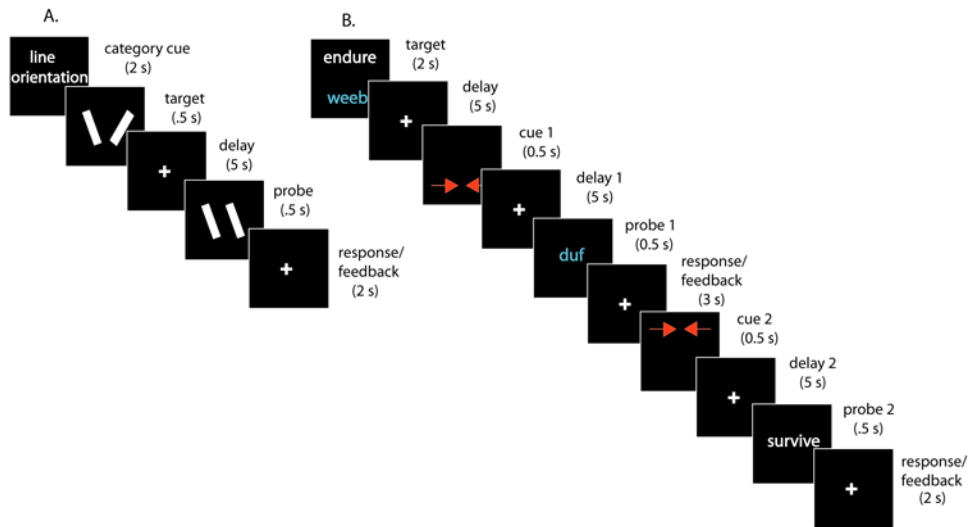
Figure 1. Behavioral task schematic
In Phase 1, participants performed a short-term memory task (A) for items from one of three categories: line segments (visual memory), vowel sounds in pseudowords (phonological memory), or word meanings (semantic memory). In Phase 2, participants performed a two-item short term memory task (B) for items drawn from the same three categories – the two items present on each trial were always from different categories. A cue indicated the behaviorally relevant item before each of the two probes, with half of the trials cuing the same item twice (repeat trials) and half cuing the two items in turn (switch trials, as in the example diagrammed above).
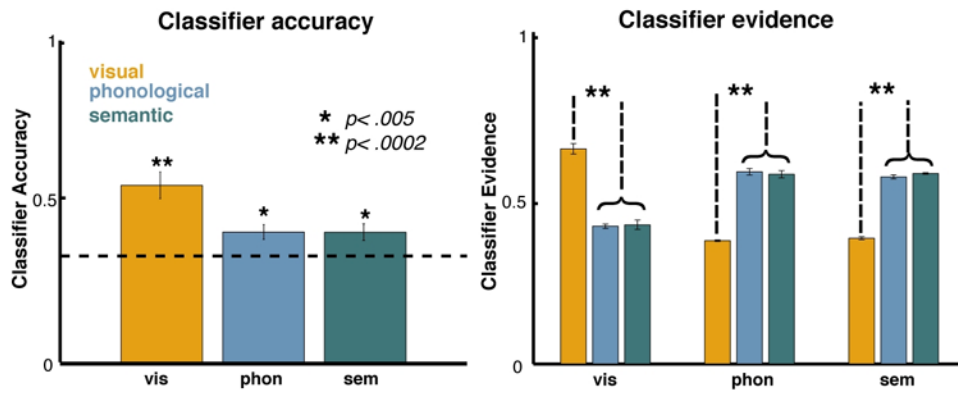
Figure 2. Phase 1 classifier cross-validation
Classifier accuracy (percent of classifications in which evidence for the correct trial type was highest) is plotted separately for each trial type (visual, phonological, semantic). Classifier evidence for each category (color-coded) is also plotted for each trial type. *p<0.005, **p<0.0002
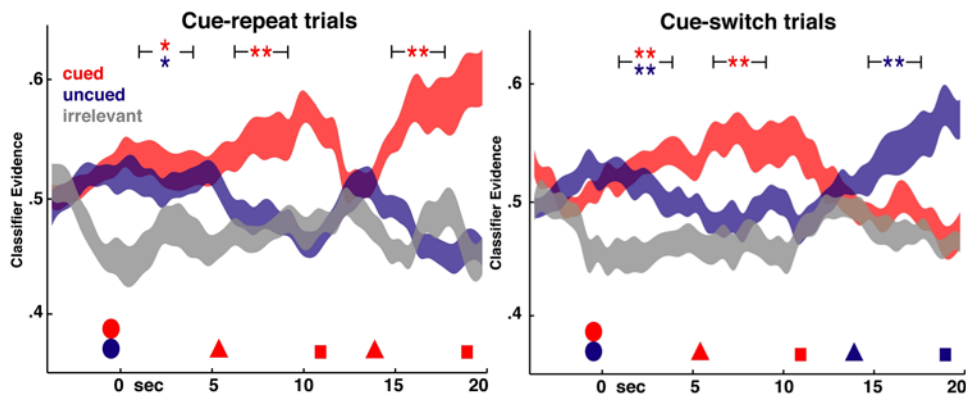
Figure 3. Phase 2 decoding of cue repeat and cue switch trials
Mean classifier evidence values are plotted separately for cue repeat (left) and cue switch trials (right). The visual, phonological and semantic categories were collapsed across all trials into new categories defined by the first cue: cued (red) is the category indicated by the first cue, uncued (blue) is the category of the other stimulus which is not selected by the first cue, and absent (gray) is the category not present on that trial. Time is represented on the horizontal axis, with stimulus presentation (circles) from -2 to 0 s, the first cue (triangle) at 5 s, the first probe (square) at 10.5 s, the second cue (triangle) at 14 s and the second probe (square) at 19.5 s. Width of the ribbons represents mean ± SEM, with spline interpolation to create continuous curves. Width of the brackets surrounding significance markers denotes extent of delay period used for statistical analysis: see methods for details. *p < 0.05, **p < 0.005.
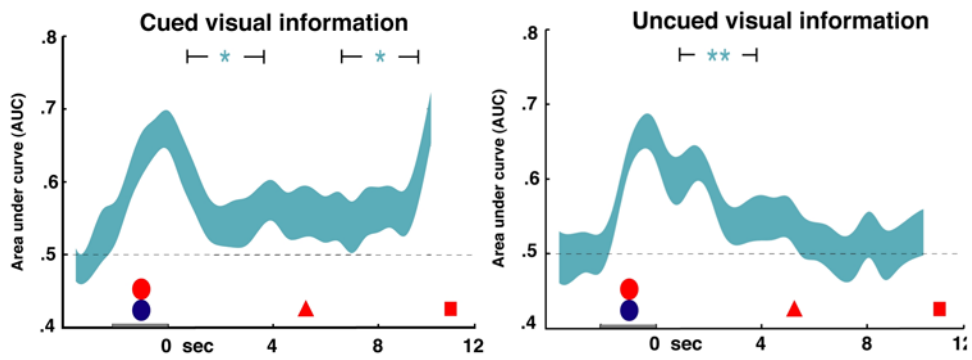
Figure 4. Phase 2 cross-validation decoding of cued and uncued visual information
Cross-validation decoding was performed within the Phase 2 task. Classifier sensitivity to the visual category (AUC) is plotted for each k-fold cross validation analysis time window, averaged across subjects. The width of the ribbon corresponds to the standard error of the AUC, calculated across subjects. Width of brackets surrounding significance markers indicates extent of delay period used for statistical analysis – see methods for details. Time is represented on the horizontal axis, with stimulus presentation (circles) from -2 to 0 s, the first cue (triangle) at 5 s, and the first probe (square) at 10.5 s. *p<0.05, **p<0.005.