Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load

Thomas C. Sprague,1,* Edward F. Ester,2 and John T. Serences1,2
1Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093-0109, USA
2Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, USA

Summary

Working memory (WM) enables the maintenance and manipulation of information relevant to behavioral goals. Variability in WM ability is strongly correlated with IQ [1], and WM function is impaired in many neurological and psychiatric disorders [2, 3], suggesting that this system is a core component of higher cognition. WM storage is thought to be mediated by patterns of activity in neural populations selective for specific properties (e.g., color, orientation, location, and motion direction) of memoranda [4–13]. Accordingly, many models propose that differences in the amplitude of these population responses should be related to differences in memory performance [14, 15]. Here, we used functional magnetic resonance imaging and an image reconstruction technique based on a spatial encoding model [16] to visualize and quantify population-level memory representations supported by multivoxel patterns of activation within regions of occipital, parietal, and frontal cortex while participants precisely remembered the location(s) of zero, one, or two small stimuli. We successfully reconstructed images containing representations of the remembered—but not forgotten—locations within regions of occipital, parietal, and frontal cortex using delay-period activation patterns. Critically, the amplitude of representations of remembered locations and behavioral performance both decreased with increasing memory load. These results suggest that differences in visual WM performance between memory load conditions are mediated by changes in the fidelity of large-scale population response profiles distributed across multiple areas of human cortex.

Results

To assess the functional role that population codes in different visually responsive occipital, parietal, and frontal regions of interest (ROIs) play in spatial working memory (WM), we presented participants (n = 4, four scanning sessions each) with two target stimuli (Figure 1A) followed by a postcue instructing them to remember the location(s) of zero (R0), one (R1), or two (R2) stimuli. In behavioral testing sessions performed outside of the scanner, participants used a mouse click to indicate the exact position of the remembered target. During scanning, participants performed a two-alternative forced-choice (2AFC) discrimination task in which they compared the position of a probe stimulus to that of the corresponding remembered target stimulus (Figure 1A). We chose to test precise memory for spatial positions using either a recall task (outside the scanner) or a “same/different” task (during scanning) so that participants were required to encode exact spatial positions rather than use a verbal code or only encode a single dimension (e.g., “8 o’clock,” “far to the left”).

Behavioral performance on the analog recall task performed outside the scanner revealed lower mnemonic precision when two target locations were remembered compared to when a single target location was remembered (Figure 1C; p < 0.001, resampling test). During scanning, response accuracy did not significantly differ across set size conditions, although three out of four participants performed slightly worse with increasing set size (Figure 1D, p = 0.174, resampling test; see the Experimental Procedures). However, response times (RTs) were significantly longer when two stimuli were remembered compared to when a single stimulus was remembered (Figure 1E; p < 0.001, resampling test). Increased RTs during scanning suggest that memory representations in the R2 condition were degraded and therefore less accessible during behavioral report, consistent with previous observations of increased RTs after manipulations that impair spatial WM (e.g., [17]). Together, the behavioral data recorded inside and outside of the scanner are consistent with a degraded representation of each remembered location in the R2 condition compared to the R1 condition.

To characterize neural responses associated with WM maintenance, we first compared averaged blood-oxygenation-level-dependent (BOLD) functional magnetic resonance imaging (fMRI) responses in a set of functionally defined occipital (V1–hV4 and V3A), parietal (IPS0–IPS3), and frontal (sPCS; thought to be the human homolog of macaque frontal eye fields [18, 19]) ROIs as a function of memory load. We replicated previous reports that BOLD responses in frontal and parietal ROIs were larger on R2 trials compared to R1 trials [6, 20, 21] (Figure S1 available online). Interestingly, in early visual areas (V2–V3a and hV4) we observed a larger mean BOLD amplitude on R0 trials compared to R1 or R2 trials (Figure S1B, p < 0.001, resampling test). We also observed similar results using a complementary exploratory analysis in which we searched for any voxels with increased activation for larger memory loads (Figure S1C).

Next, we used a multivariate image reconstruction technique based on a spatial encoding model [16] to reconstruct remembered locations in spatial WM based on the pattern of activation across all voxels within each ROI (Figure 2). In contrast to analyses that focus solely on mean signal intensity (Figure S1), neural firing rates, or multivariate classification accuracy, this analysis uses an independently estimated model of the spatial sensitivity profile across all voxels in each ROI to transform BOLD activation patterns into an image of the remembered stimulus position(s) carried by those patterns (Figure 2; Experimental Procedures). Importantly, this analysis provides additional information compared to some other methods such as univariate population receptive field (pRF) [22] estimation or multivariate linear classification [9]: by yielding a reconstructed image of the remembered stimulus location(s), covert information held in WM can be directly visualized, quantified, and...
related to behavior [16]. These reconstructions can be thought of as an image of the spatial WM contents in visual field coordinates (rather than coordinates relative to the cortical surface), and we interpret the focal bright spots found at target positions as target representations.

Spatial WM reconstructions computed based on patterns of delay-period activation from occipital (V1–hV4/vV3A), parietal (IPS0–IPS3), and frontal (sPCS) cortex revealed highly robust representations of remembered target positions on R1 trials, but not on R0 trials (Figure 3; see Figure S2A for data from individual participants), suggesting that these images reflect memory-related activation changes rather than lingering sensory signals. Furthermore, reconstructed images contain representations of both remembered target locations on R2 trials that were robust in many occipital and posterior parietal ROIs (Figure 3C; V1–V3A, hV4, IPS0, and IPS1) but became less separable in anterior parietal and frontal ROIs (IPS2–IPS3 and sPCS). The relative decline in separability of R2 target representations in these anterior parietal and frontal ROIs may reflect the rather small screen size that we used relative to the large size of spatial RFs typical of these ROIs [23, 24]. Finally, we examined the temporal structure of WM reconstructions from all ROIs over the course of the entire trial. We could readily reconstruct images of both remembered locations during target presentation when the positions were encoded into WM, but we could only reconstruct images of locations held in WM during the delay interval (Movie S1).

Next, we sought to quantify how spatial WM reconstructions differ across ROIs and under different memory loads. To do so, we rotated and shifted the reconstruction on each trial to a common reference location such that the target positions were in alignment and averaged all coregistered reconstructions together (Figure 2E; see Movie S2 for coregistered reconstructions through time). Then, because the target position across all trials was now aligned, we quantified attributes of the averaged target representation by fitting a 2D surface (Figures 4A and 4B) characterized by several independent parameters (see Figures S4A–S4D for a demonstration that these parameters reflect dissociable properties of target representations). The size parameter reflects the spread (full-width half-maximum, FWHM) of the delay-period target representation: an increased fit size would reflect a less spatially precise representation of the remembered target location (note that here and elsewhere, we use “spatial” with reference to visual field space, not cortical space). The amplitude parameter reflects the height of the target representation over baseline: increased fit amplitude would correspond to a more prominent representation of the target over baseline activation not related to the target location. The baseline parameter reflects the non-spatially-selective response amplitude (i.e., a constant offset across the entire reconstructed visual field): a change in baseline reflects a change in mean signal amplitude across an entire ROI that does not carry spatial information and thus does not directly change the spatial information content of the reconstruction.

Increasing memory load did not change the size of the best-fit surfaces to the target representations within WM reconstructions that were based on activation patterns in occipital and posterior parietal ROIs (Figure 4D; V1–IPS0; all statistics were computed via nonparametric resampling methods and Bonferroni corrected for multiple comparisons; Table S1; see the Experimental Procedures). However, fit surface size did increase with memory load in anterior parietal (IPS2–IPS3) and frontal (sPCS) ROIs. Note that in these ROIs, we did not observe strongly disjoint target representations during R2 trials (Figure 3C), so these size increases may partially reflect an inability to separately quantify the representation of each location. It is likely that a larger display and more stimulus separation would enable a more accurate reconstruction and quantification of each remembered target representation in these anterior parietal and frontal areas (like in the early visual and posterior IPS ROIs). We evaluated the possibility that observed size increases may be partially an artifact of coregistering reconstructions and averaging over target positions on R2 trials, even if the “true” target representations are constant in size, by simulating reconstructions under the null assumption that target representations were equal in size.
WM Reconstructions Degrade with Memory Load

Please cite this article in press as: Sprague et al., Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.07.066

Figure 2. Inverted Spatial Encoding Model for Reconstructing the Contents of Spatial WM

(A) Each participant was scanned for three to four independent spatial mapping runs (for encoding model estimation per session (see the Supplemental Experimental Procedures). Participants performed a challenging spatial WM task in which they determined whether a probe stimulus (500 ms) was in the exact same position or a slightly different position from a remembered target position (500 ms; 2AFC; see [16]). During the brief delay period (3,000 ms), a flickering checkerboard stimulus was presented near the remembered target position. This stimulus was irrelevant to the task performed by the participant but was used to drive large sensory responses to estimate a voxel-level encoding model used for computing reconstructions in the main task (see C–E). We adjusted difficulty on a run-by-run basis to maintain vigilance and equate performance across participants and sessions (73.738% ± 1.819% accuracy, mean ± SEM).

(B) We presented the mapping stimulus at each of 36 positions arrayed across a 6 × 6 square grid (one trial per position per run). To estimate spatial sensitivity profiles for each voxel, we predicted the response of each of 36 hypothetical “information channels” (spatial filters) to each stimulus used in the training runs [16]. Then, we took the measured response of each voxel and the predicted hypothetical channel responses to each stimulus position and used ordinary least-squares linear regression to estimate the contribution of each information channel to the signal observed in each voxel. This step is performed on each voxel independently (see the Supplemental Experimental Procedures, Equation 3).

(D) For each collection of voxels for which we computed reconstructions (ROIs, Figures 3 and 4), all voxels from all ROIs, Figure 4) we computed a mapping from voxel space into channel space (Supplemental Experimental Procedures, Equation 4). In contrast to “population receptive field” analyses [22], this step is multivariate and must be performed using all voxels that contribute to the image reconstruction. Using the computed linear mapping, the measured activation pattern across all voxels is transformed into “information space”—the amount each channel must have been active in order to produce the measured voxel activation pattern. A “raw” reconstruction can be computed for any single observation (e.g., one fMRI volume from area V1) by computing a sum of the spatial filters that define the information channels weighted by the estimated channel responses (right panel).

(E) When computing average reconstructions across all trials (Figures 4C and S2B), we coregistered different target positions on each trial to a common location by first rotating the spatial filters around the fixation point such that the target lies along the Cartesian x axis, then shifting the filter centers horizontally such that the target is positioned 3.25° from fixation along the x axis (white dot in reconstructions shown in Figures 4C and S2B). For R0 and R2 trials, this is done for each remembered target, and the coregistered reconstructions aligned to each target are averaged. Importantly, this coregistration procedure enables us to average the representations of spatial WM targets that appeared at different positions in the display on different trials.
Figure 3. Reconstructed Contents of Spatial WM Measured using Delay-Period Patterns of Activation

Image reconstructions for all target position arrangements during remember zero (A), remember one (B), and remember two (C) conditions from each ROI. Each reconstruction is computed using spatial filters that have been rotated around the fixation point and flipped over the horizontal meridian such that there are four possible target arrangements (top panel; dashed yellow circles indicate remembered target[s]). Targets appeared uniformly within each of these four windows. Early visual (V1–hV4) and parietal (IPS0–IPS1) ROIs carry precise target representations over the delay interval of a single remembered position (remember one; B) or both remembered positions (remember two; C). Reconstructions from anterior parietal (IPS2–IPS3) and frontal (sPCS) ROIs carry moderately precise target representations when a single position is maintained in WM, but they are not as disjoint when both positions are simultaneously held in WM (IPS, intraparietal sulcus; sPCS, superior precentral sulcus, human homolog to macaque frontal eye fields [18,19]). Additionally, despite a significant reduction in average BOLD response during the delay period in occipital ROIs (Figure S1), reconstructions contain robust representations of remembered stimuli. The color map is identical across all panels. See also Figure S1, Figure S2A for spatial reconstructions from each individual participant, and Movie S1 for temporal unfolding of reconstructions across the duration of the trial.

Discussion

Here, we employed an image reconstruction approach implemented using a multivariate inverted encoding model [8,16,25–28] to reconstruct the contents of spatial WM based on activation patterns in occipital, parietal, and frontal regions of human cortex. Prior studies have used measures like
classification accuracy to correlate behavioral performance with the discriminability of neural activation patterns [6, 13]. Although these analyses have many advantages due to a relative lack of model assumptions, changes in decoding accuracy may result from many different types of neural response pattern modulation [25, 29]. In contrast, by assuming a set of spatial basis functions, our method allows us to assess whether each region encoded information about the location of a remembered stimulus (e.g., [5, 30]), as well as to visualize and quantify the characteristics of these covert representations of target locations and relate different aspects of these quantified representations to behavioral performance (e.g., [8, 16, 25, 27]). In addition, these findings reinforce the importance of measuring the effect of cognitive manipulations on population-level estimates of mnemonic representations rather than on particular properties of the underlying neural generators, as these population-level representations can be more closely associated with cognition and behavior than activity changes in single neurons or voxels [3, 16, 25–29, 31–33]. These image reconstruction and quantification analyses revealed lower amplitude and, in some anterior parietal and frontal ROIs, broader target representations with increasing memory load (Figure 4). From an information-theoretic perspective, response variability (i.e., intertrial variability in the reconstructed images) has two components: signal entropy, which is variability associated with experimental manipulations (remembered location), and noise entropy, which is variability not associated with experimental manipulations. The decrease in target representation amplitude under increased memory load should lead to less variability that is related to the remembered location(s) and thus to a decrease in the signal entropy and information about the remembered location. An increase in target representation size should also decrease signal entropy, as increased size leads to
more overlap between target representations for different locations, which would decrease the ability of the population code to discriminate between locations. In contrast, baseline shifts should not strongly influence information content as an additive shift in the entire reconstruction does not change signal entropy [14, 16, 34] (Figure S3). Thus, the observation of higher amplitude target representations corresponds to higher information content of population codes about a spatial position [14–16, 32–34] (Figure S3) and may be a consequence of changes in delay-period neural gain associated with neurons tuned to remembered locations [14]. In addition, modest increases in target representation size in anterior IPS and sPCS may reflect poorer mnemonic fidelity within particular ROIs, echoing previous results that the dispersion (analogous to size here) of reconstructed profiles of remembered features (e.g., orientation) correlates with behavioral performance [5, 25, 27]. However, future work using larger spatial stimulus arrays may help to more accurately disentangle and characterize multiple WM representations in anterior IPS and sPCS.

We were able to reconstruct the covert contents of spatial WM not only in occipital [4, 6–10, 13] and posterior parietal regions [10, 13], but also in anterior parietal and frontal cortex [5, 11]. These widespread modulations raise the possibility that distributed WM representations can be optimized to differentially contribute to complementary sensory (e.g., target localization) and motor (e.g., eye movements, reaches) behaviors. Consistent with this idea, a recent demonstration that induced alpha oscillations (which are often thought to reflect synchronized activity of large-scale cortical networks [35]) measured with scalp EEG can be used to reconstruct remembered orientations also suggests that long-range, interacting representations across much of human cortex support the maintenance of information in WM [27]. The successive representations of spatial position reported here may thus allow for a common coordinate system with which low-level stimulus features (such as spatial position and color) that are represented in occipital cortex are bound with spatial motor plans (such as eye movements and arm reaches [36]) that are more closely associated with representations in parietal and frontal cortex.

Quantifying Target Representations in WM Reconstructions

We fit a surface to each reconstruction that was allowed to vary in its size, amplitude, and base line (Figures 4A and 4B). Its center was constrained to be the position, in visual field coordinates, with the highest local reconstruction amplitude (local average within a 0.5° radius).

Statistics

For group-level analyses (Figures 1C–1E, 4D, and S1B), we combined data from all participants within a given ROI and memory load condition and resampled all trials with replacement and computed a mean measurement value for that resampling iteration (Figure 1C, behavioral recall error; Figure 1D, behavioral accuracy; Figure 1E, response time; Figure 4D, target representation fit parameters; Figure S1B, mean BOLD signal). We repeated this procedure 1,000 times to produce a resampled distribution of each measured value for each memory load condition. We computed p values for each ROI and each parameter as the two-tailed probability of observing an effect in the opposite direction of the mean effect observed. Comparisons are Bonferroni corrected across ROIs for each parameter (Figure 4D, ten comparisons) or across all comparisons performed (Figure S1B, 30 pairwise comparisons). All error bars are 95% confidence intervals derived from these resampled distributions unless indicated otherwise (Figure S1A).

For exploratory individual-participant analyses (Figure S2C), we performed an identical procedure but resampled only across each participant’s data when computing confidence intervals.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, four figures, one table, and two movies and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.07.066.

Author Contributions

T.C.S., E.F.E., and J.T.S. developed the experiment protocol and wrote the manuscript. T.C.S. and E.F.E. collected data. T.C.S. analyzed data. J.T.S. supervised the project.

Acknowledgments

We thank Miranda Scolari and Mary Smith for assistance in developing parietal cortex mapping protocols, Anna Byers for assistance with data collection, Sirawaj Itthipuripat, Vy Vo, and Alexander Heitman for discussion, and Sirawaj Itthipuripat, Vy Vo, and Stephanie Nelli for comments on the manuscript. This work was supported by a NSF Graduate Research Fellowship to T.C.S., NIH T32-MH020002-12 to E.F.E., and NIH R01-MH-092345 to J.T.S. Data from each cortical region of interest and scripts required to produce results as presented in this manuscript are available upon request.

Received: April 30, 2014
Revised: June 13, 2014
Accepted: July 25, 2014
Published: September 4, 2014

References


