WORKING MEMORY

Reactivation of latent working memories with transcranial magnetic stimulation

Nathan S. Rose,[1,2]* Joshua J. LaRocque,[1,3] Adam C. Riggall,[1,5] Olivia Gosseries,[1,4] Michael J. Starrett,[1] Emma E. Meyering,[1] Bradley R. Postle[1,5]*

The ability to hold information in working memory is fundamental for cognition. Contrary to the long-standing view that working memory depends on sustained, elevated activity, we present evidence suggesting that humans can hold information in working memory via "activity-silent" synaptic mechanisms. Using multivariate pattern analyses to decode brain activity patterns, we found that the active representation of an item in working memory drops to baseline when attention shifts away. A targeted pulse of transcranial magnetic stimulation produced a brief reemergence of the item in concurrently measured brain activity. This reactivation effect occurred and influenced memory performance only when the item was potentially relevant later in the trial, which suggests that the representation is dynamic and modifiable via cognitive control. The results support a synaptic theory of working memory.

The ability to mentally retain information in an accessible state, to manipulate it, and to use it to guide behavior is a critical building block for cognition. It has long been assumed that the neural basis for this working memory (WM) ability is elevated and persistent neuronal firing (*1*). This assumption has been called into question by recent proposals that information can be held in WM via synaptic mechanisms that do not require sustained, elevated brain activity (*2*–*4*).

Building on theoretical frameworks that information can be held in WM in one of several states of activation (*5*, *6*), we recorded neural activity while participants performed a multistep task in which two items were presented as memoranda for each trial. A cue indicated which item would be tested by the impending recognition memory probe, followed by the probe, then by a second cue, and then a second probe (Fig. 1). There was equal probability following the first cue, but not the second, that the uncued item might be needed for an ensuing memory judgment. This procedure moves the uncued item into a different state than the cued item, which, by definition, is in the focus of attention (*7*). Cognitive theories refer to the intermediate state of this unattended memory item (UMI) as "activated long-term memory" (LTM) (*5*, *6*).

For experiment 1, multivariate pattern analysis (MVPA) showed evidence for an active representation of the UMI that dropped to baseline levels (Fig. 2) (*7*–*9*). This suggests that information in WM (but outside of focal attention) can be maintained in a latent state via mechanisms other than sustained, elevated activity. Although a similar drop-to-baseline pattern is observed when participants are instructed to drop information

[1]Department of Psychiatry, University of Wisconsin, Madison, WI 53706, USA. [2]Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA. [3]Neuroscience Training Program, University of Wisconsin, Madison, WI 53706, USA. [4]Coma Science Group, University of Liège, 4000 Liège, Belgium. [5]Department of Psychology, University of Wisconsin, Madison, WI 53706, USA.
*Corresponding author. Email: nrose1@nd.edu (N.S.R.); postle@wisc.edu (B.R.P.)

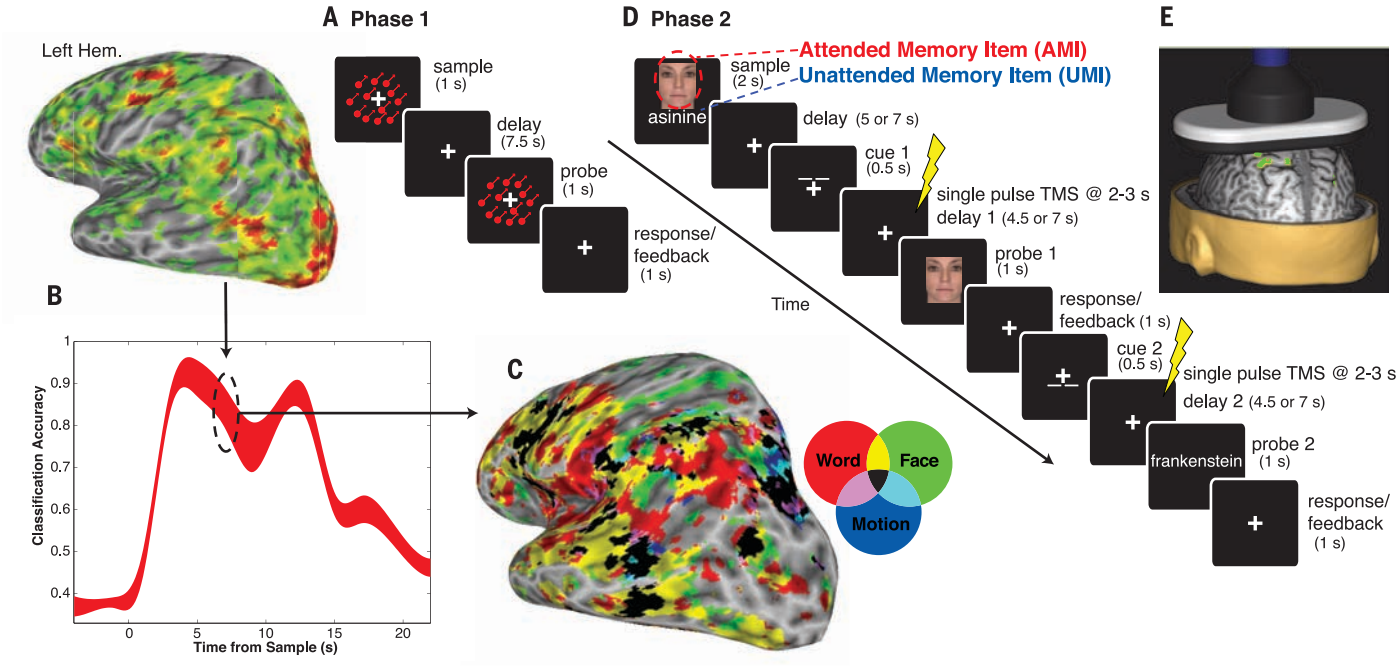


Fig. 1. General procedure. (**A**) In phase 1, functional magnetic resonance imaging (fMRI) data were acquired while participants performed a one-item delayed-recognition task for words, faces, or directions of motion; these data were used for multivariate pattern analysis (MVPA). (**B**) Classifiers trained on the delay period were used for subsequent analyses. For experiment 1, these classifiers were used to decode fMRI activity from phase 2 (Fig. 2). (**C** and **D**) For experiments 2 and 3, they were used in a whole-brain searchlight conjunction analysis to generate participant-specific maps of category-sensitive areas (C); nonoverlapping areas were used for transcranial magnetic stimulation (TMS) targeting in phase 2 (D). (**E**) In phase 2, single pulses of TMS were delivered during the postcue delay periods.

from WM (*10*, *11*), here the UMI remained in WM because, when so instructed by the second cue, participants accurately reactivated it and used it to evaluate the final probe (Fig. 2B).

In three additional experiments, we tested the hypothesis that if a UMI is encoded in a distributed pattern of synaptic weights and held in a state that is more accessible than trial-irrelevant information, the readout from a nonspecific burst of activity filtered through this network might reveal this latent representation (*2*) (fig. S1). This would be consistent with the idea that networks in the posterior cortex can be dynamically configured as matched filters to encode behaviorally relevant information (*3*, *4*, *12*, *13*).

For experiments 2 and 3, participants performed the phase 2 WM task (Fig. 1) while we recorded electroencephalography (EEG) and applied single-pulse transcranial magnetic stimulation (TMS) 2 to 3 s after the cue. For experiment 2, we targeted brain regions identified from the phase 1 MRI task as preferentially supporting MVPA decoding for one category, but not the other two. MVPA of the spectrally transformed EEG data from only the phase 2 task detected reliable evidence for an active representation of both memory items across the initial portion of the trial, until the onset of the first cue, at which point decoding accuracy remained elevated for the attended memory item (AMI) but dropped to the baseline for the UMI (*14*).

After a single pulse of TMS, there was a brief recovery of MVPA decoding of the UMI—a "reactivation effect"—before it returned to baseline and remained there while the cued item was tested [P = 0.01; Bayes factor (BF) = 3.64 against the null] (Fig. 3A). TMS affected neither broadband decoding of the AMI nor recognition memory judgments (fig. S4). When we analyzed bandpass-filtered data, the TMS reactivation effect was isolated to signal from the beta band (fig. S5) and was associated with a transient period of above-chance decoding performance for both the UMI and the AMI. The TMS reactivation effect was specific for information that was in WM on that trial, because above-chance MVPA performance, as assessed with the AUC (area under the curve) analysis, necessarily means that TMS did not activate a representation of the category that was irrelevant on that trial.

In experiment 2, we administered blocks of trials with TMS targeting one of the category-selective regions, but we varied, on a trial-by-trial basis, which category was the AMI and which was the UMI. Each block included trials for which the UMI belonged to the targeted region's preferred category, and trials for which it did not. A TMS reactivation effect was observed (Fig. 3B) whether or not TMS targeted the UMI's category-preferred region, although the effect was larger and more prolonged when it did (BF = 4.02 for targeted sites, 1.72 for nontargeted sites). This finding suggests that WM is supported by heightened connectivity between cortical networks that represent all trial-relevant information (AMI and UMI) relative to trial-irrelevant information (*15*, *16*).

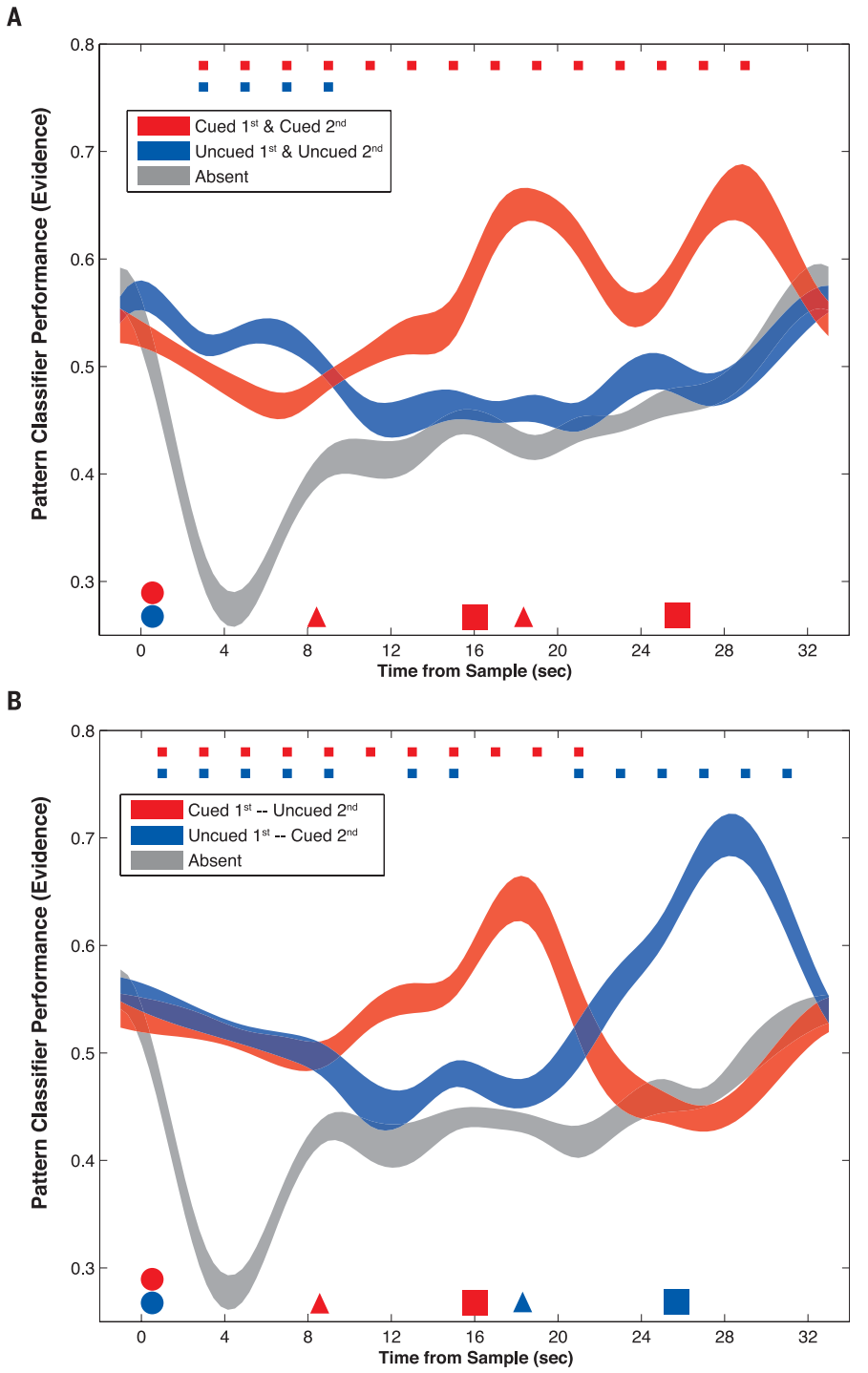


Fig. 2. Experiment 1 fMRI decoding (train phase 1, test phase 2): Classifier evidence as a function of an item's status, collapsed across stimulus category. After stimulus presentation (red and blue circles), delay-period classifier evidence for both items was elevated relative to the empirical baseline of evidence for the category that was not presented on that trial ("absent," gray). Upon presentation of the first cue (red triangle), evidence for the cued category (red) remained elevated, but for the uncued category (blue) dropped to baseline. (**A**) After the first probe (red square), on half the trials the second cue designated that the same item would be tested by the second probe, and evidence for the two categories remained the same relative to baseline. (**B**) When the second cue designated the previously uncued item, evidence for the two categories reversed for the remainder of the trial. Color-coded small squares at the top of each plot indicate P < 0.01; line width reflects SEM.

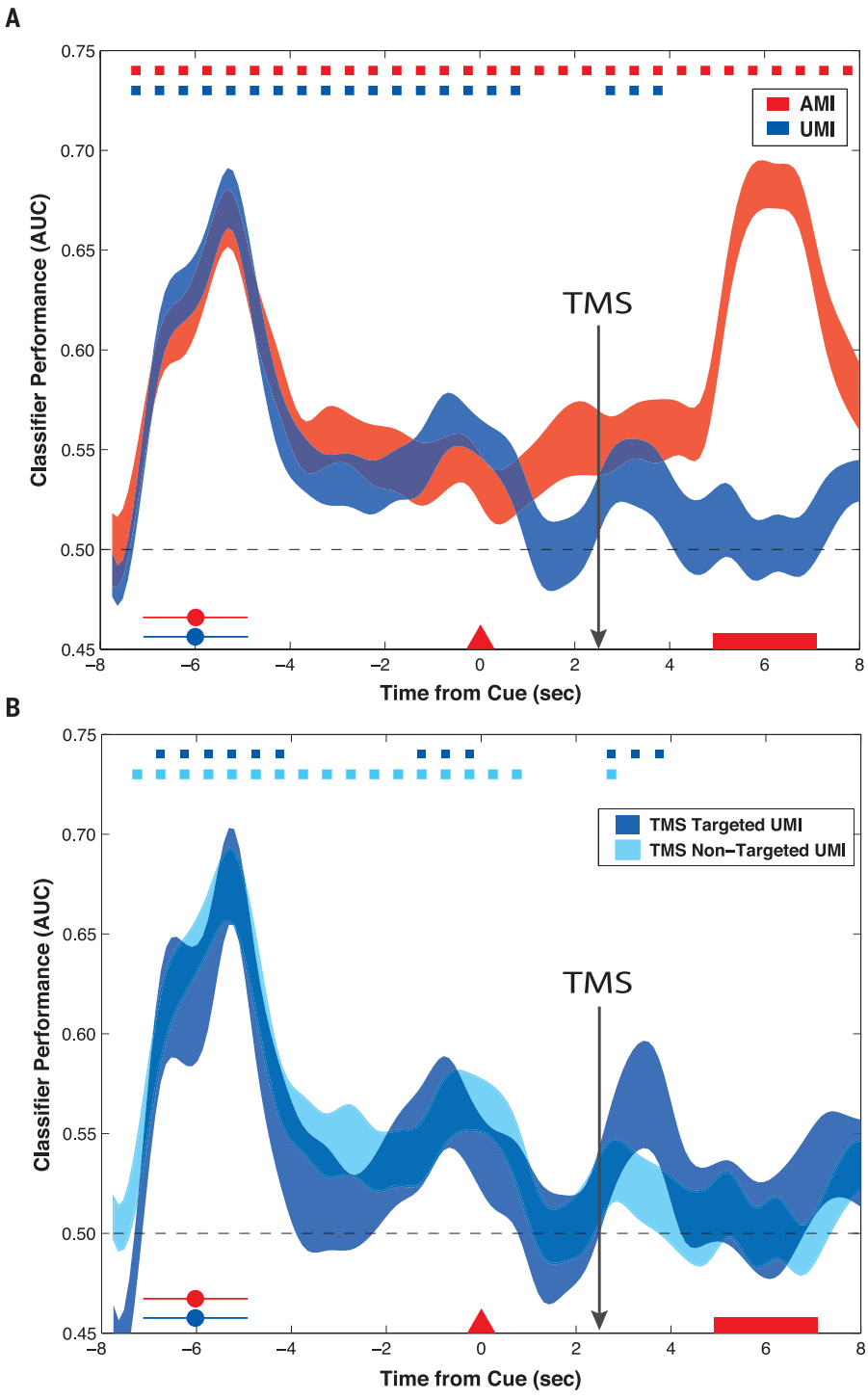


Fig. 3. Experiment 2 EEG decoding (train and test on phase 2 data): Classifier accuracy (area under curve, AUC) as a function of an item's status at the time of the first cue, collapsed across stimulus category. AUC reflects classifier sensitivity to discriminating between evidence for the AMI or UMI relative to the absent category. (**A**) Classification time series of the AMI and UMI upon stimulus presentation (red and blue circles), the first cue (red triangle), TMS, and first probe (red rectangle), averaged over N = 18 sessions, 2952 trials (decoding ends where the AMI and UMI switched on 50% of the trials). (**B**) Decoding UMIs as a function of whether TMS targeted that item's phase 1–defined region or a different category's region. Color-coded small squares at the top of each plot indicate $P < 0.05$; line width reflects SEM.

Retrocues that inform subjects that they can drop an item from memory result in a rapid loss of multivariate evidence for the no longer relevant item (*11*, *17*). Nonetheless, proactive interference from stimuli presented on previous trials indicates that the brain retains a residual trace of such recent, but no longer relevant, information (*18*). An important test of state-based models of WM is whether there is a functional distinction between UMIs (putatively held in a state of activated LTM) and dropped information (no longer in WM). In experiment 3, with a different group of participants, we also administered TMS after the second cue, after which the uncued item would no longer be relevant on the trial, and at which point it should have the same status as an irrelevant item. If the TMS reactivation effect is a consequence of an item being maintained in a privileged state, it should only be observed when that item is still potentially relevant for the trial. We also jittered the onset of TMS between 2 and 3 s after the cues (*14*) and standardized TMS by targeting the same region on every trial for all participants—an MVPA-defined region in the right precuneus known to be critical for the top-down control of visual attention (*19*) (Fig. 4A).

For the first half of the trial, the results from experiment 3 replicated those from experiment 2 (Fig. 4B), with a robust TMS reactivation effect for the UMI (BF = 9.8 against the null). For the delay period following the second cue, however, there was no evidence for significant decoding of the uncued item following the TMS pulse (BF = 3.4 in favor of the null). These results suggest that UMIs are maintained in a different state than are items that have been dropped from WM, and that the mechanisms that maintain latent representations in WM are dynamic and modifiable via cognitive control (*20*).

Because our design entails decoding at the category level, it does not rule out the possibility that the TMS reactivation effect reflects a general reinstatement of category context (*21*), rather than the temporary activation of the UMI itself. The idea that the representation of the UMI itself drives this effect would be strengthened by demonstrating that TMS can influence recognition memory decisions on this task. If the TMS reactivation effect reflects a temporary reinstatement of the UMI back into the focus of attention, participants should have more difficulty rejecting the UMI as a lure when probing their memory of the AMI.

In experiment 4, we presented recognition memory probes that matched the AMI on 50% of trials; of the 50% of nonmatch probes, 30% were drawn from the same category as the AMI, and a critical 20% matched the UMI (*14*). Participants were instructed to reject memory probes that did not match the AMI. Critically, only for the first probe was there an increased proportion of false alarms to the UMI for TMS relative to no-TMS trials (Fig. 3C, P = 0.01, BF = 3.48) (*14*).

Our results have important implications for the understanding of WM at many levels. They

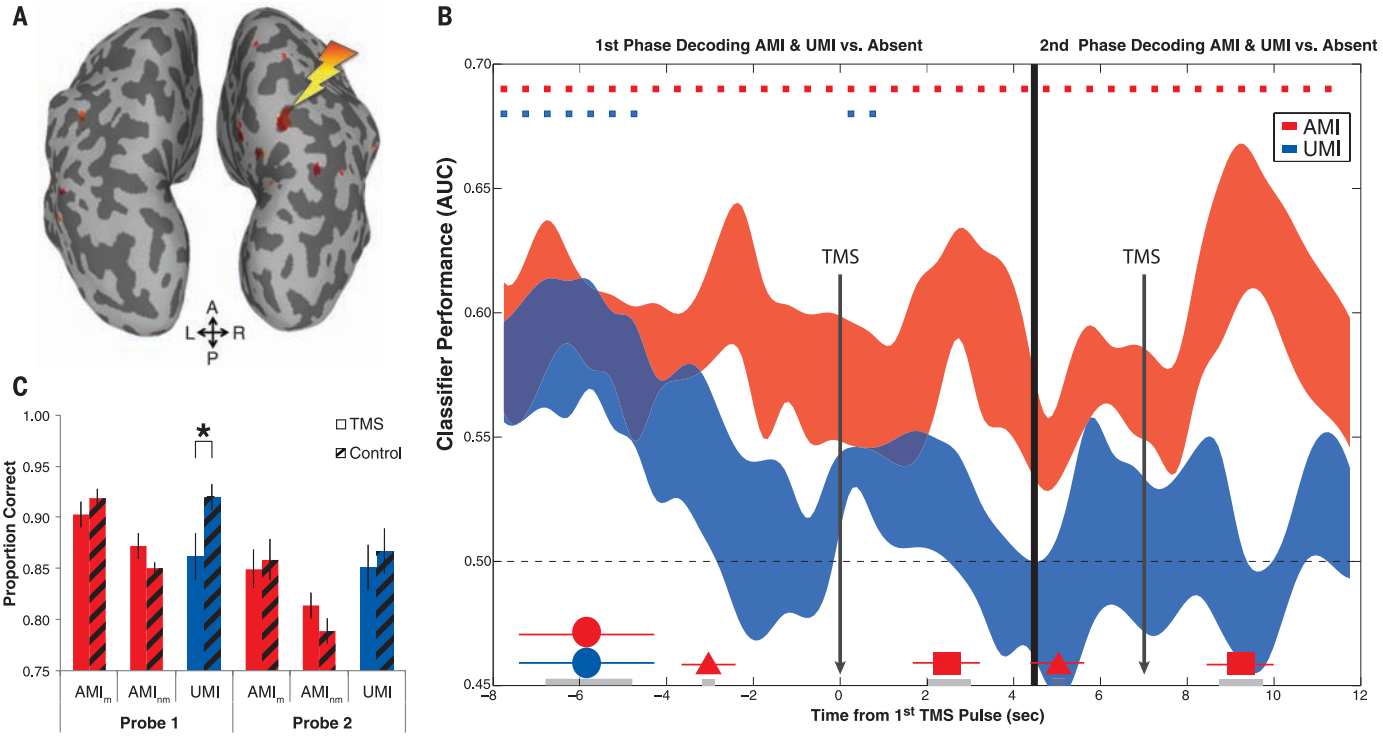


Fig. 4. Results from experiments 3 and 4. (**A**) The MVPA-defined TMS target for experiments 3 and 4 (right precuneus). A, anterior; L, left; R, right; P, posterior. (**B**) Classification time series from experiment 3 showing TMS reactivation of the UMI after the first cue, when the UMI was still relevant (left), but not after the second cue, when the UMI was no longer relevant on the trial (right) averaged over 1152 trials. Color-coded small squares at the top of each plot indicate $P < 0.05$; line width reflects SEM. (**C**) Experiment 4 recognition memory for AMI match probes (AMI_m), AMI nonmatch probes (AMI_{nm}), and UMI (nonmatch) probes. *P < = 0.01; error bars denote SEM.

provide neural evidence for at least two levels of WM that are distinct from the default state of LTM representations (*5*, *6*). They are inconsistent with models positing just one level of WM storage (*22*, *23*). They also suggest that instead of "activated LTM," a more apt label for the second level of WM would be "prioritized LTM." Information can be held in WM in latent "activity-silent" traces (*11*, *20*). What might be the physiological bases of such representations? Computational models of WM have proposed that short-term synaptic plasticity could be the basis for the transient formation of weight-based networks that can represent information over short time periods (*2*, *24*).

Our results provide empirical evidence for the existence of a short-term plasticity mechanism that is likely to be fundamental to a wide range of cognitive functions involving attentional selection (*25*) and may provide the building blocks for long-term potentiation mechanisms that support LTM (*26*). Therefore, our findings introduce a potential avenue for reactivating and strengthening representations that underlie many classes of high-level cognition.

REFERENCES AND NOTES

1. P. S. Goldman-Rakic, *Neuron* **14**, 477–485 (1995).
2. G. Mongillo, O. Barak, M. Tsodyks, *Science* **319**, 1543–1546 (2008).
3. Y. Sugase-Miyamoto, Z. Liu, M. C. Wiener, L. M. Optican, B. J. Richmond, *PLOS Comput. Biol.* **4**, e1000073 (2008).
4. M. J. Wolff, J. Ding, N. E. Myers, M. G. Stokes, *Front. Syst. Neurosci.* **9**, 123 (2015).
5. N. Cowan, in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake, P. Shah, Eds. (Cambridge Univ. Press, 1999), pp. 62–101.
6. K. Oberauer, *Psychol. Learn. Motiv.* **51**, 45–100 (2009).
7. J. J. LaRocque, J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, *J. Cogn. Neurosci.* **25**, 127–142 (2013).
8. J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, *J. Cogn. Neurosci.* **24**, 61–79 (2012).
9. J. J. LaRocque, A. C. Riggall, S. M. Emrich, B. R. Postle, *Cereb. Cortex* 10.1093/cercor/bhw283 (2016).
10. E. F. Ester, D. E. Anderson, J. T. Serences, E. Awh, *J. Cogn. Neurosci.* **25**, 754–761 (2013).
11. A. C. Riggall, B. R. Postle, *J. Neurosci.* **32**, 12990–12998 (2012).
12. B. Y. Hayden, J. L. Gallant, *Front Neurosci* **7**, 18 (2013).
13. N. E. Myers *et al.*, *eLife* **4**, e09000 (2015).
14. See supplementary materials on *Science* Online.
15. B. Kundu, J. Y. Chang, B. R. Postle, B. D. Van Veen, *Neuroimage* **114**, 320–327 (2015).
16. T. G. Lee, M. D'Esposito, *J. Neurosci.* **32**, 15458–15466 (2012).
17. T. B. Christophel, M. N. Hebart, J. D. Haynes, *J. Neurosci.* **32**, 12983–12989 (2012).
18. S. Monsell, *Cognit. Psychol.* **10**, 465–501 (1978).
19. D. M. Beck, N. Muggleton, V. Walsh, N. Lavie, *Cereb. Cortex* **16**, 712–717 (2006).
20. M. G. Stokes, *Trends Cogn. Sci.* **19**, 394–405 (2015).
21. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, *Science* **310**, 1963–1966 (2005).
22. B. McElree, *Psychol. Learn. Motiv.* **46**, 155–200 (2006).
23. A. Baddeley, *Annu. Rev. Psychol.* **63**, 1–29 (2012).
24. V. Itskov, D. Hansel, M. Tsodyks, *Front. Comput. Neurosci.* **5**, 40 (2011).
25. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
26. Y. Dudai, *Annu. Rev. Psychol.* **55**, 51–86 (2004).

ACKNOWLEDGMENTS

We thank J. Samaha, A. Sheldon, B. Kundu, J. Lewis-Peacock, and J. Johnson for assistance and helpful discussions. Supported by NIH grant MH095984 (B.R.P.). The data are stored at curate.nd.edu. Author contributions: N.S.R., B.R.P., J.J.L., and A.C.R. designed the research; N.S.R., O.G., M.J.S., and E.M.M. conducted the research; N.S.R., J.J.L., and A.C.R. analyzed the data; and N.S.R., B.R.P., and J.J.L. wrote the manuscript. The authors declare no conflict of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1136/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
Movie S1
References (*27*–*38*)

3 August 2016; accepted 31 October 2016
10.1126/science.aah7011

www.sciencemag.org/content/354/6316/1136/suppl/DC1

Supplementary Materials for

Reactivation of latent working memories with transcranial magnetic stimulation

Nathan S. Rose,* Joshua J. LaRocque, Adam C. Riggall, Olivia Gosseries, Michael J. Starrett, Emma E. Meyering, Bradley R. Postle*

*Corresponding author. Email: nrose1@nd.edu (N.S.R.); postle@wisc.edu (B.R.P.)

Published 2 December 2016, *Science* **354**, 1136 (2016)
DOI: 10.1126/science.aah7011

This PDF file includes:

Other supplementary material for this manuscript includes:

Supplemental Material

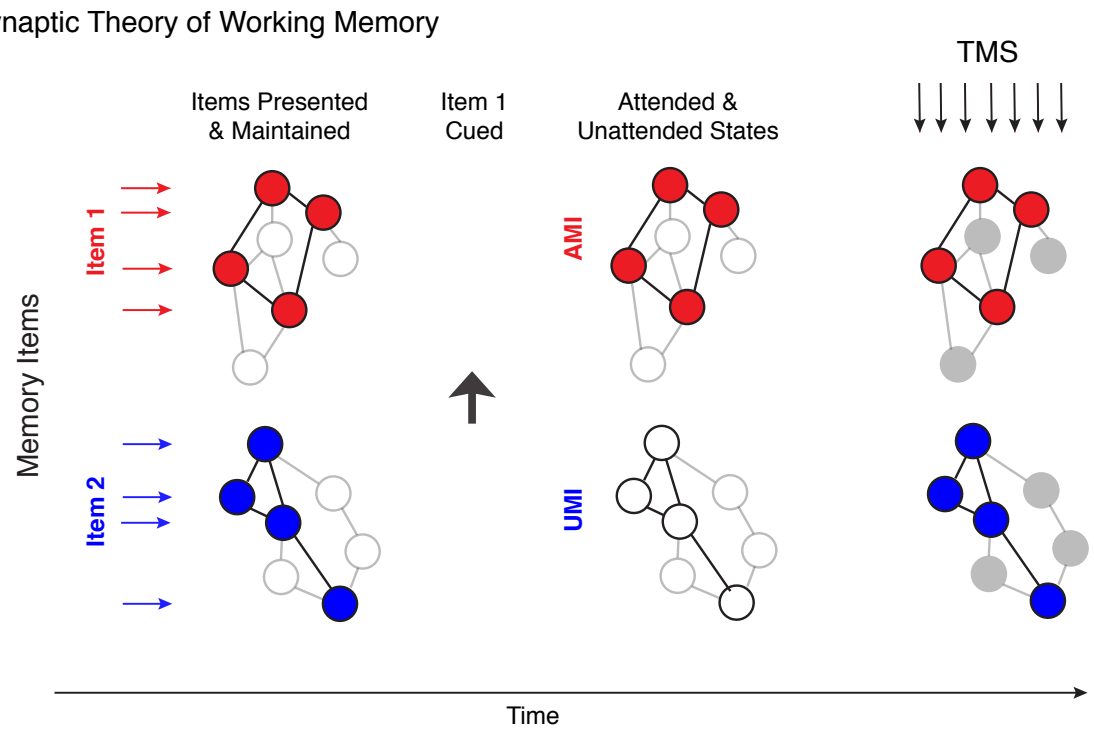


Fig. S1. Illustration of the Synaptic Theory of Working Memory. When memory items are presented and retained, their representations are activated. Attentional prioritization results in sustained activation of attended memory items (AMIs), but latent, synaptic-based representations of unattended memory items (UMIs) in "prioritized long-term memory"; an exogenous burst of energy via transcranial magnetic stimulation (TMS) can reactivate the latent, potentiated representations of UMIs.

Methods

Participants

Participants (ages 18-34 yrs) were recruited from the undergraduate and graduate student community of the University of Wisconsin–Madison and were paid for their participation. All had normal or corrected-to-normal vision, no reported history of neurological disease, no other contraindications for MRI or TMS, and all gave written informed consent in accordance with the local Institutional Review Board. Experiment 1 enrolled ten participants who performed both task phases in the MRI scanner, as well as 14 participants who performed

the tasks in a behavioral-only testing session (Fig. S1). Experiment 2 had nine participants for the Phase 1 MRI session and six of these participants participated in all three of the TMS-EEG sessions in Phase 2, with 164 trials in each session (N=18 sessions, 2,952 trials). One participant was excluded due to an incidental finding on the MRI; one participant dropped out of the study after the phase 1 MRI session; and one participant was excluded from analysis due to chance level behavioral performance on Phase 2.

Experiment 3 had six participants who participated in both the Phase 1 MRI session and a Phase 2 TMS-EEG session with 192 trials (N=1,152 trials). Experiment 4 had 20 participants perform the Phase 2 task with TMS (N=7,200 trials) and 23 control participants that performed the Phase 2 task in a behavioral-only testing session (N=8,280, 360 trials each). Data were excluded from analysis for one control participant due to a programming error, and one TMS and two control participants due to a misunderstanding of the task instructions.

Phase 1 – MRI session

Localizer task. In the first phase of each experiment, participants performed a one-item delayed-recognition task in the MRI scanner that required the short-term retention of a face, a word, or a direction of motion (see Fig. 1), and we applied MVPA to identify category-selective regions for subsequent targeting with TMS. On each trial of the task, a fixation cross appeared for 1.5 sec, followed by target stimulus (1 sec), delay period (7.5 sec), recognition probe stimulus (1 sec), response period and feedback (1 sec), followed by a variable inter-trial interval jittered between 6 and 12 sec in steps of 2 s with a rectangular distribution. Participants performed four runs of 18 trials each (72 total; 24 per category). Trials of the experimental

conditions in fMRI experiment (and all subsequent experiments) were randomly mixed within blocks of trials, with short breaks for rest between blocks.

Participants were to indicate if the probe did or did not match the target stimulus in terms of the features of the face, phonology of the word (rhyme), or direction of the moving dots by pressing button 1 or 2 of the MRI compatible response box, respectively. For the recognition probe, the face was either an exact match or the face with 50-70% shared genetic features morphed using FaceGen Modeller Software; the word was either a rhyme or non-rhyme; the patch (15° in diameter) of coherently moving dots (selected to be off the cardinal axes to minimize verbal recoding) were moving in either the same direction or a direction rotated clockwise or counterclockwise randomly between 5-45°.

Data acquisition and preprocessing. Whole-brain images were acquired with a 3-T MR scanner at the Lane Neuroimaging Laboratory at the University of Wisconsin-Madison. For all subjects, a high-resolution T1-weighted image was acquired with a fast spoiled gradient-recalled-echo sequence (8.132 ms TR, 3.18 ms TE, 12° flip angle, 156 axial slices, 256 X 256 in-plane, 1.0 mm isotropic). A gradient-echo, echo-planar sequence (2 s TR, 25 ms TE) was used to acquire data sensitive to the BOLD signal within a 64 X 64 matrix (39 sagittal slices, 3.5 mm isotropic).

The fMRI data were preprocessed offline using the Analysis of Functional NeuroImages (AFNI) software package (*27*). All volumes were spatially aligned to the first volume of each run using a rigid-body realignment and corrected for slice time acquisition. Linear, quadratic, and cubic trends were removed from each run to reduce the influence of scanner drift. For

univariate analyses, data were spatially smoothed with a 4 mm FWHM Gaussian kernel and warped to Talairach space (*28*). For MVPA, data were left in their native space (i.e., unsmoothed) and z-scored separately within each run for each voxel.

fMRI analyses. For univariate analysis, the sample, delay, and probe events within each trial were modeled separately as a boxcar of the corresponding duration. These regressors were convolved with a canonical hemodynamic response function and entered into a general linear model (GLM) for analysis using AFNI. MVPA was performed using the Princeton Multi-Voxel Pattern Analysis (www.pni.princeton.edu/mvpa) toolbox and custom routines in MATLAB. Preprocessed MRI data from individual trial time points were used to train a classifier to distinguish between the three stimulus categories. First, leave-one-trial-out cross-validation with L2 regularized logistic regression was used to classify stimulus category on each TR based on the signal from the top 2000 voxels identified by the omnibus F-test. The λ penalty term was set to 25 (based on prior analyses (*11*)) in order to reduce the contribution of less informative voxels to classification. Classifier accuracy was based on the proportion of trials in which the classifier chose the correct category label. Separate classifiers were trained and tested on each time point in the trial in the leave-one-trial-out cross-validation scheme. The result of this procedure is a time course of decoding accuracy for the entire trial, which can reveal the evolution of a memory representation over time including a delay period (Fig. 1B).

Then separate whole-brain searchlight analyses were conducted for each subject (Experiments 2 and 3) to classify both the presence vs. absence of each category and the presence of each category vs. each of the other categories on one post-sample delay-period TR

(time = 7 s) with a 3 mm sphere. The purpose of this analysis was to identify category-sensitive regions that could accurately discriminate one category from the others (e.g., face vs. not a face) at a high level (e.g., >75%) yet could not discriminate the other categories from one another (e.g., word vs. motion 40-60%) (Fig. 1C, Fig. S2). Note that we selected TMS targets based on the peak decoding accuracy of non-overlapping areas. For Experiment 2, the target was defined separately for each category for each individual.

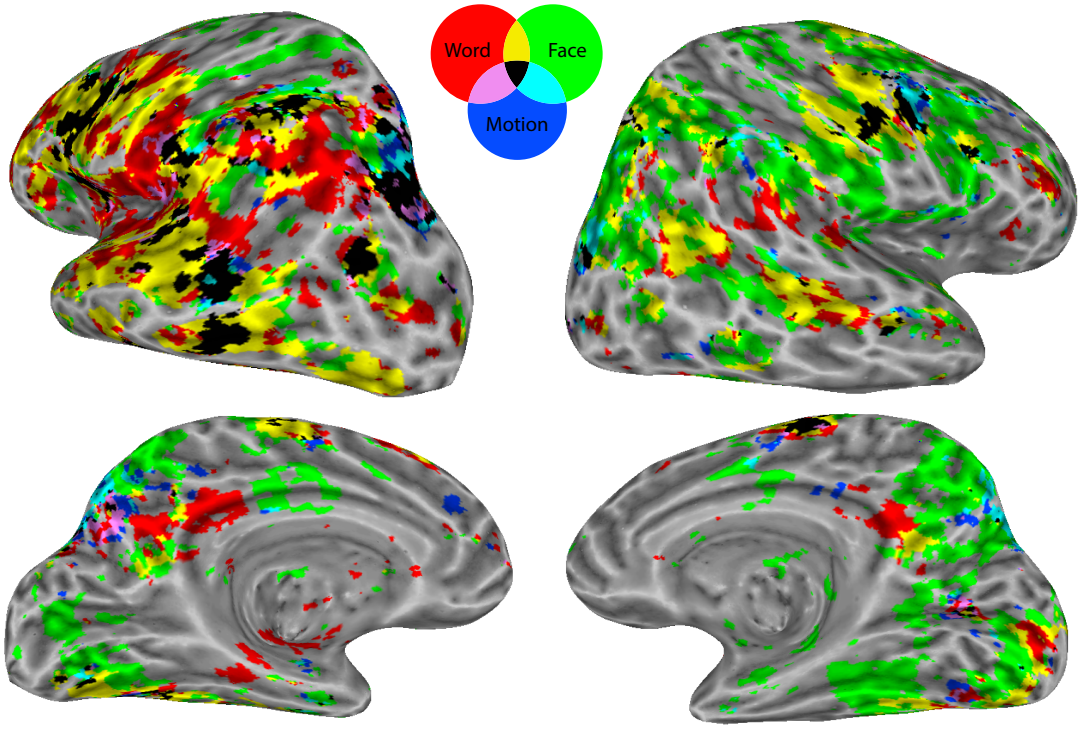


Fig. S2. Results of the wholebrain searchlight analysis that was used to select non-overlapping areas to target with TMS for a representative subject. The color-coding depicts areas that could decode one category from the others (e.g., word vs. not word in red) at >70% accuracy; overlapping areas are indicted according to the color-coded Venn diagram.

The MVPA-defined TMS target for Experiments 3 and 4 was the right precuneus, (MNI 11 -81 44) near the bank of the occipito-parietal sulcus in posterior parietal cortex—the

motion-selective area for all subjects in Experiment 2 (>70% wholebrain searchlight classification accuracy, t>9.0, p<.0001).

Phase 2 – fMRI (Experiment 1), Simultaneous TMS/EEG (Experiments 2 & 3), or Behavioral only (Experiment 4)

Experimental task

Participants performed a two-item delayed recognition task with two retrocues and two recognition probes per trial. A fixation cross appeared (.5 sec) and then two items from two of the three categories were presented as to-be-remembered items (Exp. 1 = 1 sec; Exps. 2, 3, & 4 = 2 sec), followed by an initial delay period (Exp. 1 = 7 sec; Exp. 2 = 5 sec; Exps. 3 & 4 = 2 sec), cue (.5 sec), a first post-cue delay period (4.5 sec), probe (1 sec), response & feedback (1.5 sec), a second cue (.5 sec), a second post-cue delay period (4.5 sec), probe (1 sec), response & feedback (1.5 sec), and an inter-trial interval randomly jittered between 2 and 4 sec. A single pulse of TMS was delivered during the first post-cue delay period at either 2.5 s after the cue offset (Exps. 2 & 4) or randomly jittered between 2-3 sec after cue offset in steps of 50 ms (Exp. 3) and then again during the second post-cue delay period (Exps 3 & 4, with the same timing parameters as the first post-cue delay period). All tasks were programmed in Matlab with Psychtoolbox functions to control stimulus presentation and record response accuracy and timing.

TMS targeting and stimulation

TMS was targeted with a Navigated Brain Stimulation (NBS) system that uses infrared-based frameless stereotaxy to coordinate the position of the coil and the participant's head according to the individual's high-resolution MRI overlaid with the TMS target for each category from the MVPA of the Phase 1 localizer task. TMS was delivered with a magnetic stimulator fit with a focal single-pulse, figure-of-eight stimulating coil. NBS allows measurement of the electrical field induced by TMS at the targeted cortex using a model of the participant's head, the coil position, and the estimated distance from the coil to the cortical target. TMS was delivered to the individually defined targets to achieve an estimated intensity at the stimulation target of 90–115 V/m (50–68% of stimulator output, depending on the thickness of the subject's scalp, cortex and depth of the target). The coil was oriented along the sagittal plane to induce an anterior-posterior direction of current, with individual adjustments to minimize artifact. Stimulator intensity, coil position, and coil orientation were held constant for each participant for the duration of each session. To mask the sound of the TMS coil's discharge, participants listened to white noise through earbuds during testing at a titrated volume such that the participant could not detect the click using a staircase thresholding procedure. The stimulation parameters were in accordance with published TMS guidelines.

EEG recording

EEG was recorded with a 60-channel cap and TMS-compatible amplifier, equipped with a sample-and-hold circuit that held amplifier output constant from 100 μs before stimulation to 2 ms after stimulation. Electrode impedance was kept below 5 kΩ. The reference electrode was placed superior to the supraorbital ridge. Eye movements were recorded with two additional

electrodes placed near the outer canthi of each eye. The EEG was recorded between 0.1 and 350 Hz at a sampling rate of 1450 Hz with 16-bit resolution.

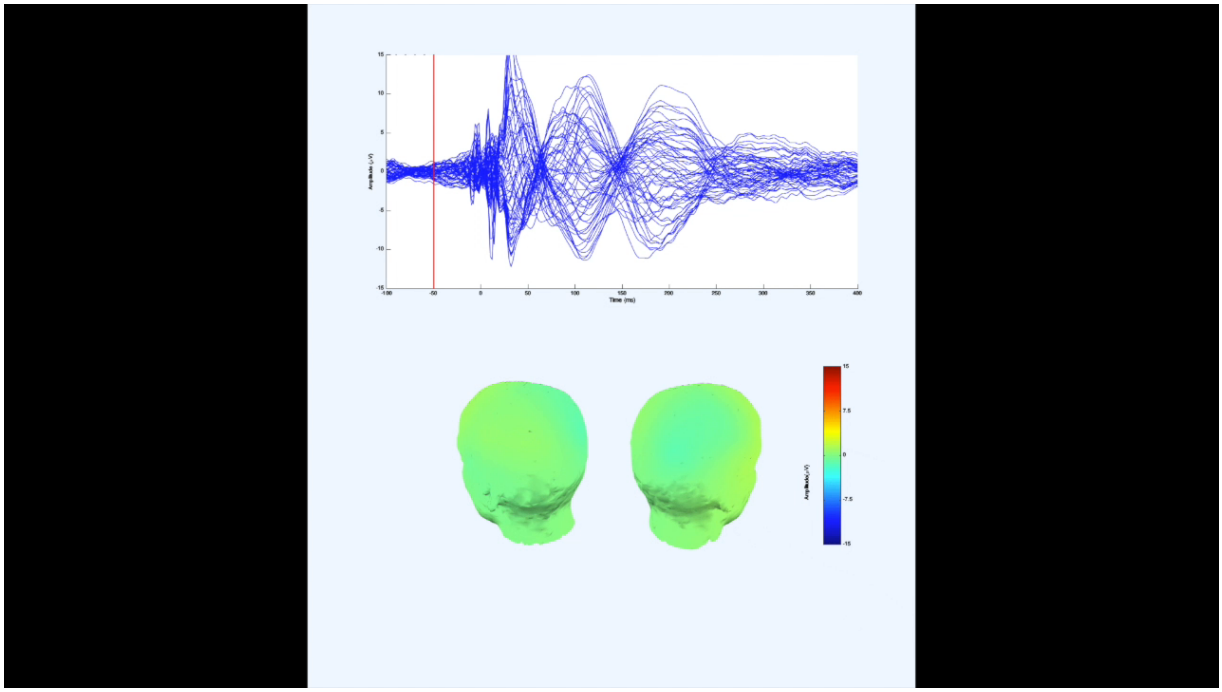
EEG preprocessing

Data were processed offline using EEGLAB (*29*) and Fieldtrip (*30*) toolboxes in MATLAB, downsampled to 500 Hz, and bandpass filtered between 1 and 100 Hz with a notch filter centered at 60 Hz. A spherical spline interpolation was applied to electrodes exhibiting excessive noise. In order to minimize the TMS spike artifact, data were interpolated from 0-30 ms around the TMS pulse for Experiment 2. A TMS pulse was applied on all trial types of the TMS-EEG experiments, which rules out an artifactually-based explanation of a TMS-induced reactivation effect. Nonetheless, in order to minimize the TMS spike artifact, data were interpolated from 0-30 ms around the TMS pulse for Experiment 2. Although this had no effect on the MVPA decoding of power in the binned frequency data, filtering produced some small distortions at the edges of the interpolated window (Movie S1). To better retain the shape and timing of the waveform around the TMS pulse for Experiment 3, a median filter was applied to -5 to 20 ms. Reassuringly, the results were not appreciably different (*31*) (for further details on preprocessing steps for TMS-EEG analysis, see http://nigelrogasch.github.io/TESA/). Independent components analysis was used to identify and remove components reflecting residual muscle activity, eye movements, blink-related activity, and residual TMS-related artifacts. Finally, the data were rereferenced to the average of all electrodes.

A Morlet wavelet transform was performed on the data using the Fieldtrip toolbox (*30*). 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 as in (*7*). This transform resulted in spectral power values at each of 34 frequencies and 60 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. As in (*7*), this temporal smoothing procedure was necessary to minimize the noise of the dynamic EEG signal. These data were the features used for all subsequent pattern classification analyses.

Movie S1. The movie shows the scalp topography of the TMS-evoked response for one of the stimulation sites for a representative subject.



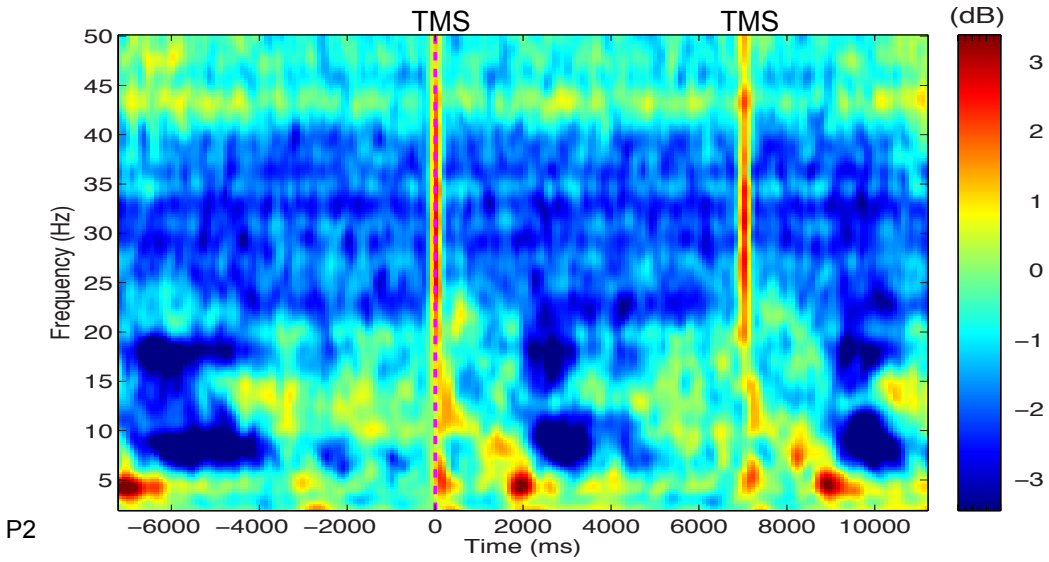


Fig. S3. The time frequency representation spectrogram for the whole trial timecourse from Experiment 3 time-locked to the first TMS pulse in electrode P2. The figure shows event related spectral perturbations associated with the presentation of the stimuli (~5 Hz) and administration of TMS across most frequencies. The spectral data depict the features that were used in the EEG decoding analyses.

MVPA decoding

MVPA was performed in MATLAB using the EEG Analysis Toolbox (*32*) (code.google.com/archive/p/eeganalysis-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 set to 1 based on prior research (*7*). Because participants performed the Phase 1 task in the MRI scanner, the EEG decoding analysis scheme considered each 0.5 sec time point of the Phase 2 task as a separate training exemplar, so that every trial yielded 40 exemplars. Each feature was z-scored across all trials and time points. The leave-one-out cross-validation scheme trained a classifier on data

from 191 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.

To measure the classifiers' ability to distinguish trials when a given category was present (and either cued or uncued) from trials in which the category was absent, we used the area under the receiver operating characteristic curve (AUC) as a metric (*33, 34*). Accuracy was determined by scoring the classifier's chosen category against the actual category, the conventional measure of accuracy. AUC was computed separately for each stimulus condition (i.e., for AMI and for UMI). For the AMI condition, for each stimulus category (i.e., faces, motion, words) we computed the proportion of "hits" (e.g., for faces, the proportion of trials when a face was the AMI and the classifier identified "face" as being present in the signal) vs. "false alarms" (again for faces, the number of trials when a face was not in the memory set yet the classifier identified "face" as being present in the signal). The analogous procedure was used for the UMI condition. An AUC of greater than 0.5 indicates sensitivity to the category of interest. Statistical significance of sensitivity was assessed with both Bayesian analyses and one-tailed, one-sample t-tests of AUC values averaged across the three categories versus the zero sensitivity value of 0.5, using data from the three sessions for each of the six participants. Classifier evidence was averaged separately across cue-switch and cue-repeat trials for the initially cued, initially uncued, and absent categories (Fig. 3 and Fig. 4, left panel). Note that in Fig. 4, the analysis involved decoding the UMI on the first cue period of the trial and that, after the first memory probe, the UMI would become the AMI in the second cue period on half of the trials (i.e., cue switch trials). For Experiment 3, in which TMS was also administered in the second post-cue delay period, the analysis was repeated to decode the category of the item

that was cued second or uncued second (or absent) during the latter half of the trial (Fig. 5, right panel).

Hypotheses regarding the reactivation effect on decoding accuracy were tested with Bayesian analyses to specifically compare evidence favoring the null versus the alternative hypothesis using the formula advocated by (*35*), with the unit-information scaled-prior conservatively set to the default value. In Experiment 2, for the timepoint immediately following the TMS pulse, classification of the UMI significantly differed from chance, $t(17) = 2.42$, $p = .01$, which resulted in a Bayes Factor (BF) of 3.64, which is considered moderate evidence in favor of the alternative (i.e., against the null) (*35*). Additionally, the reactivation effect for the TMS targeted category, $t(17)=2.53$, $p=.01$, $BF=4.02$, was larger and more prolonged than for the TMS non-targeted category, $t(17)=2.0$, $p=.03$, $BF=1.72$.

In Experiment 3, for the timepoint immediately following the first TMS pulse, classification of the UMI significantly differed from chance, $t(5) = 4.35$, $p < .005$, resulting in a Bayes Factor of 9.8, which is considered rather strong evidence against the null (*35*). In contrast, for the timepoint immediately following the second TMS pulse, when the UMI was no longer relevant, classification of the UMI did not differ from chance, $t(5) = 0.2274$ ($p=.42$), resulting in a Bayes Factor of 3.4 in favor of the null. For further confirmation, analyses were also conducted using the Bayes Factor calculations recommended by (*36*). For the timepoint immediately following the first TMS pulse, with the upper bound of the uniform population distribution set to the maximum classification of the UMI before the first cue (59% classification accuracy), the Bayes Factor was 3,035.84 against the null. In contrast, for the timepoint immediately following the second TMS pulse, with the upper bound of the uniform population

distribution set to the maximum reactivation effect of the UMI for all subjects following TMS after the first cue (64% classification accuracy), the Bayes Factor for the timepoint immediately following TMS after the second cue was 3.33 in favor of the null.

Supplemental Results

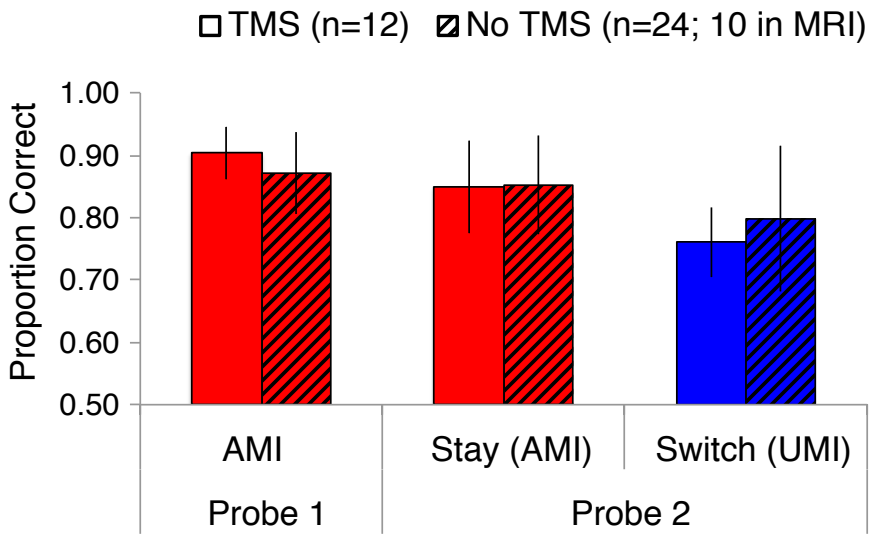


Fig. S4. Mean proportion of correct recognition memory responses on Probe 1 and Probe 2 on the phase 2 task for participants with or without single-pulse TMS during the delay period(s). Relative to Probe 1, recognition accuracy was slightly poorer on probe 2, especially for cue-switch trials (i.e., testing the item that was the UMI during the initial post-cue delay period, but there was no effect of TMS on recognition memory performance $F<1.0$.

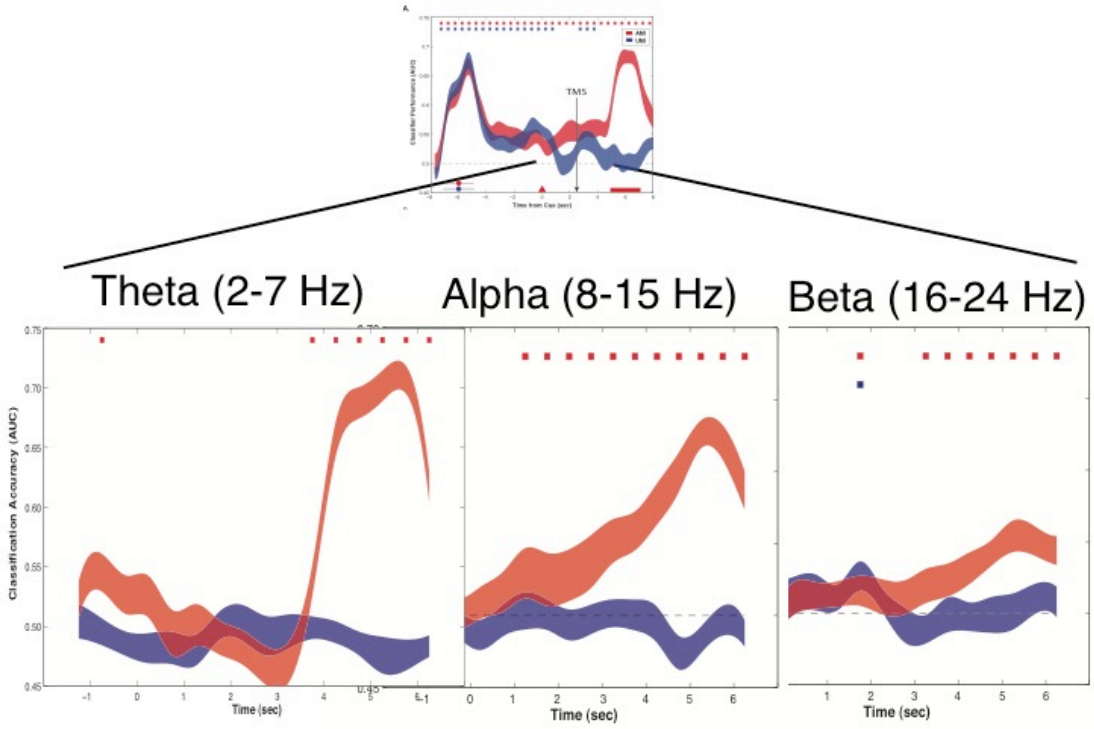


Fig. S5. Features that are important for decoding category specific representations of the AMI and UMI from the EEG: the AMI could be decoded from oscillations in the theta-band during the recognition memory probe period; the AMI could be decoded from alpha-band activity during the delay and memory probe phases; both the AMI and UMI could be decoded from the beta-band following administration of TMS. Thus, oscillations in the beta-band underlie the TMS-induced reactivation effect.

Supplemental Analyses for Experiment 4. To examine the effect of TMS on the ability to reject the UMI nonmatch probes, a mixed, repeated-measures ANOVA on recognition performance on nonmatch probes with stimulus type (UMI, AMInm), probe number (1, 2), and TMS condition was conducted on the data presented in Fig. 5C. There was a significant main effect of stimulus type, $F(1,39)=15.37$, $p<.001$, as well as a significant interaction between stimulus type

and TMS, F(1,39)=6.87, p=.01, because the UMI nonmatch probes were easier to reject than the AMI nonmatch probes for all contrasts except for Probe 1 with TMS, meaning that there was an increase in false alarms to the UMI probes following TMS on the first recognition memory test when the UMI was still potentially relevant on the trial (Probe 1: TMS UMI=.86, AMInm=.87, p=.35; Control UMI=.92, AMInm=.85, p<.001; Probe 2: TMS UMI=.85, AMInm=.81, p<.05; Control UMI=.87, AMInm=.79, p<.001). There was also a significant main effect of probe number because recognition memory performance was better overall for Probe 1 than Probe 2, F(1,39)=26.93, p<.001.

The critical contrast between Probe 1 UMIs for TMS vs. Controls, t(39)=2.33, p=.01, was also supported Bayseian analysis: Bayes Factor = 3.50 in favor of the alternative; for Probe 2 UMIs for TMS vs. Controls, t(39)=0.48, p=.32, the Bayes Factor was 2.24 in favor of the null.

Source Localized Analyses of the TMS-Evoked Response. In an attempt to clarify the anatomical substrates of working memory contents, we conducted source localization analyses of the TMS evoked response on the data from Experiment 2, in which unique TMS targets were utilized for each stimulus category. For each of the targeted stimulus categories, synthetic measures of the brain's response to TMS were computed separately for trials in which the targeted category was represented in each state (attended, unattended, absent). The significant current density (SCD) and significant current scatter (SCS) were computed using the methods detailed in (*37*), and the perturbational complexity index was computed using the methods detailed in (*38*).

First, epochs from -100 to 400 ms post-TMS were source localized using the 3-sphere BERG forward model, modeled using 3,000 meshes derived from each subject's anatomical

MRI. The global SCD measure is the sum of the absolute amplitude of all significant TMS-evoked currents (in microamperes per square millimeter) observed between 0 and 400 ms post-TMS in all cortical regions, identified using a nonparametric statistical procedure with n=70,000 bootstrapped samples relative to the prestimulus baseline (-100-0) (see (*37*) for further details). Global SCS is the sum of the geodesic distances (expressed in mm) between the stimulated brain region and significant current sources in distal brain regions. PCI uses the Lempel-Ziv complexity algorithm on the SCD and SCS measures to gauge the amount of non-redundant information in the matrix of significant currents scaled from 0 to 1 (see (*38*) for further details).

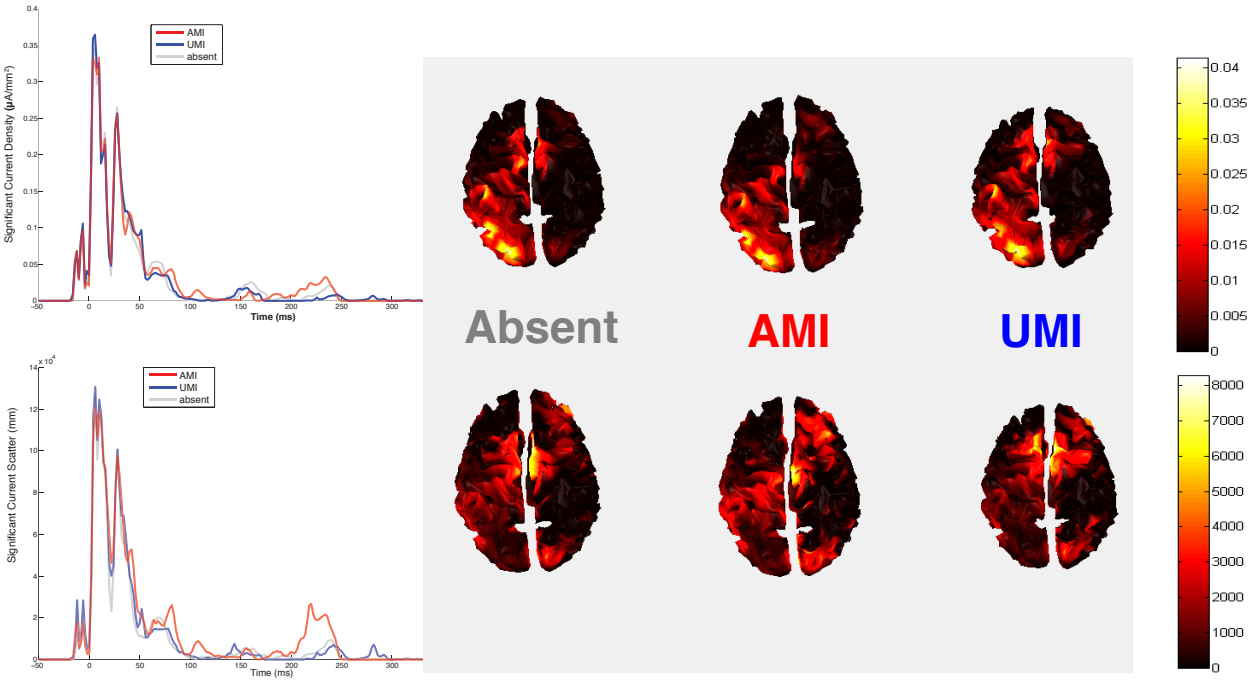


Fig. S6. Timeseries (left panel) and source localization (brain maps in the right panel) of the average significant current density (top row) and significant current scattering (bottom row) of the TMS-evoked response on trials in which the targeted category was the attended memory item (AMI), the unattended memory item (UMI), or absent.

Table S1. Synthetic measures of the TMS-evoked response.

Stimulus	Absent	Attended	Unattended
Significant Current Density (μA/mm^2)			
Motion	6.9 (2.0)	6.0 (1.3)	5.3 (1.2)
Face	5.4 (1.8)	7.9 (2.8)	8.2 (2.7)
Word	7.7 (3.2)	7.5 (2.7)	8.9 (3.0)
Mean	6.7 (1.9)	7.1 (1.6)	7.4 (2.1)
Significant Current Scatter (mm)			
Motion	4.57E+06 (1.22E+06)	3.50E+06 (6.47E+05)	3.22E+06 (7.84E+05)
Face	3.19E+06 (1.01E+06)	3.73E+06 (8.40E+05)	4.03E+06 (9.71E+05)
Word	4.31E+06 (2.01E+06)	4.32E+06 (1.85E+06)	5.17E+06 (2.29E+06)
Mean	4.02E+06 (1.41E+06)	3.85E+06 (1.11E+06)	4.14E+06 (1.35E+06)
*Perturbational Complexity Index (PCI)**			
Motion	0.442 (0.036)	0.418 (0.023)	0.405 (0.032)
Face	0.402 (0.041)	0.376 (0.033)	0.378 (0.056)
Word	0.399 (0.050)	0.393 (0.052)	0.425 (0.050)
Mean	0.415 (0.037)	0.396 (0.032)	0.403 (0.040)

These synthetic measures of the TMS-evoked response failed to distinguish between any of the WM states of representation. All main effects and interactions on the global SCD, SCS, and PCI measures failed to reach significance (*p*s>.05). Although these synthetic measures have been successfully used to discriminate between varying states of consciousness, e.g., coma or non-REM sleep vs. wakefulness (*38*), or fixation vs. the perceptually identical delay period of a WM task (*37*), it seems unlikely that these global measures of the brain's response to TMS during WM retention could distinguish between different states of accessibility for an individual item that is represented either in an attended state or an unattended state from items that were attended on previous trials. Thus, the MVPA decoding methods were more sensitive to detecting multivariate patterns of brain activity associated with stimulus-specific representations than the synthetic measures of voltage on the scalp in source-localized space.

To address more structural hypotheses about the precise anatomical substrates of representations in working memory, future work should more directly probe for the presence of stimulus specific representations in sensory or associative cortices. More invasive methods than those used in the current work may be necessary to sufficiently address such hypotheses.

References

1. P. S. Goldman-Rakic, Cellular basis of working memory. *Neuron* **14**, 477–485 (1995). Medline doi:10.1016/0896-6273(95)90304-6

2. G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008). Medline doi:10.1126/science.1150769

3. Y. Sugase-Miyamoto, Z. Liu, M. C. Wiener, L. M. Optican, B. J. Richmond, Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLOS Comput. Biol.* **4**, e1000073 (2008). Medline doi:10.1371/journal.pcbi.1000073

4. M. J. Wolff, J. Ding, N. E. Myers, M. G. Stokes, Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* **9**, 123 (2015). Medline doi:10.3389/fnsys.2015.00123

5. N. Cowan, in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake, P. Shah, Eds. (Cambridge Univ. Press, 1999), pp. 62–101. doi:10.1017/CBO9781139174909.006

6. K. Oberauer, Design for a working memory. *Psychol. Learn. Motiv.* **51**, 45–100 (2009). doi:10.1016/S0079-7421(09)51002-X

7. J. J. LaRocque, J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, Decoding attended information in short-term memory: An EEG study. *J. Cogn. Neurosci.* **25**, 127–142 (2013). Medline doi:10.1162/jocn_a_00305

8. J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* **24**, 61–79 (2012). Medline doi:10.1162/jocn_a_00140

9. J. J. LaRocque, A. C. Riggall, S. M. Emrich, B. R. Postle, Within-category decoding of information in different attentional states in short-term memory. *Cereb. Cortex* 10.1093/cercor/bhw283 (2016). doi:10.1093/cercor/bhw283

10. E. F. Ester, D. E. Anderson, J. T. Serences, E. Awh, A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* **25**, 754–761 (2013). Medline doi:10.1162/jocn_a_00357

11. A. C. Riggall, B. R. Postle, The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* **32**, 12990–12998 (2012). Medline doi:10.1523/JNEUROSCI.1892-12.2012

12. B. Y. Hayden, J. L. Gallant, Working memory and decision processes in visual area v4. *Front Neurosci* **7**, 18 (2013). Medline doi:10.3389/fnins.2013.00018

13. N. E. Myers, G. Rohenkohl, V. Wyart, M. W. Woolrich, A. C. Nobre, M. G. Stokes, Testing sensory evidence against mnemonic templates. *eLife* **4**, e09000 (2015). Medline doi:10.7554/eLife.09000

14. See supplementary materials on *Science* Online.

15. B. Kundu, J. Y. Chang, B. R. Postle, B. D. Van Veen, Context-specific differences in fronto-parieto-occipital effective connectivity during short-term memory maintenance. *Neuroimage* **114**, 320–327 (2015). Medline doi:10.1016/j.neuroimage.2015.04.001

16. T. G. Lee, M. D'Esposito, The dynamic nature of top-down signals originating from prefrontal cortex: A combined fMRI-TMS study. *J. Neurosci.* **32**, 15458–15466 (2012). Medline doi:10.1523/JNEUROSCI.0627-12.2012

17. T. B. Christophel, M. N. Hebart, J. D. Haynes, Decoding the contents of visual short-term memory from human visual and parietal cortex. *J. Neurosci.* **32**, 12983–12989 (2012). Medline doi:10.1523/JNEUROSCI.0184-12.2012

18. S. Monsell, Recency, immediate recognition memory, and reaction time. *Cognit. Psychol.* **10**, 465–501 (1978). doi:10.1016/0010-0285(78)90008-7

19. D. M. Beck, N. Muggleton, V. Walsh, N. Lavie, Right parietal cortex plays a critical role in change blindness. *Cereb. Cortex* **16**, 712–717 (2006). Medline doi:10.1093/cercor/bhj017

20. M. G. Stokes, 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015). Medline doi:10.1016/j.tics.2015.05.004

21. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963–1966 (2005). Medline doi:10.1126/science.1117645

22. B. McElree, Accessing recent events. *Psychol. Learn. Motiv.* **46**, 155–200 (2006). doi:10.1016/S0079-7421(06)46005-9

23. A. Baddeley, Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* **63**, 1–29 (2012). Medline doi:10.1146/annurev-psych-120710-100422

24. V. Itskov, D. Hansel, M. Tsodyks, Short-term facilitation may stabilize parametric working memory trace. *Front. Comput. Neurosci.* **5**, 40 (2011). Medline doi:10.3389/fncom.2011.00040

25. R. Desimone, J. Duncan, Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995). Medline doi:10.1146/annurev.ne.18.030195.001205

26. Y. Dudai, The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* **55**, 51–86 (2004). Medline doi:10.1146/annurev.psych.55.090902.142050

27. R. W. Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996). Medline doi:10.1006/cbmr.1996.0014

28. J. Talairach, P. Tournoux, *Co-planar Stereotaxic Atlas of the Human Brain. 3-D Proportional System: An Approach to Cerebral Imaging* (Thieme, 1988).

29. A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004). Medline doi:10.1016/j.jneumeth.2003.10.009

30. R. Oostenveld, P. Fries, E. Maris, J. M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011). Medline doi:10.1155/2011/156869

31. N. C. Rogasch, C. Sullivan, R. H. Thompson, N. S. Rose, N. W. Bailey, P. B. Fitzgerald, F. Farzan, J. C. Hernandez-Pavon, Analysing concurrent transcranial magnetic stimulation and electroencephalographic data: A review and introduction to the open-source TESA software. *Neuroimage* 10.1016/j.neuroimage.2016.10.031 (2016). doi:10.1016/j.neuroimage.2016.10.031

32. N. W. Morton, M. J. Kahana, E. A. Rosenberg, G. H. Baltuch, B. Litt, A. D. Sharan, M. R. Sperling, S. M. Polyn, Category-specific neural oscillations predict recall organization during memory search. *Cereb. Cortex* **23**, 2407–2422 (2013). Medline doi:10.1093/cercor/bhs229

33. E. L. Newman, K. A. Norman, Moderate excitation leads to weakening of perceptual representations. *Cereb. Cortex* **20**, 2760–2770 (2010). Medline doi:10.1093/cercor/bhq021

34. T. Fawcett, An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006). doi:10.1016/j.patrec.2005.10.010

35. J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey, G. Iverson, Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009). Medline doi:10.3758/PBR.16.2.225

36. Z. Dienes, Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5**, 781 (2014). Medline doi:10.3389/fpsyg.2014.00781

37. J. S. Johnson, B. Kundu, A. G. Casali, B. R. Postle, Task-dependent changes in cortical excitability and effective connectivity: A combined TMS-EEG study. *J. Neurophysiol.* **107**, 2383–2392 (2012). Medline doi:10.1152/jn.00707.2011

38. A. G. Casali, O. Gosseries, M. Rosanova, M. Boly, S. Sarasso, K. R. Casali, S. Casarotto, M. A. Bruno, S. Laureys, G. Tononi, M. Massimini, A theoretically based index of

consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* **5**, 198ra105 (2013). Medline doi:10.1126/scitranslmed.3006294