Year: 2013

Decoding attended information in short-term memory: an EEG study

Larocque, Joshua J ; Lewis-Peacock, Jarrod A ; Drysdale, Andrew T ; Oberauer, Klaus ; Postle, Bradley R

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 fMRI study has suggested that this activity may be more related to attention than to retention. Specifically, a multivariate pattern analysis failed to find evidence that information that was outside the focus of attention, but nonetheless in STM, was retained in an active state. Here, we replicate and extend this finding by querying the neural signatures of attended versus unattended information within STM with electroencephalograpy (EEG), a method sensitive to oscillatory neural activity to which the previous fMRI study was insensitive. We demonstrate that in the delay-period EEG activity, there is information only about memory items that are also in the focus of attention. Information about items outside the focus of attention is not detectable. 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.

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

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

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 fMRI study has suggested that this activity may be more related to attention than to retention. Specifically, a multivariate pattern analysis failed to find evidence that information that was outside the focus of attention, but nonetheless in STM, was retained in an active state. Here, we replicate and extend this finding by querying the neural signatures of attended versus unattended informa-

tion within STM with electroencephalograpy (EEG), a method sensitive to oscillatory neural activity to which the previous fMRI study was insensitive. We demonstrate that in the delay-period EEG activity, there is information only about memory items that are also in the focus of attention. Information about items outside the focus of attention is not detectable. 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 function of STM and working memory, is central to much of human behavior. Decades of theoretical (Baddeley & Hitch, 1974; Hebb, 1949) and physiological (Vogel, McCollough, & Machizawa, 2005; Curtis & D'Esposito, 2003; Fuster, 1973) 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 working memory 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.

Among the models that explicitly consider the relationship between attention and memory, there are several that explain STM and working memory as emerging from an interaction between attention and long-term memory (LTM; Oberauer, 2002; Cowan, 1988). In these models, memory representations being actively utilized to guide behavior occupy a central, capacity-limited component [called "the focus of attention" (Cowan, 1988) or "the region of direct

access" (Oberauer, 2002); here, we use the term "focus of attention"]. Information outside the focus of attention which still needs to be remembered may be transiently retained as activated LTM, which is distinct from the immense network of latently stored LTM. An important caveat is that the meaning of the term "activated" in these cognitive models need not correspond to the meaning of activation in a neurophysiological sense (that is, elevated neuronal firing and/or elevated oscillatory power). For this manuscript, in which we investigate levels of neural activity underlying various states of "activation" in a theoretical STM model, we will use the term "prioritized" instead of "activated" when discussing theoretical models and reserve "activation" for neurophysiology.

The crucial aspect of these prioritized LTM models is that they dissociate the retention of information—which can be accomplished in any of the prioritized states of LTM—from attention to information—which is a capacity-limited resource that can be applied only to a small subset of highly prioritized representations. A recent study (Lewis-Peacock et al., 2012) explored the neural correlates of these different hypothesized states of STM by acquiring fMRI data while participants performed a delayed-recognition task that allowed for the dissociation of attended memory items (AMI) from unattended memory items (UMI). 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 that permits information to be extracted from high-dimensional data sets such as fMRI

[1]University of Wisconsin, Madison, [2]Princeton University, [3]University of Zurich

(Pereira, Mitchell, & Botvinick, 2009; Polyn, Natu, Cohen, & Norman, 2005). The authors first 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. After an initial delay period, the memory item that was to be probed first was retroactively cued (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 neural 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 & 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 neural activity related to 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 working memory, the virtual totality of which have been interpreted in the context of the signal intensity assumption. One important caveat, however, is that these new findings were derived solely from fMRI data. The possibility exists that UMIs are retained by an active neural mechanism to which the BOLD signal of fMRI is not sensitive. To address this concern, this study was designed to replicate the critical features of Lewis-Peacock et al. (2012), with the exception that we measured neural activity with the 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 (Johnson, Sutterer, Acheson, Lewis-Peacock, & Postle, 2011; Palva, Kulashekhar, Hämäläinen, & Palva, 2011; Jensen & Tesche, 2002). Furthermore, EEG is more temporally precise than fMRI, which permits more nuanced interpretations of the time course of delay-period activity. We applied MVPA to delay-period spectral data 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 to magnetoencephalography (MEG; Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010).

As was the case with Lewis-Peacock et al. (2012), in this 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 active trace account of STM (Curtis & D'Esposito, 2003; Baddeley & Hitch, 1974) and with previous fMRI studies (Nee & Jonides, 2008; Lepsien & Nobre, 2007). 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, we would conduct a second analysis using pattern classifiers that were trained and tested solely on the two-item retention intervals, utilizing a k-fold cross-validation approach.

A second goal of our experiments was to further explore the dynamics and neural bases of the removal of information from the focus of attention. The work presented here draws on prior work using retrocuing to experimentally manipulate the effective memory load. Because RTs increase with the size of the currently maintained memory load (Sternberg, 1969), RTs can be used as a gauge for memory load. Oberauer and colleagues have shown with lists of words (Oberauer, 2001), numbers (Oberauer, 2002), and mixed-domain lists (Oberauer & Göthe, 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). At short CPIs (100–500 msec), RTs 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 sec), 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 the focus of attention before the uncued items had been removed. This logic has been used to estimate that the removal process takes approximately 1 sec per item to complete. The items were removed from the focus of attention into a higher-capacity secondary storage layer in STM (which we will refer to as prioritized [originally "activated"] LTM; Oberauer, 2002; 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 removal process. The first of these behavioral experiments (Experiment 2a) used a task with stimuli drawn from the same category on each trial to establish a baseline estimate of removal time for these stimuli. The next experiment (2b) differed from the first only in that the two memory sets were drawn from different categories on each trial to examine cross-category effects of the removal process. The final behavioral experiment (2c), 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 of the removal process with an independent neural estimate. 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 removal effect, and therefore, we predicted that the deactivation time estimate from the EEG experiment should correspond to the removal 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 removal 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 AMI versus UMI in STM. To interpret the EEG data from the Phase 2 task, we applied pattern 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 removal 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 removal of UMIs exerts a residual performance cost, behavioral Experiment 2c compared RTs on trials that required the removal of a UMI set versus trials that did not (i.e., trials that consisted of only a single set of AMIs).

Experiment 1: Participants

Eighteen participants (12 women, average age = 22 years) were recruited from University of Wisconsin, 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 (Figure 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-sec presentation of the instruction to "Blink now." Next, a fixation cross appeared for 2 sec, followed by a category cue (2 sec), target stimulus (0.5 sec), delay period (5 sec), probe stimulus (0.5 sec), response period (1.5 sec), and feedback (1 sec). 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 (because the Phase 2 task, using identical stimuli, did not provide category cues), 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
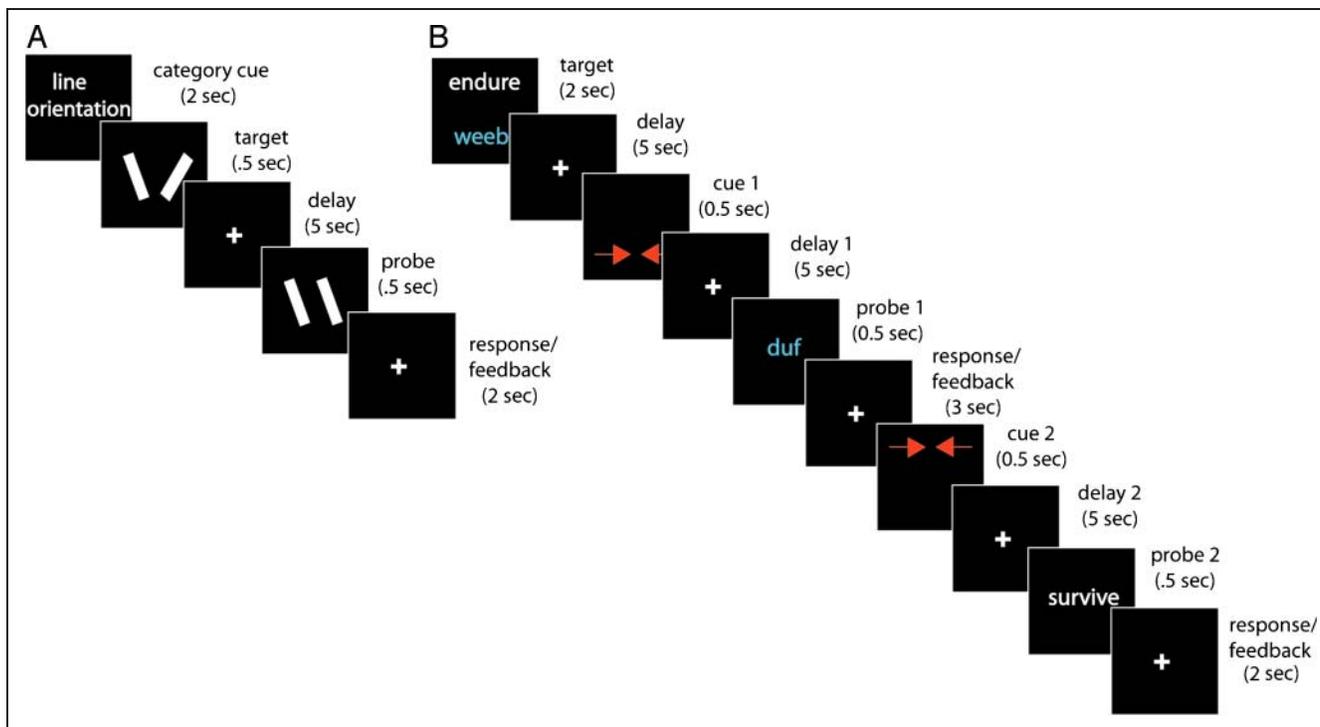
Figure 1. Behavioral task schematic. In Phase 1, participants performed an STM task (A) for items from one of three categories: line segments (visual memory), vowel sounds in pseudowords (phonological memory), or word meanings (semantic memory). In Phase 2, participants performed a two-item STM task (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).

nonmatch. Participants had a total of 2.5 sec 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 nonmatch probe stimuli, one of the line segments differed in orientation from the target stimulus by at least 30°. For pseudowords, to be a match, the vowel sound in the probe had to match the vowel sound of the target. For words, to be a match the probe word had to be synonymous with the target word. These different matching criteria were designed to elicit different encoding strategies, so that different mental codes would be employed for each of the three categories (visual, phonological, and semantic, respectively). The rationale for this design choice was to increase the chances of distinguishing the three trial categories from one another with MVPA, so that separate readouts for multiple items would be possible. 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 (Figure 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 sec), two inward-facing red arrows appeared at the top or bottom position (each with $p = .5$) to cue which item would be tested by the first memory probe. After a second delay period (5 sec) and the first probe, participants 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 participants could not simply forget the item that was initially uncued, because there was a 50% probability that this item would be the target of the second probe. Feedback was provided after each response as in the Phase 1 task. Participants 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 (Psychology Tools, Inc., Pittsburgh, PA) projected on an LCD

monitor situated approximately 24 in. in front of the participant. RTs 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 participant interpreted the vowel sound of a pseudoword differently than we intended, that participant 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 (Eugene, OR). The impedance of each electrode was kept below 75 kΩ, and the sampling rate was 500 Hz.

EEG Preprocessing

The data were processed off-line using the EEGLAB (Delorme & Makeig, 2004) and Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) toolboxes in MATLAB (Mathworks, Natick, MA). First, 72 channels along the face, ears, and neck were removed. These channels are most susceptible to noise because of poor contact and non-neural signals because of 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 1 per subject, 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 intertrial intervals. 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 participant. 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 to 20 Hz and every other integer from 22 to 50 Hz was used with a fixed, Hanning-tapered window of 0.5 sec. This transform resulted in spectral power values at each of 34 frequencies and 185 channels, sampled every 0.5 sec, 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-sec 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

MVPA was performed in MATLAB using the EEG Analysis Toolbox (Morton et al., in press; 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 four 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 preclassification 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 sec of the delay period from each trial was used. The first (0–0.5 sec) and last (4.5–5.0 sec) time points were excluded 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 sec window; therefore, the data point centered at 0.75 sec averaged over data from −0.5 to 2 sec (thus including the stimulus), and the data point centered at 4.25 sec was an average of data from 3 to 5.5 sec (thus including the probe). We included these points 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 sec 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 sec 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 16 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" versus "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 sec each, 1 sec 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 that 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 postcue delay period, this sliding window analysis would permit the different active representations to be classified separately (i.e., based on separate training data), and there-

fore, 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 (Newman & Norman, 2010; Fawcett, 2006). An AUC of greater than 0.5 indicates sensitivity to the category of interest. Statistical significance of sensitivity was assessed with one-tailed, one-sample t tests of AUC values averaged across the middle 3 sec of each delay period versus 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 removed 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 removal time in this first postcue 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 (baseline) was performed at each time point; the point at which the uncued evidence was no longer significantly greater than baseline (at $p < .05$) 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: Participants

For Experiment 2a, 16 participants (right-handed; native English speakers; 13 women; ages 18–27 years) were recruited from the undergraduate and medical campuses of the University of Wisconsin, Madison. An additional 22 participants (right-handed; native English speakers; 13 women; ages 18–21 years) were recruited for Experiment 2b, and separate group of 21 participants (right-handed; native English speakers; 17 women; 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 Removal

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 (Figure A1A). The rationale was to systematically vary the timing between the retrocues and the recognition probe 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 sec) 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 (two words/pseudowords or three line segments; high load), and set sizes were varied orthogonally for the two sets. After 0.7 sec of a blank screen, a retrocue, two inward-facing red arrows, appeared at the top or bottom position (each with $p = .5$) 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 sec and was followed immediately by a recognition probe. Participants 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 sec, 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 counterbalanced. Each block represented one combination of first cue duration and second cue duration, and blocks were presented in one of four pseudorandom orders for each subject. All blocks of one stimulus category were completed before advancing to the next category. The second session was completed between 1 day and 2 weeks after completion of the first.

Experiment 2b: Mixed-category Stimulus Removal

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 (Figure A1B). 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 pseudowords) were completed before advancing to the next combination.

Experiment 2c: Residual Effects of Stimulus Removal

Trials were configured as in Experiment 2a, except for three modifications: (1) only nouns were used as stimuli, (2) recognition probes were simple match/nonmatch comparisons rather than synonym judgments, and (3) memory items were presented either as two sets of two (four items in total) or as one set of two items (Figure A1C). The timing of cues and probes was identical for both conditions, but there were no UMIs to remove 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 sec). Trials from the six conditions of this design were presented in random order. The experiment was completed in one session lasting 50 min.

Behavioral Experiments: Data Collection

All experiments were implemented with E-Prime software version 2.0. RTs 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 an RT was at least three standard deviations above or below the mean RT 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% ± 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 $2 \times 2 \times 6$ repeated-measures ANOVA with the factors of Probe Order (first, second), Trial Type (cue repeat, cue switch), and Stimulus Type (each of the six possible pairwise combinations), main effects of Probe Order, $F(1, 17) = 11.89$, $p = .003$, Trial Type, $F(1, 17) = 5.2, p = .036$, and Stimulus Type, $F(5, 85) = 5.31, p = .0003$, 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 order by stimulus type was significant, $F(5, 85) = 4.88, p = .0006$. The three-way interaction was nonsignificant. An identical repeated-measures ANOVA for RT on correct trials was performed, and significant main effects of Trial Type, $F(1, 17) = 6.42, p = .021$, and Stimulus Type, $F(5, 85) = 3.36$, $p = .008$, were noted. Participants were faster to respond to repeat trials (first probe, 917 msec, $SEM = 42$ msec; second probe, 846 msec, $SEM = 35$ msec) than to switch trials (first probe, 909 msec, $SEM = 41$ msec; second probe, 933 msec, $SEM = 43$ msec). Pairwise comparisons with Bonferroni correction indicated that participants were much faster at visual–semantic trials (853 msec, $SEM = 22$ msec) than they were at phonological–visual (927 msec, $SEM = 28$ msec), phonological–semantic (927 msec, $SEM = 24$ msec), or semantic–phonological (931 msec, $SEM = 24$ msec). The semantic–phonological RTs were also significantly slower than the visual–phonological responses (862 msec, $SEM = 26.3$ msec). Of the interactions, Probe Order × Trial Type, $F(1, 17) = 11.9, p = .003$, Probe Order × Stimulus Type, $F(5, 85) = 3.46, p = .0005$, and

Trial Type × Stimulus Type, $F(5, 85) = 3.87, p = .003$, were significant.
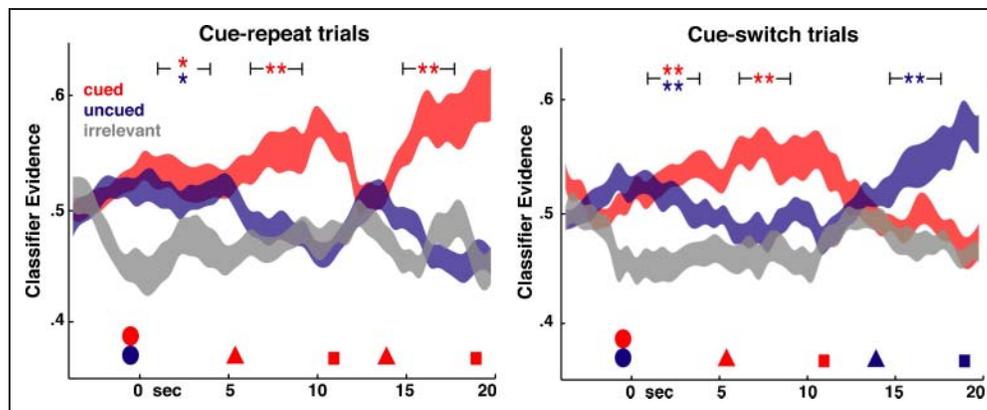
MVPA Results: Phase 1 Training

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 (Figure 2). The overall accuracy was 45.3%, and classification accuracies within each stimulus category were all significantly above chance (category accuracy ± SEM: visual $55.0 \pm 4.2\%$, phonological $40.4 \pm 2.3\%$, semantic $40.4 \pm 2.6\%$). Classifier performance for two participants was below chance (31%, 30%), and so these participants' data were excluded from subsequent analyses that used classifiers trained on Phase 1 data.

In examining the category level performance (Figure 2, left), 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 (Figure 2, right). 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

Figure 2. Phase 1 classifier cross-validation. Classifier accuracy (percentage 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. *$p < .005$, **$p < .0002$.
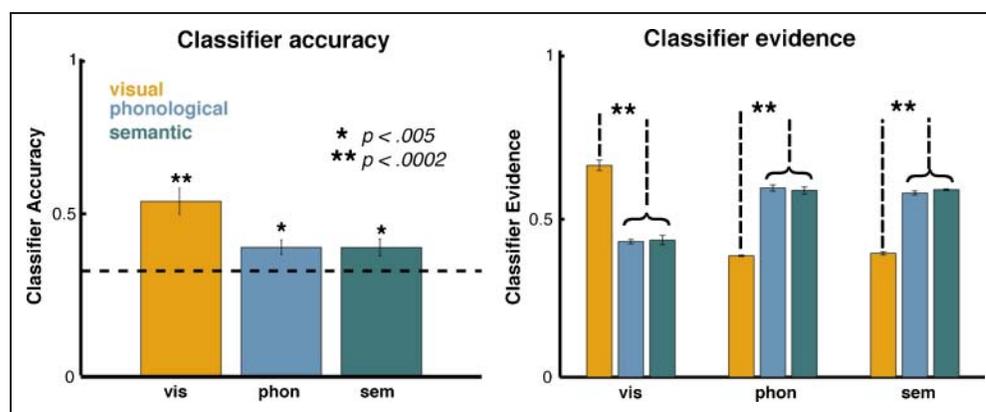
Figure 3. Phase 2 decoding of cue repeat and cue switch trials. Mean classifier evidence values are plotted separately for cue repeat (left) and cue switch trials (right). The visual, phonological, and semantic categories were collapsed across all trials into new categories defined by the first cue: 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 sec, the first cue (triangle) at 5 sec, the first probe (square) at 10.5 sec, the second cue (triangle) at 14 sec, and the second probe (square) at 19.5 sec. 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. *$p < .05$, **$p < .005$.

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 (Figure 3). The evidence during the first delay period after the stimulus presentation clearly differentiated the categories that were present from the one that was absent ($p < .005$, $t > 3.1$ for both categories on cue-switch trials; $p < .05$, $t > 1.9$ 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: $p = .09$, $t = 1.4$; repeat trials: $p = .25$, $t = 0.68$), whereas evidence for the cued item's category was still clearly above baseline ($p < .005$, $t > 3.5$ for both switch and repeat trials). In the cue repeat trials, the same pattern was evident after the second cue (cued vs. baseline, $p = .002$, $t = 3.4$; uncued vs. baseline, $p = .8$, $t = −1.01$). 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 ($p = .002$, $t = 3.4$) and the previously cued category dropped to baseline ($p = .09$, $t = 1.4$).

As a final attempt to detect any evidence for the UMI, we focused on cue switch trials in which visual information was initially uncued, and subjects were able to correctly respond to the second probe (which was to the initially uncued line orientation stimulus). These trials provided evidence for a line orientation UMI (the category for which classification success was highest) when there was behavioral evidence for successful retention of the stimulus. We compared this UMI evidence estimate to the baseline of visual category evidence, taken from those trials without any visual information (phonological–semantic trials). A one-tailed, paired *t* test comparing these evidences during the delay period did not show a significant difference ($p = .09$, $t = 1.4$), confirming the result from the original analysis.

We also used the train on Phase 1/test on Phase 2 data to derive an estimate of the removal 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 $p < .05$ level was 1.25 sec after the cue onset (i.e., the time window centered 1.25 sec 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 Figure 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 precue delay period, the classifier was clearly sensitive to the presence of visual information ($p < .05$, $t > 1.8$ for both analyses). After the cue, the classifier was still sensitive to the presence of visual information when it was cued ($p = .002$, $t = 3.3$), but not when it was uncued ($p = .4$, $t = 0.21$).

Experiment 2 Behavioral Results

A summary of the accuracies, RTs, and omnibus ANOVAs can be found in Tables A1 and A2 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 versus small uncued memory sets, but that at longer cue durations, this effect would dissipate. The ANOVA supports our hypothesis by demonstrating a significant CPI × uncued set size interaction, $F(2, 14) = 17.4$, $p < .001$ for Experiment 2a; $F(2, 14) = 3.62$, $p = .047$ for Experiment 2b (Table A1). Critically, planned t tests revealed that, after a cue duration of 2 sec for same-category stimulus sets (Experiment 2a) and 4 sec for mixed-category stimulus sets (Experiment 2b), RTs were not statistically different for trials with large uncued sets versus trials with small uncued sets (see Figure A2). That is, after 4 sec in both experiments, RTs to probes became insensitive to the number of currently irrelevant memory items. Similar results for Experiment 2c are shown (Figure A2 and Table A3) for each cue duration comparing high-load versus no-load conditions for the uncued memory set. After 4 sec, the RT for relevant stimuli was insensitive to whether irrelevant stimuli had also been presented at the beginning of the trial.

DISCUSSION

This 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 (Olivers, Peters, Houtkamp, & Roelfsema, 2011; Oberauer, 2002; Cowan, 1988). 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. In addition, they applied MVPA to time-domain ERPs, 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.

Results from the initial precue delay period are particularly interesting because the decoding analyses indicated that information was present for both items being retained. This result is interesting in light of recent electrophysiological work in primates suggesting that, in the pFC, neuronal activity underlying two-item retention was not predictable from activity associated with retaining

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 participants. The width of the ribbon corresponds to the standard error of the AUC, calculated across participants. 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 sec, the first cue (triangle) at 5 sec, and the first probe (square) at 10.5 sec. *p < .05, **p < .005.
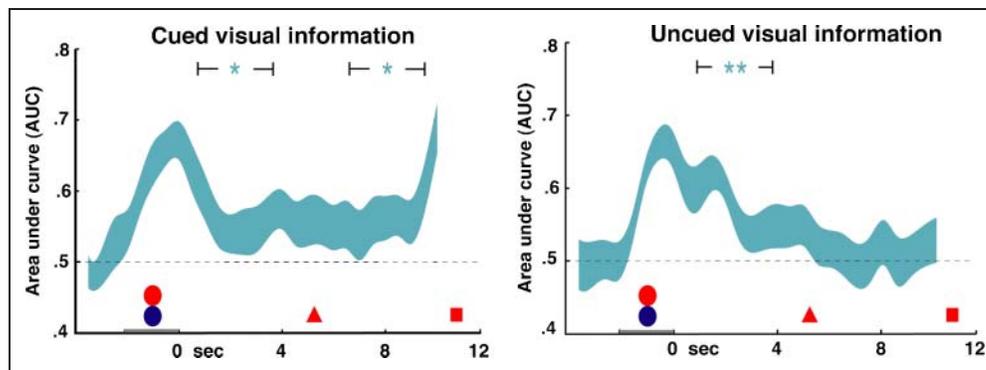
either item alone (Warden & Miller, 2007). Our findings with fMRI (Lewis-Peacock et al., 2012) and EEG, in contrast, are consistent with findings from extracellular data indicating that a neuronal representation can be stable at the population level although an individual neuron's contribution may not be stable (Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Crowe, Averbeck, & Chafee, 2010). Our results from the precue delay period are also intriguing because, during this part of the task, there was no experimental control of the focus of attention, leaving open the question of how attention interacted with the two items in memory. There are several accounts that may explain the successful decoding of both items. 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, consistent with the theories of Oberauer (2002) and Cowan (1995). A second possibility is that the focus of attention is limited to a single item (McElree, 1998), and 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. 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.

Potential objections to the interpretation of the primary analysis and the previous fMRI results (Lewis-Peacock et al., 2012) arise because of the fact that the classifiers used to decode the two-item memory task (Phase 2 task) were trained on a task (the Phase 1 task) in which subjects needed to remember only one item. First, it is possible that the patterns of neural activity underlying one-item retention are qualitatively different from those underlying two-item retention. This concern can be addressed, in both this

study and Lewis-Peacock et al. (2012), with the decoding results from the initial, precue delay period of the Phase 2 task. In this delay period, 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. A second potential objection is that the patterns of neural activity associated with AMIs differ from those associated with UMIs. In the Phase 1 task, the single memory item was presumably retained within 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) and not to UMIs. If this were the case, then there would still be the possibility of a separate neural signal that corresponds to retained but unattended information (UMIs). To address this concern, we performed a follow-up analysis using classifiers trained and tested on Phase 2 data, using a *k*-fold cross-validation procedure (Figure 4). In this analysis, classifiers were trained to distinguish trials in which visual information was present and uncued versus 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 evident in the delay-period activity only when visual information was in the focus of attention.

A key feature of this study was the higher temporal resolution afforded by EEG compared with 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 removal of items from the focus of attention takes place over a timescale of approximately 1–2.5 sec (Oberauer, 2001, 2005). Our own estimates of this removal process,

obtained from behavioral paradigms (Experiments 2a and 2b) 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 sec, depending on whether the memory sets were drawn from the same or from different categories.

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 sec. To compare these results with the behavioral estimates of removal 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) versus low-load (single UMI) conditions. Such relative estimates reflect the time required to remove the larger UMI sets, and therefore, we must correct for the number of items in these sets to derive an estimate for removing a single UMI. In Experiment 2b, the influence on RTs of the UMI set size had disappeared by 4 sec. (Note that we did not test CPIs in between 2 and 4 sec, thus 4 sec is likely an overestimation of the removal time.) Our high-load UMI set sizes were a mixture of two (for words and pseudowords) and three (for line segments) items, so we can approximate a high-load set size of 2.33 items. Dividing 4 sec (removal time for the high-load UMI set) by 2.33 items yields a per-item estimate of approximately 1.7 sec, which agrees relatively well with our EEG-derived neural estimate of 1.25 sec. The foregoing estimate relies on the assumption that memory items are removed serially; to compare the removal time courses without relying on this assumption, we also fitted an exponential decay function to the set-size effect at various CPIs (for the behavioral data) and to the classifier evidence for the UMI at postcue time points. With this approach, there was excellent agreement between the rate constants for the behavioral ($r = -0.386 \text{ sec}^{-1}$, 95% CI -4.38 to 3.6 sec^{-1}) and EEG (-0.424 sec^{-1}, 95% CI -0.684 to -0.164 sec^{-1}) exponential fits. A limiting factor in this comparison is that only three time points were available for the exponential fit to the behavioral data, which markedly increased the uncertainty in the rate constant value.

Independent of the relation of the behavioral results to the EEG estimate of the removal time, the results from the three behavioral experiments provide unique insight into the removal process. Previously, removal times for lists of words and numbers had been estimated to be approximately 0.33–1.0 sec based on RTs to a probe appearing at various intervals after a retrocue (Oberauer, 2001, 2002, 2005). Our estimate of 1.7 sec per item is reasonably consistent with these results. Experiment 2b goes beyond previous demonstrations of removal from working memory by showing a cross-category effect of set size 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 at the shortest cue–probe interval.

This implies that memory items from different domains to some extent compete for the limited capacity of the focus of attention.

Our findings support theories of working memory that distinguish between a central, capacity-limited component (referred to here as the focus of attention) and a prioritized part of LTM to which memory contents not currently needed can be outsourced (Oberauer, 2005; Cowan, 1995). However, an irony of terminology arises in our suggestion that prioritized LTM (also referred to as "activated" LTM) 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 prioritized LTM leaves open the question of what might differentiate prioritized 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 that 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 that could create such a transient synaptic network.

Both the theoretical terminology (prioritized or "activated" LTM) and candidate molecular mechanism (potentiated synaptic weights in neurons without above-baseline activity) we have applied to UMIs suggest a close kinship with LTM. A valid theoretical question is whether the short-term retention of UMIs might be a type of LTM, rather than STM; a valid practical question is how this would affect the conclusions drawn from this study. First, the practical question; even if one conceptualizes the removal of attention from UMIs as effecting a coincident shift from STM to LTM, it is certainly the case that performing the task required the short-term retention of information no longer present in the environment. Therefore, our conclusion—that the short-term retention of information outside the focus of attention does not require an active trace—remains intact. To the theoretical question, if UMIs are to be considered a form of LTM, then STM would only contain items in the focus of attention. This would conflate the cognitive construct of STM with that of attention, propagating the conflation of neural substrates of STM and attention (which we argue has existed for decades) into the theoretical realm. This would erode the utility of the construct of STM. We prefer to retain the simple, behavioral definition of STM as the retention of information no longer present in the environment, acknowledging the models (Oberauer, 2002; Cowan, 1988) and experimental studies (Lewis-Peacock & Postle, 2008;

Ruchkin, Grafman, Cameron, & Berndt, 2003) suggesting that STM draws on the representations of LTM. (As an aside, for two decades, the "*n*-back" task has been the undisputed workhorse of human working memory research; Kirchner, 1958). It has never been suggested that the UMIs in that task (e.g., the previous item in the 2-back task) are outsourced to LTM.)

This 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 result converges with the analogous finding of Lewis-Peacock et al. (2012), which demonstrated that only items inside the focus of attention were maintained in a state detectable by MVPA in fMRI data. Taken together, these results suggest that an active trace of short-term retention is only present when attention is also allocated to the retained information.

APPENDIX
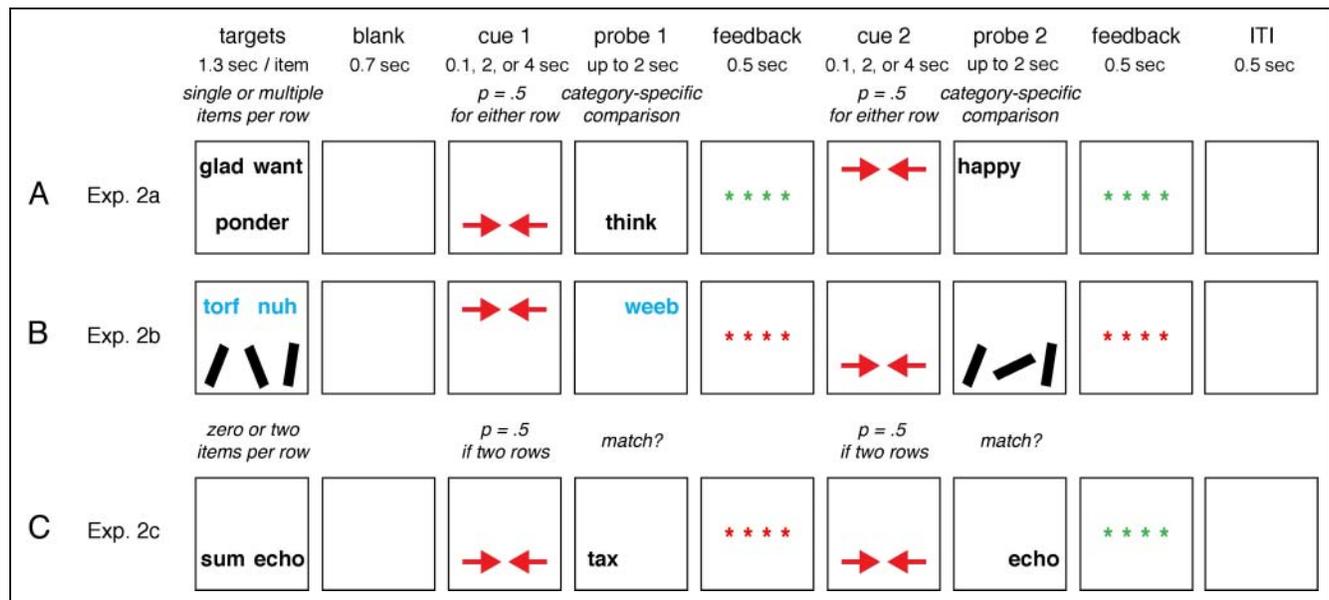
Figure A1. Task procedures. (A) Experiment 2a: Participants performed short-term recognition of words, pseudowords, or line segments. Both sets of stimuli for each trial were drawn from the same category. (B) Experiment 2b: Participants performed the same task but with each stimulus set drawn from a different category. (C) Experiment 2c: Participants performed short-term recognition for trials with either 2 sets of 2 stimuli each (4 total) or 1 set of 2 stimuli.
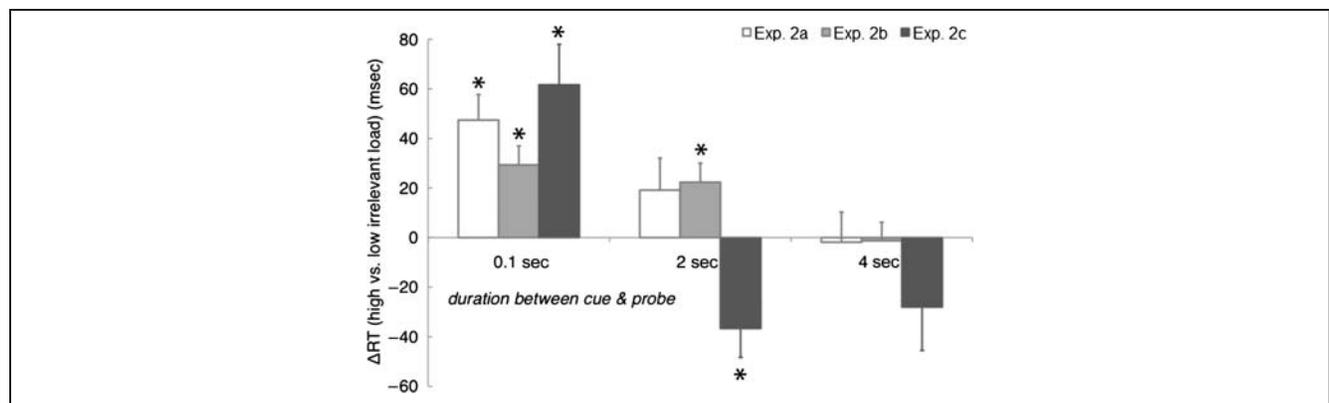
Figure A2. Removal effects on RT for behavioral experiments. Data are grouped along the horizontal access according to the length of time between the onset of the cue and the onset of the probe (0.1, 2, or 4 sec). The vertical axis shows the difference in RTs (from the first recognition probe) between conditions of high- and low-load of uncued (irrelevant) stimuli. For Experiments 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 Experiment 3, however, the high-load condition had two stimuli and the low-load condition had none (i.e., no removal of UMIs was required).

Table A1. ANOVA Results for RTs to the First Probe in Behavioral Experiments 2a, 2b, and 2c

Effect	F (df_effect, df_error)	p
A. Experiment 2a		
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)	.696
Stimulus Category × Cued Set Size	2.44 (2, 14)	.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)	.768
Cued Set Size × Uncued Set Size	6.20 (1, 15)	.025*
Stimulus Category × Cued Set Size × Uncued Set Size	0.66 (2, 14)	.531
CPI × Cued Set Size × Uncued Set Size	0.56 (2, 14)	.581
Stimulus Category × CPI × Cued Set Size × Uncued Set Size	0.46 (4, 12)	.766
B. Experiment 2b		
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)	.568
Stimulus Category Pair × Uncued Set Size	1.28 (2, 14)	.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)	.492
Stimulus Category Pair × Cued Set Size × Uncued Set Size	2.79 (2, 14)	.086
CPI × Cued Set Size × Uncued Set Size	4.78 (2, 14)	.021*
Stimulus Category Pair × CPI × Cued Set Size × Uncued Set Size	1.12 (4, 12)	.382
C. Experiment 2c		
CPI	0.012 (1, 20)	.913
Uncued Set Size	1.528 (2, 19)	.242
CPI × Uncued Set Size	16.086 (2, 19)	< .001*

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.

*$p_{(\text{one-tailed})}$ < .05.

Table A2. Pairwise *t* Tests for RTs to First Probe for Behavioral Experiments 2a and 2b

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

Data are mean RTs for each condition, with standard deviations in parentheses. CPI = cue–probeinterval, the duration between the onset of the cue and the onset of the probe stimulus.

**p < .01.

Table A3. Pairwise *t* Tests for RTs to First Probe, for Behavioral Experiment 2c

CPI	2/0	2/2	t(Paired)
0.1 sec	864.09 (142.37)	925.68 (141.30)	**(−) 3.76
2 sec	936.65 (152.35)	900.01 (154.05)	** 3.14
4 sec	918.52 (161.53)	890.53 (140.83)	1.60

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.

**p < .01.

REFERENCES

Baddeley, A. D., & Hitch, G. (1974). Working memory. In H. B. Gordon (Ed.), *Psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Crowe, D. A., Averbeck, B. B., & Chafee, M. V. (2010). Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *The Journal of Neuroscience, 30,* 11640–11653.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7,* 415–423.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Erickson, M. A., Maramara, L. A., & Lisman, J. (2009). A single brief burst induces GluR1-dependent associative short-term potentiation: A potential mechanism for short-term memory. *Journal of Cognitive Neuroscience, 22,* 2530–2540.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27,* 861–874.

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology, 20,* 606–612.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology, 36,* 61–78.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory* (p. 335). New York: Wiley.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience, 15,* 1395–1399.

Johnson, J. S., Sutterer, D. W., Acheson, D. J., Lewis-Peacock, J. A., & Postle, B. R. (2011). Increased alpha-band power during the retention of shapes and shape-location associations in visual short-term memory. *Frontiers in Psychology, 2,* 128.

Kirchner, W. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358.

Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral Cortex, 17,* 2072–2083.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24,* 61–79.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *The Journal of Neuroscience, 28,* 8765–8771.

Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia, 50,* 470–478.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language, 38,* 225–252.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R. W., Koppenhaver, A. M., Greischar, L. L., et al. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *Neuroimage, 49,* 2416–2432.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology, 100,* 1407–1419.

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science, 319,* 1543–1546.

Morton, N. W., Kahana, M. J., Rosenberg, E. A., Baltuch, G. H., Litt, B., Sharan, A. D., et al. (in press). Category-specific neural oscillations predict recall organization during memory search. *Cerebral Cortex.*

Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 14228–14233.

Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral Cortex, 20,* 2760–2770.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 948–957.

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

Oberauer, K. (2005). Control of the contents of working memory—A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 714–728.

Oberauer, K., & Göthe, K. (2006). 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, 18,* 493–519.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15,* 327–334.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 156869.

Palva, S., Kulashekhar, S., Hämäläinen, M., & Palva, J. M. (2011). Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. *The Journal of Neuroscience, 31,* 5013–5025.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage, 45(1 Suppl.),* S199–S209.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10,* 59–63.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310,* 1963–1966.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26,* 709–728; discussion 728-77.

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PloS ONE, 5,* e14465.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438,* 500–503.

Warden, M. R., & Miller, E. K. (2007). The representation of multiple objects in prefrontal neuronal delay activity. *Cerebral Cortex, 17(Suppl. 1),* i41–i50.